IEEE Signal Processing MAGAZINE

Volume 33 | Number 5 | September 2016

COMPUTATIONAL PHOTOGRAPHY AND DISPLAYS

Interfacing with the Visual World

Recent Advances in Phase Retrieval

The MPEG Internet Video-Coding Standard

Highlights from the IEEE SP Cup 2016 Student Competition

IEEE Signal Processing Society

◆IEEE

Contents

Volume 33 | Number 5 | September 2016

FEATURES

See the "President's Message" and "Society News" in this issue!

ON THE COVER

IEEE Signal Processing Magazine is uniquely positioned to convey and embrace an evolving scope of signal processing. This issue showcases an example of a cross-disciplinary area—fascinating advances of computational photography and display. A variety of articles showcase the potential for the field to revolutionize imaging and displays and transform the way in which we capture, share, and interact with the visual world around us.

COVER IMAGE: ©ISTOCKPHOTO.COM/SCYTHER5

PG. 149

COLUMNS

PG. 176

ICIP 2016 will be held in Phoenix, Arizona, United States,
25–28 September, at the Phoenix Convention Center.

©ISTOCKPHOTO.COM/TONDA

DEPARTMENTS

2

IEEE Signal Processing Magazine

SCOPE: *IEEE Signal Processing Magazine* publishes tutorial-style articles on signal processing research and applications as well as columns and forums on issues of interest. Its coverage ranges from fundamental principles to practical implementation, reflecting the multidimensional facets of interests and concerns of the community. Its mission is to bring up-to-date, emerging and active technical developments, issues, and events to the research, educational, and professional communities. It is also the main Society communication platform addressing important issues concerning all members.

Min Wu | Editor-in-Chief | minwu@umd.edu

Blurred Boundaries

What is signal processing and what isn't? From time to time, I would come across comments related to this question from independent reviews regarding whether an article that was submitted to *IEEE Signal Processing Magazine* fit the scope. I have seen reviewers recommending an article surveying signal processing techniques for wireless communications to a communications-related publication instead, or an article related to imaging or image analysis to a computer vision-related venue.

This would have been commonly accepted several decades ago when various fields under the IEEE umbrella were well partitioned into different technical Societies. The IEEE Signal Processing Society (SPS) traces its roots to 1948 as the IEEE's first Society, with the name and scope as the Professional Group on Audio of the then Institute of Radio Engineers (IRE), the predecessor of the IEEE. Looking at the historic roster of the SPS's technical committees (TCs) (as shown in "SPS Technical Committees Then and Now: Evolved and Broadened Scope of Signal Processing"), we can see that just 40 years ago in 1976 when the Society's first flagship conference, the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), was launched, there were only five SPS TCs, and they were dealing with either signal processing

SPS Technical Committees Then and Now: Evolved and Broadened Scope of Signal Processing

Year: 1976	**Year: 2016**
Digital signal processing	Audio and acoustic signal processing
Digital measurement of noise	Bioimaging and signal processing
Electroacoustic transducers	Design and implementation of signal processing and systems
Speech processing	Image, video, and multidimensional signal processing
Underwater acoustics	Information forensics and security
	Machine learning for signal processing
Year: 1987	Multimedia signal processing
Audio and electroacoustics	Sensor array and multichannel
Digital signal processing	Signal processing for communication and networking
Multidimensional signal processing	Signal processing theory and methods
Spectral estimation and modeling	Speech and language processing
Speech processing	Industry DSP technology standing committee
Underwater acoustics	Special interest groups: Big Data, Computational Imaging, and Internet of Things
VLSI for signal processing	

theories and methods, or speech and acoustics. There was little mention of communications nor much presence of visual processing in the SPS TC structure, even though many fundamental signal processing theories and techniques had been used rather extensively in those areas and further extended and adapted to solve the problems there. Moving forward about ten years, we see that visual aspects in a more general sense of multidimensional signal processing were added to the TC list by the mid-1980s. And now, 40 years later, we see more than a dozen TCs in the SPS, with a diverse range of "new" areas explicitly embraced, including communi-

cations and networking, multimedia signal processing, biomedical-related signal processing, and information forensics and security.

Year after year, signal processing, together with other fields, have evolved, and the boundaries between several traditionally separate fields have been blurred. Many colleagues have been actively involved in multiple technical societies, fostering interaction and bringing beneficial aspects between fields. Innovations have often happened at the boundaries between traditionally separate fields.

PRESIDENT'S MESSAGE

Rabab Ward | SPS President | rababw@ece.ubc.ca

The IEEE as "Bottom-Up" or "Top-Down" Management— The Choice Is Yours

The constitutional change amendment currently proposed by the IEEE is a matter of deep concern for so many of us. It affects some key factors that are critical to the future of the IEEE as we know it, and for which we have respect, devoting endless hours volunteering to its betterment.

The IEEE, as a nonprofit association run by volunteers, has very successfully operated for many years as a "bottom-up" confederation of technical Societies and geographic units, collaborating together in the spirit of our scientific and engineering culture. Where uniform direction is needed or highest-level resource allocation is required, decisions have been escalated up to higher-level boards, such as the Technical Activities Board (TAB), with representation from every IEEE Society and Council, or the Member and Geographic Activities (MGA) Board, with representation from every IEEE geographic Region. If needed, further escalation then occurs to the level of the IEEE Board of Directors (BoD).

Contrast this with most for-profit commercial companies, which operate from the "top down." That is, corporate executives, under guidance from their BoD, direct strategy and operations. Their employees, under various tiers of management, are responsible for executing these operations and policies.

Increasingly, there have been signs that the IEEE BoD, for whatever reasons, wants to become more of a top-down organization, concentrating greater power at the top, and diminishing the role of its technical and geographic units. As evidence, resources generated by Societies are increasingly redirected without Society control for use elsewhere in the IEEE, including for overhead purposes. As a result, remaining Society resources have either become stagnant or decreased, limiting what Societies can do for their members.

The latest move to more top-down control is occurring with this year's ballot, where members are asked to vote on a constitutional amendment that will abandon dedicated seats for technical Societies and geographic units on the BoD in favor of a smaller number of board members meeting "diversity" requirements that have not been defined and can be changed at any time according to whomever happens to be serving on the BoD then.

Proponents of the change claim that a smaller board will be more nimble. Opponents claim that the checks and balances of the widely represented board that we have today are more important and have served us well. Proponents

> **It is hard to believe that we are asked to vote on this constitutional change before we know what the final new IEEE structure would be, and without knowing the new bylaws that will govern the IEEE in the future.**

argue that the current board takes too much time to run the IEEE. Opponents argue that more decisions should be delegated to existing boards, like TAB and MGA, and empower them to make resolutions independently.

Many of us joined the IEEE due to the strength of its many diverse Societies. The combined effort of 45 technical Societies and Councils is responsible for 75–80% of IEEE revenues. To diminish the various Societies' visibility and role in running the IEEE is unwise and shortsighted. If volunteers find that their ability to control the destiny of their Society is greatly reduced, many volunteers will no longer feel that they truly belong to a self-empowered Society, and it will adversely affect the morale, motivation, and enterprising spirit of the volunteers.

The recent intense effort to change the IEEE's fundamental constitution is distracting us from solving our immediate, high-priority challenges. Further, it is hard to believe that we are asked to vote on this constitutional change before we know what the final new structure would be and without knowing the new bylaws that will govern the IEEE in the future. The new bylaws are to be written

later by the BoD, but none of the new bylaws will require member vote, or even prior notification. The proposed amendment gives power to the BoD to implement any changes they wish to make, without requiring approval from us, the IEEE Members.

More than half of the governing boards of the IEEE's Societies and Councils have already spoken against the amendment, including the Computer, Communications, Power & Energy, Circuits and Systems, Electron Devices, Robotics and Automation, Solid-State Circuits, and, of course, our own Signal

> **I urge you to become more familiar with the pros and cons of the amendment, and exercise your right to vote in this critically important juncture for the IEEE.**

Processing Society. But, their decisions do not matter—only yours does, as a voting member of the IEEE.

I urge you to become more familiar with the pros and cons of the amendment, and exercise your right to vote in this critically important juncture for the IEEE. You can learn more about the amendment at https://www.ieee.org/about/corporate/election/2016_constitutional_amendment.html. The rationale for opposition to the constitutional amendment and proposed restructuring can be found at https://ieee2016blog.wordpress.com.

For background, the IEEE governing documents, including the constitution and bylaws, are available at http://www.ieee.org/about/corporate/governance/index.html.

TAB has formed a TABin2030 Committee to consider the amendment's implications. Additional materials to the pros and cons and the TABin2030 webinars and analyses can be found by visiting http://ta.ieee.org/strategic-planning/tab-in-2030. You may need to log in with your IEEE account to access the materials.

SP

IEEE SIGNAL PROCESSING CUP 2017

GLOBAL UNDERGRADUATE COMPETITION IN SIGNAL PROCESSING

MUSIC SIGNAL PROCESSING: The 2017 SP Cup competition is a real-time beat tracking challenge. The beat is a salient periodicity of a music signal. Beat tracking underlies music information retrieval research and provides for beat synchronous analysis of music. The goal of this challenge is to implement a real-time beat tracker on an embedded platform and to demonstrate the performance with a creative output such as, but not limited to, drumming, dancing, or flickering lights. It is challenging to perform beat tracking in real time because the complete signal is not available and a system is expected to perform well on a wide variety of musical input.
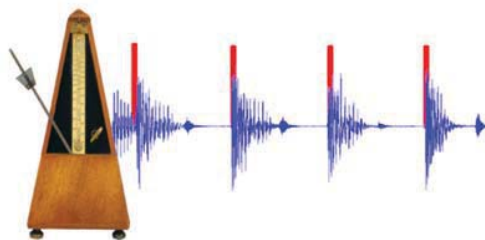
WHO CAN PARTICIPATE? Teams formed of 3 to 10 undergraduate students, at most one graduate student, and one faculty member.

PRIZES: GRAND PRIZE VALUED UP TO **$10K TOTAL**
Monetary prizes (up to $5000), plus travel grants for the top three teams to showcase their work at ICASSP 2017 – New Orleans, Louisiana, USA.

IMPORTANT DATES:
November 7, 2016:   Open competition – Stage 1 deadline.
December 21, 2016:   Open competition – Stage 2 deadline.
January 10, 2017:   Announcement of top three teams.
March 5-9, 2017:   Final competition held at ICASSP.

To learn more, visit: http://signalprocessingsociety.org/community/sp-cup/

SOCIETY NEWS

Election of Regional Directors-at-Large and Members-at-Large

Your vote is important! The election is now open for regional directors-at-large for Regions 1–6 and Region 8 (the term is 1 January 2017 through 31 December 2018) and members-at-large (term 1 January 2017 through 31 December 2019) of the IEEE Signal Processing Society (SPS) Board of Governors (BoG). Ballots, which have been mailed to SPS members, include a diverse slate of candidates for both elections, which were vetted by the SPS Nominations and Appointments Committee, as well as a space for write-in candidates. This year's election offers SPS members the opportunity to cast their votes via the web at https://eballot4.votenet.com/IEEE for up to one regional director-at-large for your corresponding Region: Regions 1–6 (United States) and Region 8 (Europe, Middle East, and Africa) and three member-at-large candidates. Ballots must be received

The candidates for member-at-large

Abeer A.H. Alwan

Homer H. Chen

Shoji Makino

Marc Moonen

Antonio Ortega

Beatrice Pesquet-Popescu

Gaurav Sharma

A. Lee Swindlehurst

Zhi (Gerry) Tian

The candidates for regional director-at-large

Regions 1–6

Radhakrishna (Radha) Giduthuri

Zhengdao Wang

Region 8

Sven Loncaric

John McAllister

Ana Isabel Pérez-Neira

at the IEEE no later than 3 October 2016 to be counted. Members must meet the eligibility requirements at the time the ballot data is generated to be eligible to vote. To be eligible to vote in this year's Society election, you had to have been an active SPS member or affiliate (excluding student member) prior to 1 August 2016. This is the date when the list of eligible Society voting members was compiled. The candidates for regional director-at-large are

■ *Regions 1–6*: Radhakrishna (Radha) Giduthuri and Zhengdao Wang
■ *Region 8*: Sven Loncaric, John McAllister, and Ana Isabel Pérez-Neira.

The candidates for member-at-large are

■ Abeer A.H. Alwan
■ Homer H. Chen
■ Shoji Makino
■ Marc Moonen

■ Antonio Ortega
■ Beatrice Pesquet-Popescu
■ Gaurav Sharma
■ A. Lee Swindlehurst
■ Zhi (Gerry) Tian.

The BoG is the governing body that oversees the activities of the SPS. The SPS BoG has the responsibility of establishing and implementing policy and receiving reports from its standing boards and committees and comprises 21 Society members: six officers of the Society who are elected by the BoG, nine members-at-large elected by the voting members of the Society, four regional directors-at-large elected locally by Society voting members of the corresponding region, as well as the Awards Board chair. The six officers are the president, president-elect, the vice president-conferences, vice president-membership, vice president-

publications, and vice president-technical directions. The executive director of the Society shall serve ex-officio, without vote.

Regional directors-at-large are SPS members who are elected locally by Society voting members of the corresponding Region via the annual election to serve on the Society's BoG as nonvoting members and voting members of the Society's Membership Board.

Members-at-large represent the member viewpoint in the Board decision making. They typically review, discuss, and act upon a wide range of items affecting the actions, activities, and health of the Society.

More information on the SPS can be found at http://www.signalprocessing society.org/.

SP

FROM THE EDITOR *(continued from page 3)*

And yet, worldwide, a number of signal processing programs and funding agencies still consider the narrow scope of signal processing as if it were the 1970s or 1980s. For example, in a country that I visited, the funding agencies and graduate programs treated image processing separately from signal processing. This unfortunately would constrain the breadth of students' preparation in their training and limit the potential source of innovations.

While we welcome and appreciate articles on core signal processing areas, our magazine is uniquely positioned to

> While we welcome and appreciate articles on core signal processing areas, our magazine is uniquely positioned to convey and embrace this evolving scope of signal processing.

convey and embrace this evolving scope of signal processing. Through engaging authors and publishing articles that reflect the crossbreed of multiple traditionally separated areas—both close and far—and bringing new advances from other areas that are of interest or benefit to signal processing professionals and vice versa, we hope the magazine will contribute to foster synergies and exchanges between areas and fields and help shape the future landscape of signal processing—the science behind our digital life.

This issue of the magazine showcases an example of a cross-disciplinary area—

a special issue on fascinating advances of computational photography and display. The timing also coincides with the IEEE International Conference on Image Processing (ICIP) to be held in Phoenix, Arizona, which features a visual technology showcase and visual innovation award.

As a final note, I would like to take this opportunity to thank all of the reviewers who have been so generous in offering their precious time and efforts in providing critical and constructive comments to the articles being considered by the magazine. We appreciate their contributions that are essential to the success of the magazine!

SP

SPECIAL REPORTS

John Edwards

Signal Processing Brings Fresh Capabilities to Digital Imaging

Research is leading to more sophisticated and useful imaging techniques

By speeding up processing and post-processing tasks, improving storage, and cutting costs, digital imaging has all but eliminated traditional film photography over the past several years. Now, with the help of increasingly sophisticated signal processing techniques, digital imaging is poised to become even more flexible and useful, generating images that would be far more difficult, if not impossible, to create with conventional analog photography.

On-the-go processing

Back in the days when film was the most popular photographic imaging medium, signs promising "same day processing" or "60-minute processing" could be seen almost everywhere. Most of today's smartphone and tablet photographers are far less patient, however. Not only do they demand to see their photos instantly, they want to have the ability to customize the images on the spot to make them look more attractive or to express a particular mood. Technology makes all of these wishes possible, but at a cost.

The seemingly inescapable drawback to mobile-device-based photo editing is that image correction and manipulation tasks are generally computationally intensive, capable of quickly draining a mobile device's relatively small battery. Several mobile apps

attempt to solve this problem by sending image files to a central server for off-site postprocessing. Yet this approach frustrates many users by creating significant delays as well as adding costs for increased data usage.

Late last year, researchers from the Massachusetts Institute of Technology (MIT), Stanford University, and San Jose, California-based Adobe Systems unveiled an experimental system that promises to slash the bandwidth consumed by server-based image-processing systems by as much as 98.5% and mobile device power consumption by up to 85%. The system works by sending a highly compressed version of the image to a central server. The server, after creating simple instructions for modifying the original image, then sends an even smaller file back to the mobile device, which uses the new data to create the enhanced image.

Applying the modifications to the original image does demand some extra processing power from the phone but not as much energy as uploading and downloading high-resolution files. In researchers' experiments, the energy savings were generally between 50 and 85%, and the time savings ranged between 50 and 70% (Figure 1).

Cloud image processing is often proposed as a solution to the limited computing power and battery life of mobile devices. "It allows complex algorithms to run on powerful servers with virtually unlimited energy supply," says Michaël

Gharbi, a graduate student in electrical engineering and computer science at MIT. "Unfortunately, this overlooks the time and energy cost of uploading the input and downloading the output images. When transfer overhead is accounted for, processing images on a remote server becomes less attractive, and many applications do not benefit from cloud offloading," noted Gharbi, who was the lead author on a recent paper describing the system. Gharbi's coauthors were his thesis advisor, Frédo Durand, an MIT professor of computer science and engineering; YiChang Shih, who received his Ph.D. degree in electrical engineering and computer science from MIT in March; Gaurav Chaurasia, a former postdoc in Durand's group who's now at Disney Research; Jonathan Ragan-Kelley, who has been a postdoc at Stanford since graduating from MIT in 2014; and Sylvain Paris, who was a postdoc with Durand before joining Adobe.

Gharbi describes the overall process as a transform recipe. "The main criterion in the design of transform recipes is that they are compact—less data to transfer over the network—but they should be able to represent faithfully many image transformations," he notes. "At equivalent quality, our recipes are much more compact than JPEG images."

Gharbi says that the process is relatively straightforward. "First, we decompose the input image into luminance and chrominance," he states. Luminance characterizes the brightness variations in
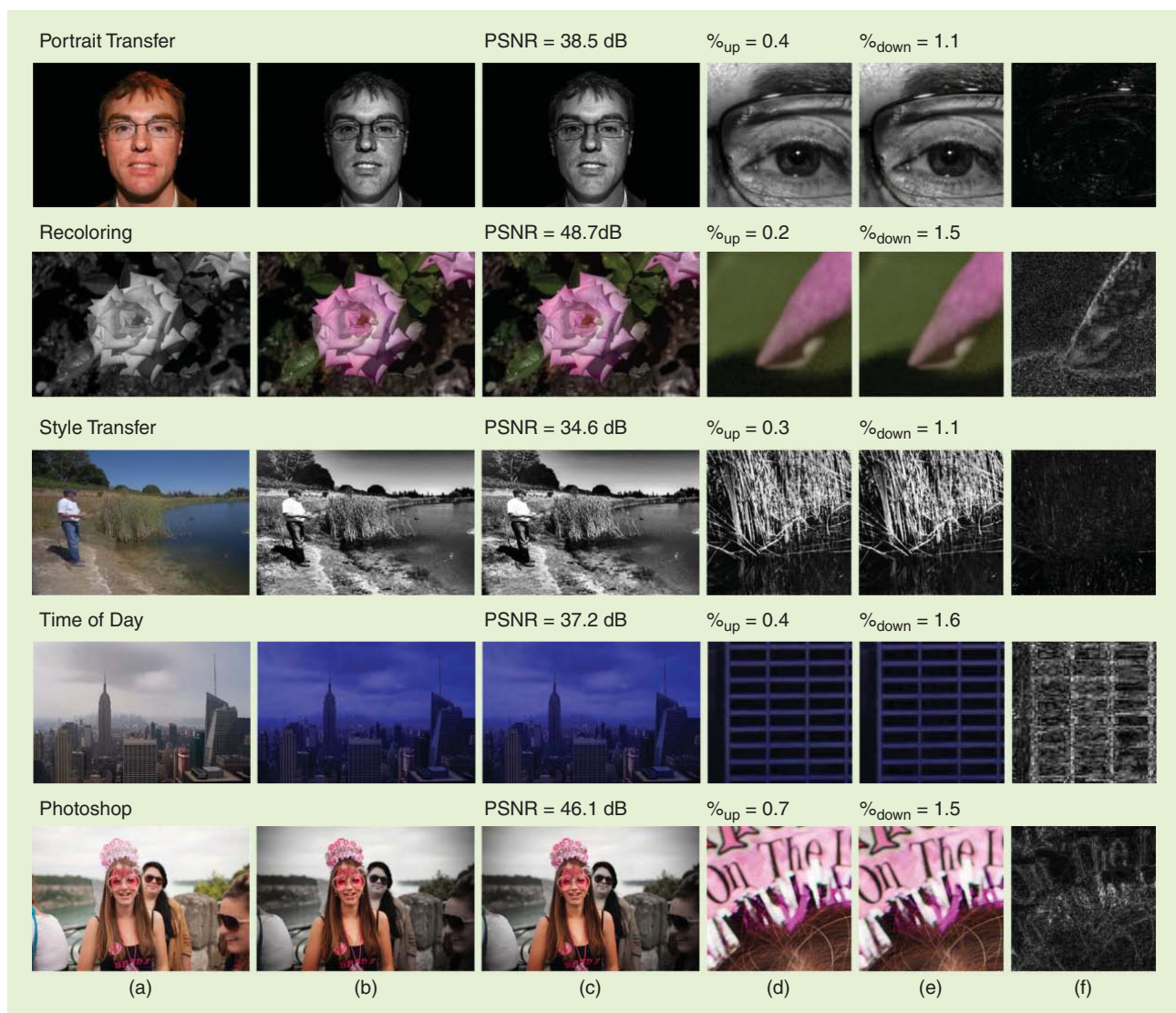
FIGURE 1. The results of various types of image processing created by an experimental system developed by MIT, Stanford University, and Adobe Systems researchers. The system promises to save bandwidth and lower mobile device power consumption. (a) Input. (b) Reference output. (c) Reconstruction. (d) Reference. (e) Highest error patch. (f) Rescaled difference.

the image (i.e., highlights and shadows), while chrominance is the colors of objects in the scene. "We process luminance and chrominance differently because the human eye is much less sensitive to errors in the chrominance than in the luminance," Gharbi says. "Therefore, we can afford lower fidelity in modeling the chrominance while favoring a good reconstruction of the luminance."

The luminance is then further decomposed into frequency bands using a Laplacian pyramid. Each band is processed differently. Low frequencies are slow variations in the image, such as the gradual shades in a sky at sunset. High frequencies are smaller details such as the

patterns of a fabric or small gravels. "We want to represent how each frequency band transforms to allow more for more complex transformations," Gharbi says. "We also add nonlinear mapping to model the luminance."

Once decomposed, the input is divided into small image patches, generally 64 x 64 pixels. "Within each patch, we use a regression model to predict the corresponding output patch," Gharbi explains. The regression is regularized by LASSO. "Now we have obtained a set of coefficients for each patch, and together they form the recipe."

The system uses conventional image compression techniques, i.e., quantiza-

tion and entropy coding, to further compress the recipe. Gharbi says that the overall process leads to results that are virtually indistinguishable from direct manipulation of the high-resolution image. Furthermore, the final bandwidth consumption is only 1–2% of what it would have been using the original phone image.

"Recipes are particularly useful as a replacement for expensive image filtering on mobile devices," notes Gharbi. "They are, essentially, a more efficient cloud off-loading procedure compared to the standard pipeline where cloud and server exchange full-resolution images."

Gharbi believes that the system could become even more useful over time as

image-processing algorithms grow increasingly powerful. "We see more and more new algorithms that leverage large databases to make a decision on the pixel. These types of algorithms don't do a very complex transform if you go to a local scale on the image, but they still require a lot of computation and access to the data, so that's the kind of operation you would need to do in the cloud," Gharbi says.

The researchers, who plan no future enhancements to the transform recipe, are now making it available to other imaging researchers via open source. "Our techniques showed that we can save energy and time on the mobile, and I believe others can build on this," Gharbi says.

Seeking clarity

Valerio Pascucci, an analysis and visualization professor at the University of Utah, doesn't have much use for thumbnail images because, for the past decade or so, he has focused almost exclusively on extreme resolution imagery, i.e., huge photographs containing anywhere from billions to hundreds of billions of pixels. Massive, high-density images are capable of providing a great deal of detail and useful information. Yet, it can take several hours to process the massive gigapixel images that intelligence analysts, physicians, engineers, and experts in many other fields rely on for critical insights. By developing software that requires only seconds to produce useful preview images, Pascucci and his coresearchers are helping such time-pressed individuals to obtain almost-

immediate access to critical visual information that, in many cases, can literally be life saving.

The technology, known as Visualization Streams for Ultimate Scalability (ViSUS), turns massive quantities of digital data into visual information that the human mind can understand. ViSUS works by sampling only a fraction of the pixels in a massive image, such as a satellite photo or a panorama made of hundreds of individual photos. According to Pascucci, the technology can produce good approximations or previews of what the fully processed image would look like.

The image-processing method can produce previews at various resolutions by taking progressively more and more pixels from the data that make up the entire full-resolution image. "The choice of algorithms was dictated by the need to visualize and analyze large image collections in real time," Pascucci says.

ViSUS also has the ability to blend hundreds to thousands of images into a single combined mosaic that can be used as a unique massive image (Figure 2). "To achieve this result, we managed to restructure a classical multiscale poisson solver into a pure coarse-to-fine approach that interactively blends a large number of images even if one can use only partial, coarse information," Pascucci explains. "While navigating through the data, more image information is acquired, and the result is incrementally improved. The approach is similar to Google Maps, where users can view more detail by zooming in on an image."

Pascucci says that updating signal processing algorithms for use in a pro-

gressive image-processing environment was essential for achieving scalability and interactivity. "The input data is typically not ready for use, and application of a traditional approach would introduce unacceptable delays," Pascucci notes. "It is, therefore, key to introduce real-time processing capabilities as part of the interactive data exploration."

A key challenge the researchers faced was changing traditional batch processing into an interactive progressive structure. "This means that dealing with larger images does not automatically mean increasing the delays in processing the data," Pascucci says.

The work by Pascucci and his team is part of a larger movement across the scientific community to come up with new ways to use and analyze information. Scientific instruments ranging from brain scanners to microscopes to telescopes are increasingly improving and generating larger amounts of high-quality data, creating the need for more efficient processing approaches. "This is a very active research project," states Pascucci. "As new data sources are developed, we find new challenges and reasons to expand our research base.

Compressed hyperspectral imaging

Conventional photographs, even those with a high resolution, only allow viewers to see what something looks like. Hyperspectral imaging, on the other hand, helps viewers to determine what something is actually made out of.

Ordinary photographic techniques create images spanning just three wavelengths of light, ranging from blue to green to red. Hyperspectral imaging can capture images across dozens or even hundreds of wavelengths. Akin to long-distance spectroscopy, such images can help viewers determine the types of materials found within an image.

Hyperspectral imaging has potential applications in areas such as security, defense, agriculture, and environmental monitoring. However, before the technique can fully enter the commercial and government mainstream, a serious challenge must be overcome, i.e., file size. If a conventional image contains millions of pixels spread across three wavelengths,

FIGURE 2. This panoramic mosaic of the Salt Lake Valley was taken by a camera mounted on a robotic panning device on top of a building at the University of Utah. It consists of more than 600 separate photographs that contain a total of 3.27 gigapixels (3.27 billion pixels) of image data. The seams between individual photos are readily apparent, as are the differences in light exposure. To edit the photos into a single, seamless, evenly exposed panorama would take hours using normal methods. ViSUS technology can complete the task in a fraction of that time.
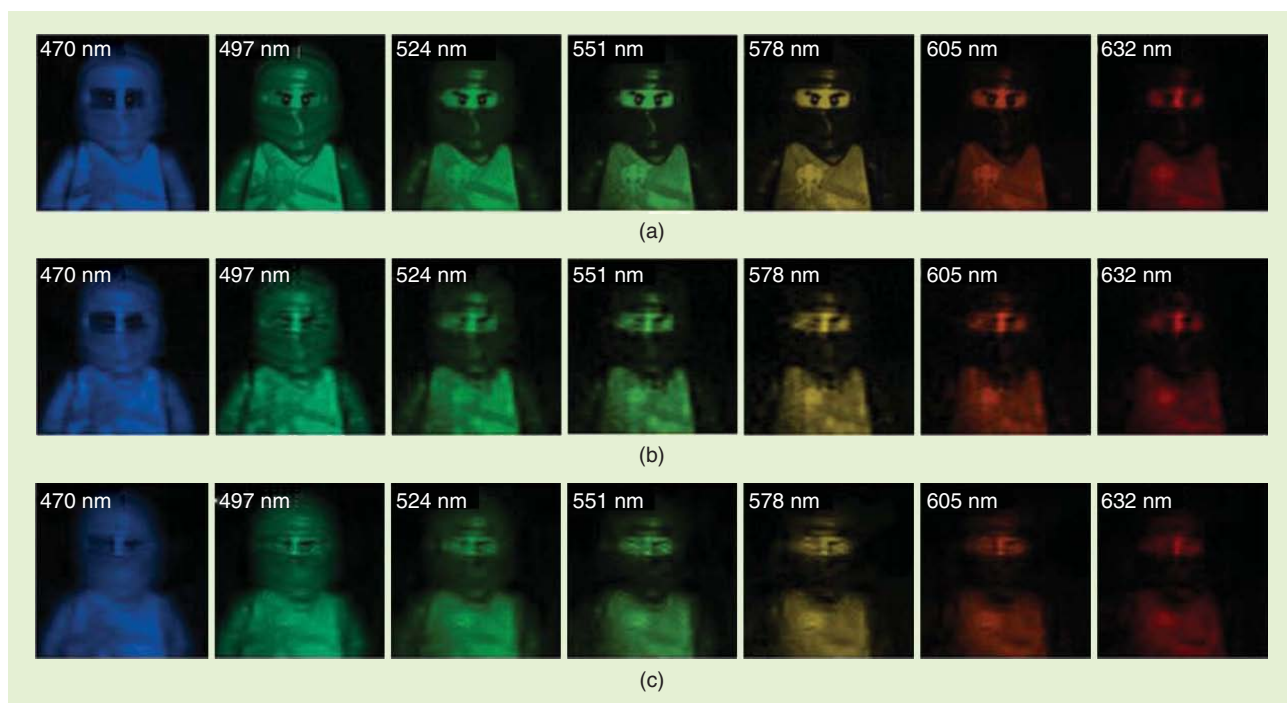
FIGURE 3. Images at wavelengths 470–632 nm within image cubes were reconstructed by the new algorithms from North Carolina State University and the University of Delaware. (a) represents the ground truth, (b) shows the output of the new algorithm, and (c) shows the output of the other algorithm.

it may consume approximately one or two megabytes of memory or storage space. Yet, a hyperspectral image file might be several orders of magnitude larger, making data storage and file transmission cumbersome, inconvenient, or, in some instances, completely impractical. Furthermore, capturing hyperspectral images across dozens of wavelengths is currently a time-consuming process, requiring conventional imaging technology to snap a series of images—each capturing a different suite of wavelengths or subsets of pixels with their corresponding wavelengths—and then combining the images into a single file. A final hyperspectral image is actually an image cube comprised of a collection of images (slices), where each two-dimensional image depicts one wavelength.

Earlier this year, researchers at North Carolina State University and the University of Delaware announced the creation of an algorithm that works with a popular compressive spectral imager to quickly acquire and accurately reconstruct hyperspectral images in less time and then store those images using less memory. According to researcher Dror Baron, an assistant professor of electrical and computer engineering at North Carolina State

University, the coded aperture snapshot spectral imager (CASSI) used in the research can acquire image data from different wavelengths simultaneously, which significantly accelerates the imaging process. Recognizing that CASSI provides significant improvements in both imaging quality and acquisition speed over conventional spectral imaging techniques, the researchers worked to further improve the system by accelerating the three-dimensional (3-D) image cube reconstruction process. They began by turning to an approximate message passing (AMP) framework.

"Our proposed AMP-3-D-Wiener used an adaptive Wiener filter as a 3-D image denoiser within the AMP framework," Baron notes. "AMP-3-D-Wiener was faster than existing image cube reconstruction algorithms and also achieved better reconstruction quality" (Figure 3).

"A big challenge was that AMP typically converts a noisy linear inverse problem where an unknown signal is observed through multiplying it by a matrix and adding noise into a much simpler problem where noise is added directly to the signal," Baron explains. The noise getting added to the signal in the new problem is supposed to be Gaussian and independent of the

signal. "However," Baron notes, "AMP has these properties when the matrix is well behaved, and, in our hyperspectral system, the matrix was very poorly behaved." This meant that the noise added to the signal in the new problem was not Gaussian and was also statistically dependent on the signal. "Therefore, we had to make some changes to the algorithm that gave a more desired performance."

"We were able to reconstruct image quality in 100 seconds of computation that other algorithms couldn't match in 450 seconds," Baron states. He's confident that computational time can be even further reduced.

The researchers' next step, Baron says, is to run the algorithm in a real-world system to gain insights into how the algorithm functions and identify potential room for improvement. "We're also considering how we could modify both the algorithm and the hardware to better compliment each other," Baron says.

Author

John Edwards (jedwards@john edwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area.

SP

FROM THE GUEST EDITORS

Amit Agrawal, Richard Baraniuk, Paolo Favaro,
and Ashok Veeraraghavan

Signal Processing for Computational Photography and Displays

Advances in imaging and displays have been the source of numerous scientific breakthroughs, as evidenced by the more than ten Nobel Prizes awarded for various inventions and discoveries in the imaging sciences. Over the last decade, imaging and displays have completely conquered our imagination, with mind-numbing statistics like more than a billion cameras being sold every year, more than a trillion photos taken each year, more than 200,000 images uploaded to Facebook alone every minute, and so on. In addition, interest and excitement around three-dimensional (3-D) displays, augmented reality displays, and virtual reality displays have grown exponentially over the last decade. This revolution in consumer imaging and displays is unparalleled and is a direct result of a decade of advancement in semiconductor fabrication technologies that have made image sensors and displays less expensive but with higher resolution every passing year.

Parallel to this imaging revolution, we have also witnessed a computing revolution with ever-more sophisticated algorithms, more computing, and storage horsepower available at the ready even on mobile devices. In addition, the ongoing cloud computing revolution is starting to provide consumers and other users with computing and data processing capabilities that have been, until now, unimaginable.

Ultra-Small Form-Factor Cameras
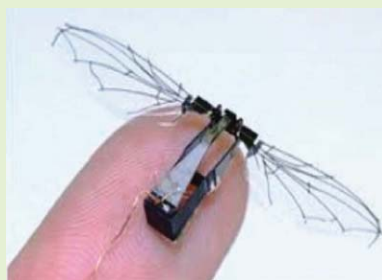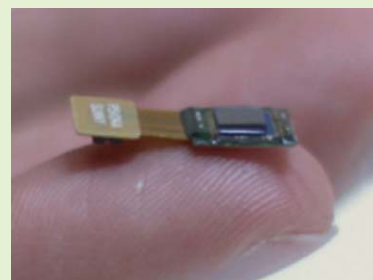
(a)   (b)

Cameras with ultra-small form factors, such as a (a) cubic-millimeter wireless sensing example surveyed in the article by Koppal on page 16) and (b) an enabling technology of lensless imaging in the article by Boominathan et al. on page 23).

Computational imaging and displays is a field at the confluence of these two fast-growing and ever-expanding disciplines. The field of computational imaging and displays seeks to create new imaging, photographic and display functionalities, and experiences that go beyond what is possible with traditional cameras and image processing tools. The key insight is that the codesign of sensor systems and signal processing algorithms to handle the sensor data provides several new degrees of design freedom, enabling im-

This issue of the magazine brings together a variety of articles covering an exciting range of application areas that showcase the potential for the field to revolutionize imaging and displays and transform the way in which we capture, share, and interact with the visual world around us.

aging and display systems to break traditional barriers. One of the most celebrated breakthroughs from this growing field is the Nobel Prize in Chemistry 2014, which was awarded for superresolution fluorescence microscopy, wherein imaging, fluorescent markers, and signal processing tools were used in concert to break the resolution limit imposed by diffraction. Another example is the emerging field of immersive 3-D displays, and virtual/augmented reality platforms such as GearVR, Hololens, and Oculus. These are examples of an early

Beyond 2-D Images

Technologies enabling advanced imaging and processing in computational photography: (a) high dynamic range imaging (see the article by Sen and Aguerrebere, page 36), (b) capturing and rendering computational appearance (see the article by Dana, page 70), (c) time-of-flight imaging (see the article by Bhandari and Raskar, page 45), and (d) light-field cameras (see the article by Ihrke et al., page 59).

breakthrough that comes out of the code-sign of imaging and computation, but there are several exciting and promising areas where similar approaches could break current barriers and unleash an era of transformative capabilities in imaging and displays for many many applications.

This special issue of *IEEE Signal Processing Magazine* brings together a variety of articles covering an exciting range of application areas showcasing the potential for the field to revolutionize imaging and displays and transform the way in which we capture, share, and interact with the visual world around us. The issue is organized into four topical areas:

- ultra-small form-factor cameras
- beyond two-dimensional images
- compressive imaging and displays
- novel applications.

Ultra-small form-factor cameras

The special issue begins with two articles that present radically small and novel form-factors for cameras, made possible by codesigning computation

with optics and sensors. The first article by Koppal presents a comprehensive overview of the current progress in this field and envisions the interaction between trillions of miniature cameras, which could provide novel sensing capabilities for agriculture, security, and health. The next article by Boominathan et al. goes even further and argues that, while lenses have remained the mainstay of conventional imaging and photography, lens-free computational imaging systems have begun to revolutionize many applications where the lenses provide undesirable form-factors and costs. This article discusses the promise and limitations of lens-based cameras and describes the body of work devoted to imaging without the use of lenses.

Computational imaging and displays is a rapidly evolving field, and we hope that the breadth and depth of coverage of this special issue will be a catalyst for identifying and tackling the major challenges facing the community.

Beyond two-dimensional images

The second set of four articles discusses how computational imaging enables modern imaging systems to go beyond capturing a two-dimensional photograph and capture multidimensional visual information, including high-dynamic-range (HDR), time-of-flight (ToF), multispectral, light field, and reflectance field. The first article in this topical area is an overview of methods for HDR imaging that exploits the joint design of optics, hardware, and processing, by Sen and Aguerrebere. Next, Bhandari and Raskar provide a thorough treatment of ToF imaging systems from a signal processing perspective, covering the history, mathematical foundation, and an overview of some recent results. Given the rapid developments and widespread adoption of ToF imagers, this is a

Compressive Imaging and Displays

Compressive imaging and displays: (a) imaging exploiting compressive sensing (see the article by Sankaranarayanan et al., page 81), (b) multispectra cameras (see the article by Cao et al., page 95), (c) millimeter wave imaging (see the article by Patel et al., page 109), and (d) factored display (see the article by Wetzstein and Lanman, page 119).

large depth of field images. Ihrke et al. illustrate the basic principles of light-field imaging by taking us through the key processing steps with more focus on compact plenoptic cameras. Finally, we end coverage of this topical area with a article by Dana that goes beyond two-dimensional images and recovers complete surface reflectance. In the article, she provides a comprehensive review of methods for capturing computational appearance, with a particular emphasis on computational imaging inspired techniques.

Compressive imaging and displays

Since the resolution, variety, and scale of visual data being collected has exploded, conventional sampling techniques based on Shannon–Nyquist are becoming impractical, resulting in the development of compressive imaging and display techniques that seek to use signal models and reconstruction algorithms to reduce the sampling requirements. The first article in this topical area is by Sankaranarayanan et al. and focuses on the importance of careful design of measurement matrices and appropriate models to perform compressive video sensing. Next up, Cao et al. consider the task of multispectral imaging—imaging with a much greater number of spectral channels than traditional trichromatic sensors,

timely and thought-provoking article. On a completely independent axis, light-field cameras have indelibly marked the field of computational photography and captured our imagination by

enabling the digital simulation of different camera settings (in the first place, digital refocusing), by animating pictures through viewpoint changes and simultaneously providing depth and

Novel Applications

Emerging applications: (a) gaze-contingent display (see the article by Stengel and Magnor, page 139) and (b) computational photography for cultural heritage (see the article by Huang et al., page 130).

thus providing spectral information at a higher level of detail. The massive scale of multispectral data—at high resolutions in the spectral, spatial, and temporal dimensions—has long presented a major challenge. This article presents an overview of computational multispectral imaging, from a signal processing perspective, covering a breadth of topics from system design, optimization to resolution, and noise analysis. While most of the focus in the special issue has been on imaging within the visible spectrum, Patel et al. expand upon how the recent developments in computational imaging have the potential to significantly improve capabilities of mm-wave imaging systems. They discuss the merits and drawbacks of available computational mm-wave imaging approaches and identify avenues of research in this rapidly evolving field. Finally, Wetzstein and Lanman show us the recent progress in compressive displays capable of producing a realistic 3-D visual experience of the world. At the core of these devices is the use of factorization techniques that provide an alternative to Lippmann's integral imaging and its inherent resolution/sampling limitations.

Novel applications

The special issue also has two articles that explore two novel application areas: digital preservation of cultural heritage and gaze-contingent computational displays. Cultural heritage provides an important opportunity for computational imaging, as it enables the noninvasive analysis of art through different imaging techniques and new ways to interact with art. Huang et al. present techniques to preserve art digitally, study and document it, and present it on modern media devices to provide new experiences. The final article in the special issue, by Stengel and Magnor, provides an overview of recent developments in computational display algorithms that exploit gaze-estimation to enhance perceived visual quality of conventional video footage when viewed on commodity monitors, projectors, or head-mounted displays.

Computational imaging and displays is a rapidly evolving field, and we hope that the breadth and depth of coverage of this special issue will not only serve as a reminder about the huge strides this community has made but also be a catalyst for identifying and tackling the major challenges facing the community.

Acknowledgments

Meet the guest editors

Amit Agrawal (amit-kagrawal@gmail.com) is a senior research scientist at Amazon Lab126. Before joining Amazon, he was a principal member of research staff at Mitsubishi Electric Research Labs, Cambridge, Massachusetts. He received his B.tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 2000 and his M.S. and Ph.D. degrees from the University of Maryland, College Park, in 2003 and 2006, respectively. His work has appeared in several technology publications including *American Scientist, Wired, New Scientist, Computer Graphics World Japan,* and *Engadget*.

Richard G. Baraniuk (richb@rice.edu) is the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University and the founder and director of OpenStax (openstax.org). He is a Fellow of the IEEE and American Association for the Advancement of Science and has received national young investigator awards from the National Science Foundation and Office of Naval Research; the Rosenbaum Fellowship from the Isaac Newton Institute of Cambridge University; the Electrical and Computer Engineering Young Alumni Achievement Award from the University of Illinois; the IEEE Signal Processing Society Best Paper, Best Column, Education, and Technical Achievement Awards; and the IEEE James H. Mulligan, Jr. Medal.

Paolo Favaro (paolo.favaro@inf.unibe.ch) is a full professor at the University of Bern, Switzerland, where he heads the Computer Vision Group. He received the laurea degree from Università di Padova, Italy, in 1999 and the M.Sc. and Ph.D. degrees in electrical engineering from Washington University in St. Louis, Missouri, in 2002 and 2003, respectively. From 2006 to 2011, he was a lecturer and then reader at Heriot-Watt University and honorary fellow at the University of Edinburgh, United Kingdom. His research fields include computer vision, computational photography, machine learning, and signal and image processing.

Ashok Veeraraghavan (vashok@rice.edu) is currently an assistant professor of electrical and computer engineering at Rice University, Texas. Before joining Rice University, he was a research scientist for three years at Mitsubishi Electric Research Labs in Cambridge, Massachusetts. He received his bachelor's degree in electrical engineering from the Indian Institute of Technology, Madras, in 2002 and his M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering at the University of Maryland, College Park, in 2004 and 2008, respectively.

SP

Sanjeev J. Koppal

A Survey of Computational Photography in the Small

Creating intelligent cameras for the next wave of miniature devices
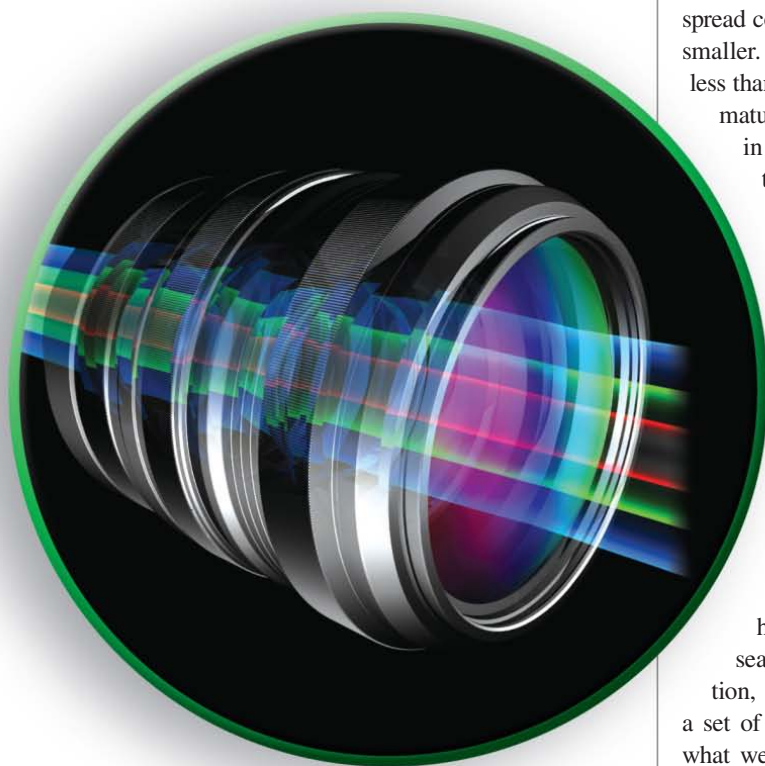
©ISTOCKPHOTO.COM/YAKOBCHUK

The sheer ubiquity of smartphones and other mobile vision systems has begun to transform the way that humans and machines interact with each other and the way that they interact with the world. Even so, a new wave of widespread computing is on the horizon, with devices that are even smaller. These are micro and nano platforms, with feature sizes less than one millimeter. These types of platforms are quickly maturing out of research labs, with some examples shown in Figure 1. These devices can potentially induce futuristic applications; for example, swarms of robotic flapping insects [29] could have applications in agriculture and security, while medical devices such as those described in [5] and [8] would enable body area networks and minimally invasive procedures. Devices such as those described in [1] are commercially available and could allow the creation of far-flung sensor networks.

Anticipating vision and imaging capabilities on these smaller platforms is a long-term prospect since, currently, none of the devices in Figure 1 even have cameras let alone full sensing systems. However, the possible impact is large since equipping tiny devices with computational cameras could help realize a new wave of applications in security, search and rescue, environmental monitoring, exploration, health, energy, and more. In this article, we outline a set of technologies that are currently converging to allow what we term computational photography *in the small*; i.e., across the millimeter, micro, and nano scales. This survey covers ongoing research that may break through existing barriers by combining ideas across computational photography, compressive sensing, micro/nano optics, sensor fabrication, and embedded computer vision. We map out the next research challenges whose solutions can propel us toward making miniature sensing systems a reality.

The broad architecture of miniature computational cameras is illustrated in Figure 1(b), where an array of (possibly heterogeneous) sensors are placed on a miniature low-power

From [29]   From [8]   From [5]   From [1]

(a)

(b)
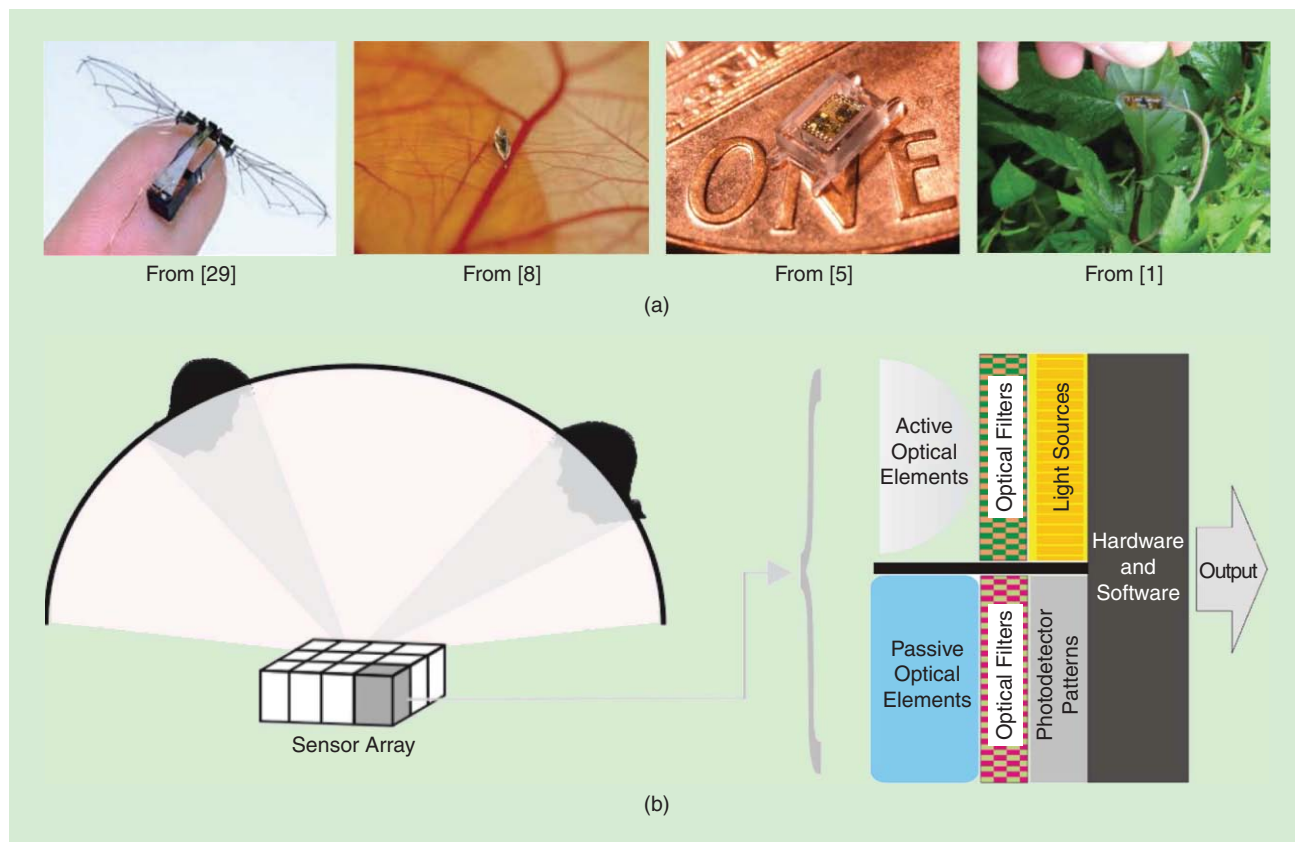
FIGURE 1. Miniature sensors: a new frontier for computational photography. In (a), a few motivating examples (images used courtesy of [1], [5], [8], and [29]) illustrate the coming, new wave of small machines that are transforming surveillance, medicine, sensor networks, agriculture, and other fields. Some, such as [1], are commercially available. However, due to restrictive power/mass budgets, none of these systems have cameras, let alone computational photography capability. If these devices could visually sense their environment, their impact would greatly increase. In this survey article, we cover relevant work in computational photography, compressive sensing, micro/nano optics, sensor fabrication, miniature displays, and embedded computer vision that together are defining the subdiscipline of computational photography in the small. In (b) we show the overall framework of such a miniature computational camera, where every sensor aspect, from optics to computing, is influenced by the visual task at hand.

platform. The design of each sensor can be optimized so that the computation is distributed across all aspects of the device, including passive optics to modulate the incoming light, active optics to project patterns onto the scene, optical filters for either polarization or wavelength as well as accompanying embedded hardware and optimized software. This comprehensive strategy can address the problem of achieving computational photography on compact devices.

Converging miniature sensor technologies: A brief history

In the last two decades, a few billion cameras became available to a large portion of humanity. This created a surge of interest and accompanying progress in a variety of imaging related technologies including, to name just a few, efficient hardware, small optical designs, miniature light-field sensors, and compact active illumination and displays.

We focus here on a brief history of three technologies in particular that have built the foundation for computational photography in the small. The first is the maturing of embedded vision sensing technologies, which includes both mass-produced low-power computing platforms from the mobile

revolution as well as specialized systems that intentionally blur the lines between computing hardware and sensing. The second is the impact of miniature optics for visual sensing, where display and imaging optics that were previously only created in research labs are now widely available. The third is the recent application of plenoptic designs to consumer cameras to allow for increased postprocessing control of photography.

Taken together, these fields have created the opportunity to make a new type of camera, as illustrated in Figure 2. This is a camera in which the visual task at hand can influence every aspect of the sensor, from the scene illumination and imaging optics to the sensing electronics and on-board processing. This allows for truly task-specific sensors that can extract every possible size, power, and mass efficiency from the system and can enable miniature computational cameras.

Embedded vision sensing and the mobile revolution

Processing images and video in real time on hand-held devices over the last two decades has resulted in a mature infrastructure for low-power vision and imaging. Dedicated imaging application-specific integrated circuits (ASICs), consisting of digital signal processors (DSPs), field-programmable gate arrays (FPGAs)
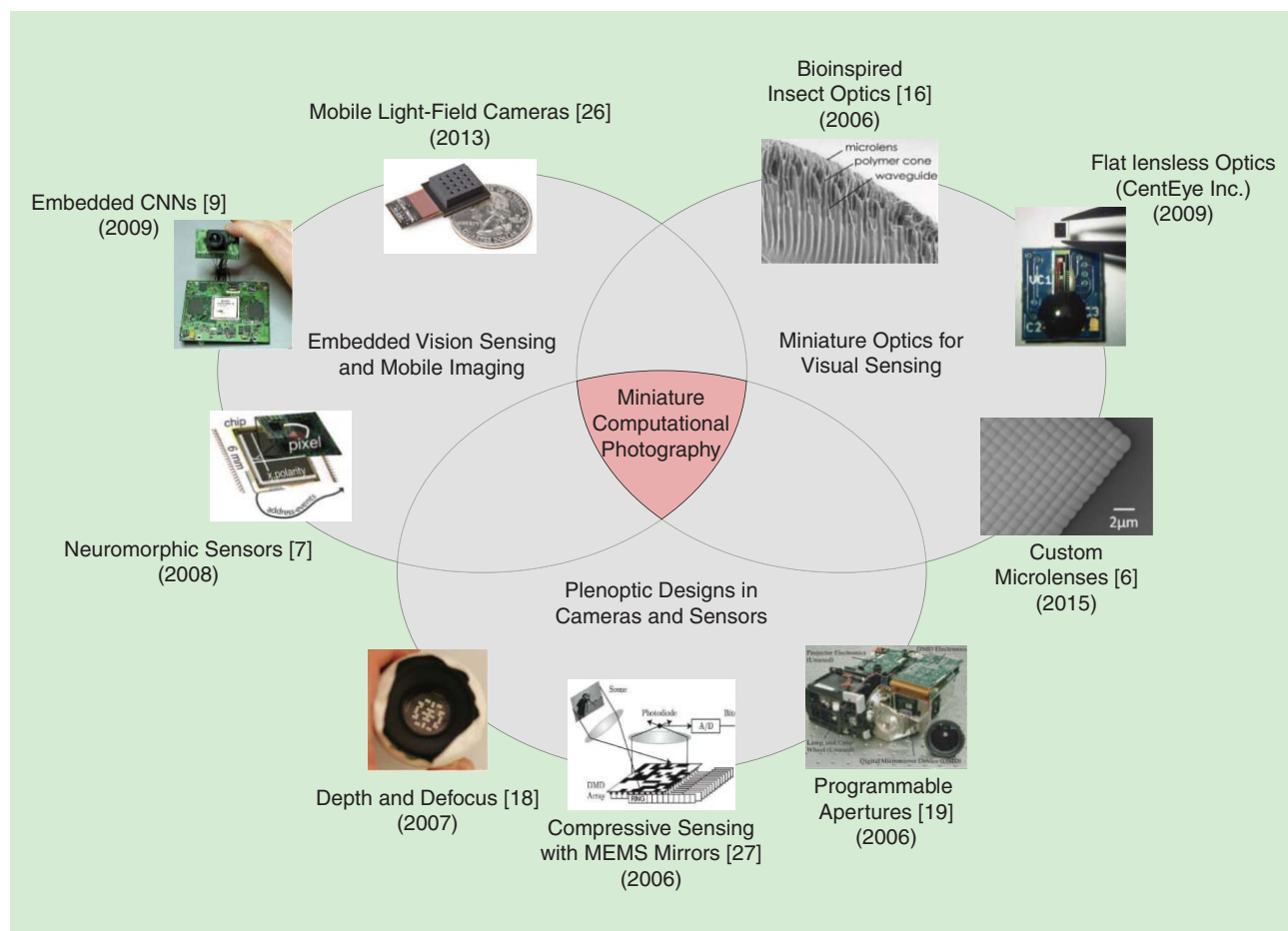
FIGURE 2. A convergence of miniature sensor technologies. We discuss the brief history of three sensor technology areas; embedded vision, miniature optics and plenoptic designs. Efforts in each area has built a library of mature techniques that allow us to build a type of camera where the energy cost of performing a visual task can influence every component in the camera architecture. [All images used with permission: [7], [9], and [27] courtesy of the IEEE; [18] and [26] courtesy of ACM; [16] courtesy of AAAC (Science); [6] courtesy of AIP; and [19] courtesy of Springer.]

and other processors are now standard in mobile devices, and much work exists in the embedded systems research community on low-power hardware support for vision [4]. For example, convolutional neural networks (CNNs) that have gained widespread use with their ability to exploit large data sets, were recently implemented on FPGA hardware with a peak power consumption of only 15 W [9]. In addition, many entrepreneurs are building mobile-scale light-field sensors [26].

The impact of vision and imaging on the mobile revolution cannot be overstated. However, as the anxiety about Moore's law suggests, such a strategy may not work for the type of extremely small devices shown in Figure 1. For such future applications, even a few watts is likely to be larger than what micro platforms are likely to support. For example, recent microscale body area networks have a per-node average power consumption of only 140 µW [14], and far-flung sensor networks have similar per-node requirements. For such scenarios,

> **In the last two decades, a few billion cameras became available to a large portion of humanity. This created a surge of interest and accompanying progress in a variety of imaging-related technologies.**

the paradigm of capture and postprocessing of images simply cannot offer enough power and mass savings.

Luckily, in addition to traditional embedded sensing research, there has been work done over the last few decades to build analogs to biological and neural architectures in vision systems. These devices perform computations at the sensor level, while photons are being converted into voltages and digitized into pixels. For example, [7] created sensors that automatically adjusted exposure pixel-wise. In this sense, these devices blur the line between sensing and computation since the sampling of voltages itself is part of the imaging algorithm. Many of these sensors have reached a mature level of development and some, such as those from Inilabs, are available commercially.

Miniature optics for visual sensing

Miniaturized optics has a long-standing impact in traditional fields such as microscopy. Micro and nano optics benefits the

rise of miniature computational photography since there now exists useful fabrication strategies [3]. However, most of the previous efforts in this area have been to create optics for generating sharp, high-quality imagery. For example, a variety of techniques exist to create micro-lenses by taking advantage of surface tension properties of PDMS and other materials that are heated and form lens shapes when in liquid form. Microlenses now form an integral part of many smartphone cameras, as they collect light within each pixel on the sensor. In research, a goal has been to create miniature optics that mimic insect eyes [16] or that offer shape control of microlenses [6].

> **Computational photography is about more than just capturing images but is also about exploiting the image-formation process to extract even more information from the world.**

While these previous efforts focus on the extremely useful goal of creating high-quality images, they cannot provide the full story. Computational photography is about more than just capturing images but is also about exploiting the image formation process to extract even more information from the world. It includes sampling the light-field, encoding the incoming light-rays and even analysis of the scene itself through filtering and optical convolutions. The fabrication technologies for creating micro-optics are useful for making computational cameras at small scales, but the design tools available require updating. For example, ray tracing softwares that model aberrations and image blurring and that assume a plano-parallel scene model are still the norm. However, geometric distortions reduce for small optics, and, instead, diffraction becomes important, posing both a challenge and an opportunity, as we will see in the next section.

Wide-angle fields of view (FOV) become important since narrow FOV miniature platforms must move to capture the surrounding visual field, which has power costs. However, wide-angle optics, while well understood at large scales, are not easily manufactured at the miniature scale. For example, miniature fish-eye lenses consist of multiple optical elements at cm scales with only 120° FOV being demonstrated. Curved mirrors allow panoramic imaging for computer vision applications and have no dispersion related problems; unfortunately, to the best of our knowledge, the state of the art for miniature mirrors does not appear to have a greater FOV than 45° [11].

Plenoptic designs in computational photography

Fourier optics [12] involves building optical systems to implement computations like Fourier transforms by, among other things, designing point spread functions (PSFs). For decades, such optical processing research resulted in the use of both coherent light and partially coherent light to build computing platforms that were meant to compete with silicon-based computers. Ten years ago, controllable PSFs began to appear in computer vision and computer graphics communities, where attenuating templates, assorted pixels and plenoptic designs created by standard photolithographic techniques, filtered scene radiance before measurement. For consumer

cameras, this allows image deblurring, refocusing [20], and depth sensing [18].

The key lesson learned by these early computational photography researchers was that important scientific questions involved the coded aperture patterns and the related decoding algorithms for images captured under these apertures. Making the coded aperture itself enjoyed the support of relatively established approaches, especially if the coded aperture in question was binary. At the millimeter scale, laser printing provided the required resolution. For smaller and more complex systems, photolithography techniques such as the 1 µm Heidelberg photomask writer could easily do the job. Therefore, many computational photography researchers became the new customers of the existing national nanotechnology infrastructure built during the 1990s and 2000s.

The plenoptic designs created by the aforementioned photolithography techniques were static and could not be changed over time. To create programmable optics, researchers took advantage of the wide availability of display related technologies for manipulating light, such as liquid crystal displays or digital micromirror devices that allow either controlled sampling of the light-field or processing of information for computer vision and image processing. Initially, these efforts required systems engineering; for example, in [19], the researchers hacked a Texas Instruments DLP projector, using it as a camera instead of a projector and whose "projected" patterns became the camera's coded aperture. Today, almost ten years later, the Texas Instruments developer kit is affordable enough that such hacking is no longer common. In fact, this availability has resulted in some of the most visible successes of compressive sensing [27] and continues to impact vision and imaging. This is a past example of the evolution and commodification of key technologies that we believe will happen in the future for many of the related areas summarized in Figure 2.

A first wave of computational photography in the small

There has been a recent surge of miniature computational cameras, and some of these are illustrated in Figure 3. The previous efforts we discuss here may lack integration, but they represent a new line of thinking that seeks to merge the intertwined technologies of plenoptic designs, miniature optics, and computational sensing in hardware and algorithms to create new types of cameras. Figure 3 depicts these on an axis of optical size and power consumption. Each of the authors cited reported their sensors' optical size, but, calculating the power footprint was more challenging since it is subject to interpretation and can change depending on the task at hand. For example, the raw images from a sensor could be used for optical flow directly, without much power consumption. However, the same sensor might require
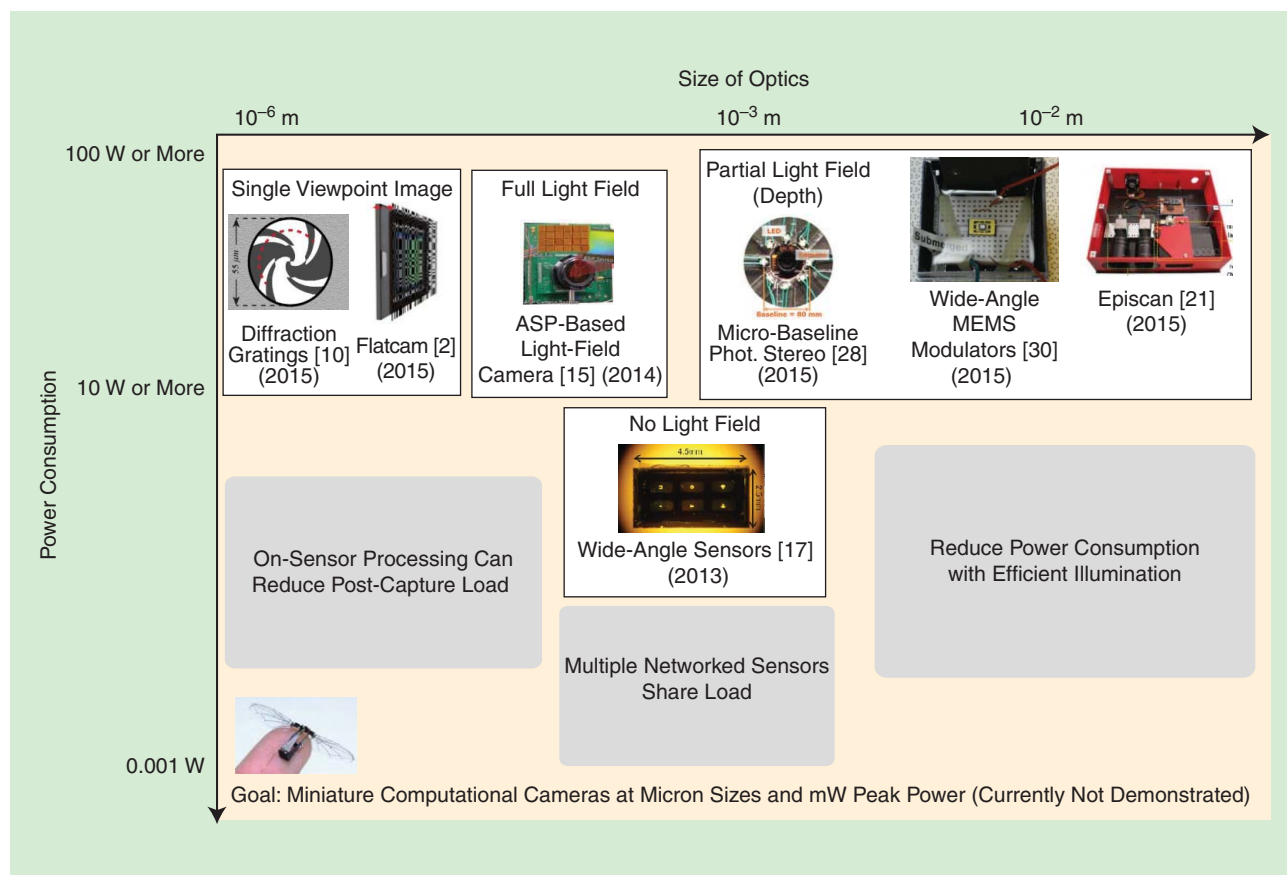
FIGURE 3. The first wave of miniature computational cameras. We organize the new wave of small computational cameras according to optical size and power consumption of the full system. Light-field cameras require powerful on-board computations, but the size of the optics and coded apertures has reached micron scales. On-board computation at millimeter scales has been proposed for vision sensors, but these do not capture the entire light-field. We illustrate the new broad steps such as applying sensor-based processing to reduce footprints, applying optical processing to share in the computational load and exploiting efficient active lighting to reduce on-board power consumption. [All images used with permission: [15], [17], [28], and [30] courtesy of the IEEE; [10] courtesy of Rambus/OSA; and [21] courtesy of ACM.]

multiple hours of PC-grade processing of the measurements to allow full light-field analysis. We picked the full power footprint required to generate the key result in each research paper.

A significant portion of this first wave of miniature computational photography has been in the realm of lensless imaging, which has long been valued due its simplicity, throughput, and potential for miniaturization. Recent novel image sensor designs recover angular information for light-field analysis [15]. Reference [10] also used lensless diffraction patterns to capture angular variations in the light field. Lensless imaging has played an important role in new types of compressive imagers [2]. Reference [17] demonstrated an angular theory of wide-angle optical processing and showed results for fiducial detection on small, autonomous robots, without needing to capture the entire light field.

Certain common ideas are shared among these first few forays into computational photography in the small. First,

> **The key lesson learned by these early computational photography researchers was that important scientific questions involved the coded aperture patterns and the related decoding algorithms for images captured under these apertures.**

diffraction is embraced, unlike much of conventional computational photography, which relies on a ray geometric model of light, albeit partially augmented with color and polarization. For example, [13] have shown the promise of adding micron-scale fabricated polarizing filters to CMOS/CCD cameras. Exploiting diffraction does not happen as in the optical processing community, where coherent or partially coherent models are used to obtain closed form solutions. Instead, to handle fully incoherent light from the real world, the relative effects of diffraction are used to infer scene properties. For example, in [15], angle sensitivity is obtained from the relative effects of a double decker layer of diffraction patterns. Another idea among these pioneering designs is the use of nonconventional optics and coded apertures fused together as one unit. For example, in [17], optical templates for detecting targets are embedded in a refractive slab, enabling the Snell's window effect, and allowing an extremely wide FOV without using fish-eye lenses.

The devices discussed above lie in the micro to millimeter scales and are passive in the sense that the coded apertures do not change over time and there is no controlled illumination projected onto the scene. This is in contrast to vision and graphics methods that use designed lighting to decode scene information and create new displays. Researchers have recently began to ask how these methods could work on miniature platforms. For example, a challenge on small devices is the inherent reduction in baseline. Reference [28] has shown how a circular setup can address some of these challenges for photometric stereo. Another direction to address the baseline issue is to move from triangulation to time-of-flight using active illumination. On the macro-scale, time-of-flight research has allowed the extraction of novel scene properties [25]. For miniature systems, trading off the modulated sources's power consumption versus the depth sensing becomes important.

One way to balance these needs and enable illumination-based sensing on small devices would be to extract a signal out of low wattage illumination. A new generation of computational illumination methods take advantage of low-power microelectromechanical systems (MEMS) mirrors that have been created for mobile hand-held projectors, such as those manufactured by Microvision, Syndiant, and Cremotech. For example, using a 5-W hand-held projector from Microvision, the authors of [21] have enabled computational illumination techniques in outdoor scenes, in the face of full sunlight. For miniature computational photography, the converse is clear; if there is no strong ambient illumination, then the same system can be made to work at orders of magnitude lower power budgets, since similar techniques of exposure synchronization and epipolar rectification can be harnessed to decrease power consumption.

While these methods prove promising, an interesting direction put forth by [30] is to engineer a wide-angle MEMS mirror modulator for enabling futuristic applications such as micro light detection and ranging (LIDAR) by demonstrating an electrothermal MEMS working in liquid for the first time. By submerging the MEMS mirror into a mineral oil whose refractive index is 1.47, a wide-angle optical scan ($>120°$) was achieved at small driving voltage (<10 V), and the scan frequency reached up to 30 Hz. The power consumption shown was 11.7 mW per degree in the mineral oil.

The next opportunities
Figure 3 depicts shaded gray regions that show the potential for further advances in efficiency and performance. For example, very few existing techniques take advantage of, say, computing in ASICs at the sensor level and many rely on conventional PC-based postimage capture processing. Task specific sampling may also reduce on-board processing; for example, a low-power face detector may have an optimal combination of thermal pixels, polarized pixels and skin

> **Miniature computational photography has great potential for applications in a variety of fields where small, networked platforms are already making an appearance, such as agriculture, security, health, and the Internet of Things.**

filter pixels to do the job. This requires exploiting the latest efforts in nano-optics, such as from [22], to use spectrally selective filters at the desired scales. Another goal is to find ways to exploit low-power programmable optical templates that use technologies such as eInk, which powers many e-readers and which remains static until sufficient energy is available for a pattern change.

Another potential opportunity is the integration of computational photography techniques with existing robotics and SLAM techniques for flying microrobots [24], floating sensors and surveillance drones. These tools could allow, for example, photometric stereo of large tourism sites or disaster zones by using varying illumination from multiple drones.

Temporal visual information at small scales can enable navigation, obstacle avoidance and optical flow; yet processing video on low-power platforms is prohibitive. CentEye (http://www.centeye.com/) has shown embedded computing based optical flow at high rates and at low resolution using embedded vision cameras. Integrating data from multiple sensors has enabled optical flow at real-time rates. For extremely fast sampling, it may be possible to exploit graded index lenses or optical fibers that can bend light in curves. Such optical elements can introduce time delay by guiding incoming scene radiance into optical loops, which can be tightly wound in a small volume, enabling, perhaps, fast capture of near simultaneous photographs without clocking at extremely high rates.

Finally, since true efficiency is only possible by having the sensing task at hand influence every part of the sensor, a fascinating question is how to distribute the work load over these different components. Should we sample and process with the optics, in such a way as to minimize the computational load? Or should we use a neuromorphic sensor to process the measurements as they are made? This suggests that design tools in the form of a compiler, to allow automatic partitioning of the computing problem into components that can be performed best by optics, coded sampling, on-board processing, or general-purpose signal processing and vision algorithms.

Toward full systems: Societal, legal, and cultural impact

We anticipate a future with trillions of networked miniature cameras. These computational cameras will be small, cheap, numerous, and capable of recovering more information about the world around them than today's conventional point-and-shoot cameras. The hypothetical impact of such devices has been discussed in many contexts, such as within the camera sensor network research community, and not all impacts may be desirable. For example, if these tiny sensors are not biodegradable, then the potential environmental impact may dwarf current concerns on e-waste. Another issue is privacy, as

miniature cameras may be discretely placed where their presence is unwanted. Blunt legal and societal restrictions to these types of small sensors may unintentionally harm the huge potential upside in terms of new applications and new platforms. Computational photography can provide answers to some of these challenges. For example, [23] proposes a new layer of optical privacy for small sensors, where optics filter or block sensitive information directly from the incident light-field before sensor measurements are made.

To conclude, we have shown that there is a confluence of technologies over the past few decades that has made the tools for enabling miniature computational photography possible. This has resulted in a recent surge of activity to build computational cameras, displays, and sensors that push the limits of size, power, weight, and mass. Miniature computational photography has great potential for applications in a variety of fields where small, networked platforms are already making an appearance, such as agriculture, security, health, and the Internet of Things. There are dangers regarding social acceptance of a trillion networked eyes around us, which can and should also be solved by computational photography research.

Author

Sanjeev J. Koppal (sjkoppal@ece.ufl.edu) received his B.S. degree from the University of Southern California in 2003. He obtained his master's and Ph.D. degrees from the Robotics Institute at Carnegie Mellon University (CMU). After CMU, he was a postdoctoral research associate in the School of Engineering and Applied Sciences at Harvard University. He is an assistant professor in the Electrical and Computer Engineering Department at the University of Florida (UF). Prior to joining UF, he was a researcher at Texas Instruments' Imaging R&D lab. His interests span computer vision, computational photography, and optics and include novel cameras and sensors, three-dimensional reconstruction, physics-based vision, and active illumination.

References

[1] Agrihouse. [Online]. Available: www.agrihouse.com

[2] M. Salman Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk. (2015). Flatcam: Thin, bare-sensor cameras using coded aperture and computation. [Online]. Available: arXiv preprint arXiv:1509.00116

[3] N. Borrelli, *Microoptics Technology: Fabrication and Applications of Lens Arrays and Devices*. Boca Raton, FL: CRC Press, 1999.

[4] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE J. Solid State Circuits*, vol. 33, no. 8, pp. 1199–1207, 1998.

[5] G. Chen, H. Ghaed, R. Haque, M. Wieckowski, Y. Kim, G. Kim, D. Fick, D. Kim, M. Seok, K. Wise, D. Blaauw, and D. Sylvester., "A cubic-millimeter energy-autonomous wireless intraocular pressure monitor," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, 2011, pp. 310–312.

[6] Y. Dan, K. Chen, and K. B. Crozier, "Self-aligned process for forming microlenses at the tips of vertical silicon nanowires by atomic layer deposition," *J. Vacuum Sci. Tech. A*, vol. 33, no. 1, pp. 01A109, 2015.

[7] T. Delbruck and C. A. Mead. "Adaptive photoreceptor with wide dynamic range," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1994, vol. 4, pp. 339–342.

[8] E. T. Enikov, M. T. Gibson, and S. J. Ritty, "Novel extrusion system for the encapsulation of drug releasing bio-medical micro-robots," in *Proc. ICME Int. Conf. Complex Medical Engineering*, 2009, pp. 1–6.

[9] C. Farabet, C. Poulet, and Y. LeCun, "An FPGA-based stream processor for embedded real-time vision with convolutional networks," in *Proc. IEEE 12th Int. Conf. Computer Vision Workshop*, Kyoto, Japan, 2009, pp. 878–885.

[10] P. R. Gill and D. G. Stork, "Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings," in *Proc. Computational Optical Sensing and Imaging, Optical Society of America*, Arlington, VA, 2013, pp. CW4C.3.

[11] C. Gimkiewicz, C. Urban, E. Innerhofer, P. Ferrat, S. Neukom, G. Vanstraelen, and P. Seitz, "Ultra-miniature catadioptrical system for an omnidirectional camera," in *Proc. Photonics Europe, Int. Soc. Optics and Photonics*, 2008, pp. 69920J–69920J.

[12] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw-Hill, 1968.

[13] V. Gruev, R. Perkins, and T. York, "CCD polarization imaging sensor with aluminum nanowire optical filters," *Opt. Express*, vol. 18, no. 18, pp. 19087–19094, 2010.

[14] B. Gyselinckx, C. Van Hoof, J. Ryckaert, R. F. Yazicioglu, P. Fiorini, and V. Leonov, "Human++: Autonomous wireless sensors for body area networks," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2006, pp. 13–19.

[15] M. Hirsch, S. Sivaramakrishnan, S. Jayasuriya, A. Wang, A. Molnar, R. Raskar, and G. Wetzstein, "A switchable light field camera architecture with angle sensitive pixels and dictionary-based sparse coding," in *Proc. IEEE Int. Conf. Computational Photography*, 2014, pp. 1–10.

[16] K. Jeong, J. Kim, and L. Lee, "Biologically inspired artificial compound eyes," *Science*, vol. 312, no. 5773, pp. 557–561, Apr. 2006.

[17] S. J. Koppal, I. Gkioulekas, T. Young, H. Park, K. B. Crozier, G. L. Barrows, and T. Zickler, "Toward wide-angle microvision sensors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 2982–2996, Jan. 2013.

[18] A. Levin, R. Fergus, and B. Freeman. "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 70, July 2007.

[19] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable imaging: Towards a flexible camera," *Int. J. Comp. Vision*, vol. 70, no. 1, pp. 7–22, 2006.

[20] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, July 2005.

[21] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, "Homogeneous codes for energy-efficient illumination and imaging," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 35, 2015.

[22] H. Park and K. B. Crozier, "Multispectral imaging with vertical silicon nanowires," *Scientific Rep*, vol. 3, pp. 2460, Aug. 2013.

[23] F. Pittaluga and S. J. Koppal, "Privacy preserving optics for miniature vision sensors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 314–324.

[24] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained mav," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2011, pp. 20–25.

[25] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nat. Commun.*, vol. 3, pp. 745, Mar. 2012.

[26] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: An ultra-thin high performance monolithic camera array," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 166, 2013.

[27] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarbotham, D. Takhar, K. Kelly, and R. Baranuik, "An architecture for compressive imaging," in *Proc. Int. Conf. Image Processing*, Atlanta, GA, 2006, pp. 1273–1276.

[28] J. Wang, Y. Matsushita, B. Shi, and A. C. Sankaranarayanan, "Photometric stereo with small angular variations," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 3478–3486.

[29] R. J. Wood, "The first takeoff of a biologically inspired at-scale robotic insect," *IEEE Trans. Robotics*, vol. 24, no. 2, pp. 341–347, 2008.

[30] X. Zhang, R. Zhang, S. Koppal, L. Butler, X. Cheng, and H. Xie, "Mems mirrors submerged in liquid for wide-angle scanning," in *Proc. 18th Int. Conf. Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, 2015, pp. 847–850.

SP

Vivek Boominathan, Jesse K. Adams, M. Salman Asif,
Benjamin W. Avants, Jacob T. Robinson, Richard G. Baraniuk,
Aswin C. Sankaranarayanan, and Ashok Veeraraghavan

Lensless Imaging

A computational renaissance
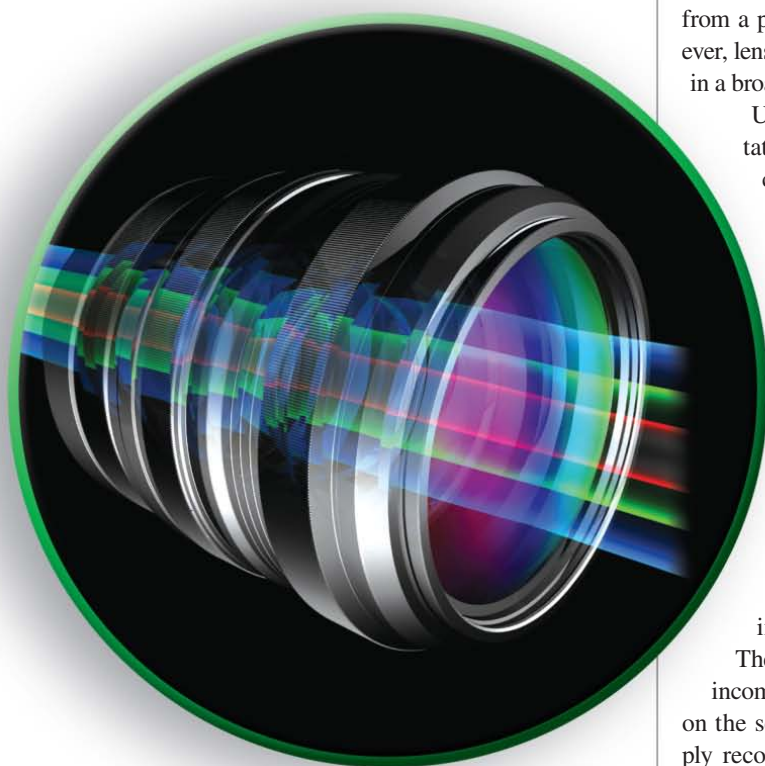
©ISTOCKPHOTO.COM/YAKOBCHUK

The basic design of a camera has remained unchanged for centuries. To acquire an image, light from the scene under view is focused onto a photosensitive surface using a lens. Over the years, the photosensitive surface has evolved from a photographic film to an array of digital sensors. However, lenses remain an integral part of modern imaging systems in a broad range of applications.

Unfortunately, lenses also introduce a number of limitations. First, while image sensors are typically thin, cameras end up being thick due to the lens complexity and the large distance required between the lens and sensor to achieve focus. For example, the thinnest mobile cameras today are approximately 5-mm thick, with the thickness increasing at larger lens aperture sizes. Second, lenses for visible light can be manufactured with inexpensive materials such as glass and plastic, but lenses for wavelengths farther into the infrared and ultraviolet spectra are either extremely expensive or infeasible. Third, lens-based cameras invariably require post-fabrication assembly, resulting in manufacturing inefficiencies.

In this article, we review a variety of alternate imaging approaches that completely eschew lenses. The primary task of a lens in a camera is to shape the incoming light wavefront so that it creates a focused image on the sensor. In the absence of a lens, a sensor would simply record the average light intensity from the entire scene. Lensless imaging systems dispense with a lens by using other optical elements to manipulate the incoming light. The sensor records the intensity of the manipulated light, which may not appear as a focused image. However, when the system is designed correctly, the image can be recovered from the sensor measurements with the help of a computational algorithm. Figure 1 shows the processes for capturing/reconstructing images in lensed and lensless systems. The simplest lensless imaging system is the pinhole camera. It is inefficient, however, since the small pinhole restricts the amount of light reaching
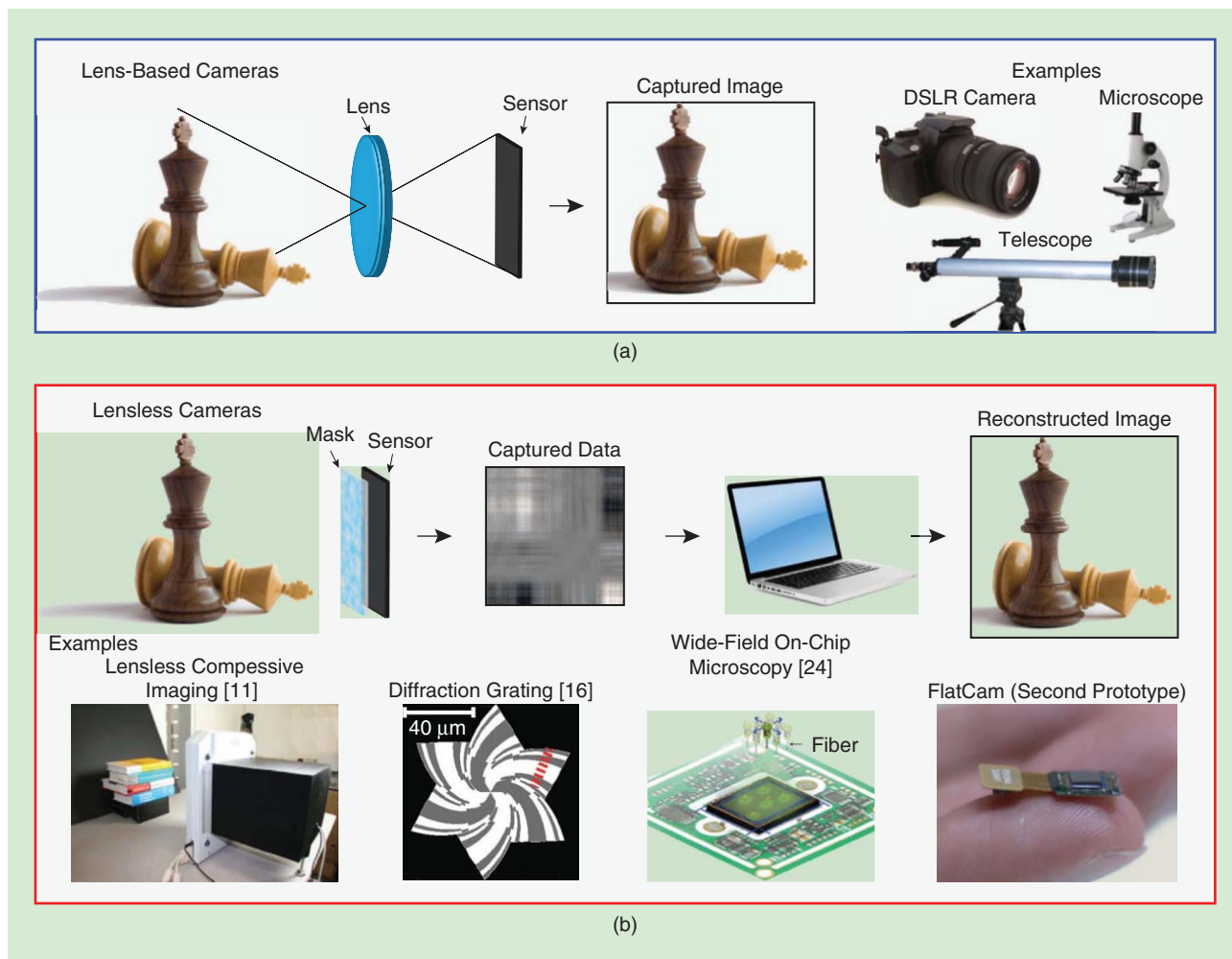
FIGURE 1. Lensed versus lensless imaging. (a) An illustration of a lens-based camera, where a lens maps the scene onto a sensor to form a clear image. A few examples of lens-based systems are shown. (b) The process of capturing an image using a lensless camera. An additional step of computation is required to reconstruct a clear image from the muddled sensor data. A few examples of lensless cameras [11], [16], [24] and our prototype based on [31] are also shown. (Figure adapted with permission from Macmillan Publishers Ltd: Nature Methods [24], copyright 2012.)

the sensor. Coded aperture cameras improve the light efficiency using a mask with an array of pinholes. The sensor measurements become a superposition of the images formed by each aperture, and the computational recovery algorithm's task is to reorganize the measurements to recover the image.

There are many benefits to going lensless;

- **Scalable fabrication.** Lensless cameras can be directly fabricated using traditional semiconductor fabrication technology. For example, a multiple-aperture mask can be fabricated either directly in one of the metal interconnect layers or on a separate wafer thermal compression that is bonded to the back side of the sensor, as is typical for back-side-illuminated image sensors [1]. Thus, lensless cameras can benefit from all of the scaling advantages of semiconductor fabrication, resulting in a low-cost, high-yield, high-performance device. In contrast, conventional cameras require inefficient post-fabrication assembly of the lens system.
- **Thin form factor.** Since the standoff distance between a multiple-aperture mask and the image sensor array need only

be a few tens to hundreds of microns, an entire lensless camera can be only a few tens to hundreds of microns thick—resulting in potentially the thinnest cameras ever produced.
- **Wavelength scaling.** Lensless cameras have been used for X-ray, gamma-ray, and astronomical imaging for decades. Lensless imaging in the visible, short-wave infrared (SWIR), and thermal wavebands is relatively new. Moreover, the technology can also be expanded to the mm-wave, terahertz, and other bands with minimal modifications, providing unmatched spectral flexibility.
- **Low cost.** While the cost of high-resolution cameras has fallen rapidly in the visible range, it remains high outside the visible range (e.g., infrared). One reason is that lenses for these wavelengths must be manufactured using expensive materials. By doing away with lenses and the need for post-fabrication assembly, lensless cameras promise significant cost reductions for imaging outside the visible spectrum.
- **Nonplanar geometries.** Lensless cameras can be adapted to arbitrary sensor geometries, including not just planar but

also cylindrical, spherical, and even flexible sensors. The compact form of spherical lensless cameras promise unmatched maneuverability in constrained environments such as endoscopy.

■ **Light throughput.** Lensless cameras can be designed to have very large input apertures, which translates into improved light efficiency and a much larger field-of-view than conventional lens-based systems.

■ **Three-dimensional (3-D) imaging.** Lensless imaging systems can extract 3-D and refocusing information in addition to two-dimensional (2-D) imaging. Although this ability is not yet competitive with existing lens-based techniques, such as light-field and time of flight, the extracted 3-D information may still be useful in some contexts, such as gesture identification.

In this article, we review the past, present, and future of lensless imaging as a shining example of the opportunities afforded by *computational imaging*, a design framework that uses computational algorithms to replace or augment imaging hardware (in this case replacing the lens). After reviewing classical and contemporary approaches to lensless imaging, we introduce and analyze a mathematical model that exposes the key issues underlying these architectures. The bulk of the article consists of a case study of the *FlatCam*, a particular mask-based lensless imager we have developed.

Early lensless imaging systems

Pinhole cameras

The very first cameras were lensless. Pinhole cameras, also known as the *camera obscura*, were discovered centuries before the invention of lenses and photography. Pinhole cameras have been well known since Alhazen (965–1039 A.D.) and Mozi (c. 370 B.C.). However, the first photograph using a pinhole camera was captured in 1850. Pinhole cameras offer a simple and elegant architecture for lensless imaging that consists of a single aperture in front of a sensor. Light from an object passes through the pinhole and forms its image on the sensor. However, a tiny pinhole is required to produce sharp images, which results in very low light throughput. As a consequence, a pinhole camera requires very long exposure times to acquire images at high quality. Lenses were introduced into cameras for precisely the purpose of increasing the size of the aperture, and thus the light throughput, without degrading the sharpness of the acquired image.

Coded aperture cameras

Coded aperture cameras extend the idea of a pinhole camera by replacing the small, single aperture with a mask containing multiple apertures [2]–[4]. Coded aperture cameras were originally invented for imaging with X-rays and gamma rays, wavelengths of light that are not easily amenable to lens-based imaging (see

"Coded Aperture in X-Ray and Gamma-Ray Imaging"). In a general coded aperture system, sensor measurements represent a superposition of the images formed behind each pinhole. The primary motivation for a coded mask is to increase the light throughput while retaining the ability to reconstruct high-resolution images. For instance, if the mask contains P pinholes, then the sensor image is the sum of P overlapping images of the scene. The signal-to-noise ratio in such an image is approximately $\sqrt{P}$ times better than a single pinhole image [2], [3].

In contrast to a single-pinhole camera, the sensor measurements of a coded aperture camera do not resemble an image of the scene. Rather, each light source in the scene casts a unique shadow of the mask onto the sensor, encoding information about locations and intensities. Consider a single light source on a dark background; the image formed on the sensor will be a shadow of the mask. If we change the angle of the light source, then the mask shadow on the sensor will shift. If we change the depth of the light source, then the size of the shadow will change. We can represent the relationship between the scene and the sensor measurements as a linear system that depends on the pattern and placement of the mask. Inverting this system using an appropriate computational algorithm will recover an image of the scene.

> **Lenses were introduced into cameras for precisely the purpose of increasing the size of the aperture, and thus the light throughput, without degrading the sharpness of the acquired image.**

The design of the mask plays an important role in coded-aperture imaging. An ideal pattern would maximize the light throughput while providing a well-conditioned scene-to-sensor transfer function to facilitate inversion. In this regard, several mask designs have been proposed in coded aperture literature, including Fresnel zone plate, random pinhole patterns, uniformly redundant arrays (URAs) [3], and their extensions. URAs are particularly useful because of two key properties: 1) almost half of the mask is open, which boosts the signal-to-noise ratio, and 2) the autocorrelation function of the mask is close to a delta function, which aids in calibration and image recovery. URA patterns are closely related to the Hadamard-Walsh functions and the maximum length sequences that are maximally incoherent with their cyclic shifts [5].

Zone plates

A zone plate can also be used to focus light and form an image using diffraction [6], [7]. A zone plate consists of concentric transparent and opaque rings (or zones). Light hitting a zone plate diffracts around the opaque regions and interferes constructively at the focal point. Zone plates can be used in place of pinholes or lenses to form an image. One advantage of zone plates over pinholes is their large transparent area, which provides better light efficiency. In contrast with lenses, zone plates can be used for imaging wavelengths where lenses are either expensive or difficult to manufacture [8], [9].

Contemporary lensless imaging systems

Recent advances in sensor technology [in particular, the conversion from analog film to digital charge-coupled device (CCD)

Coded Aperture in X-Ray and Gamma-Ray Imaging

Coded aperture cameras were originally invented for X-ray astronomy [2], [34], and they have been primarily used for X-ray and gamma-ray imaging since then [3], [4], [32], [35]. For instance, the SWIFT space telescope (see Figure S1) is a multiwavelength space telescope currently in use for observing gamma-ray bursts [36].

Image formation in a lens-based camera can be viewed as a one-to-one mapping of points at a focal plane in the scene onto a sensor. A lens is a refractive element that manipulates light wavefronts such that all the light coming from a certain direction in the scene converges to a particular location on the sensor. Visible light can be easily manipulated using transparent materials, such as glass and plastic, that have a large refractive index. Therefore, lenses for visible light are easily available at low cost.

High energy radiation beyond the visible spectrum, such as X-rays and gamma rays, are routinely acquired in radiology, screening, and astronomy applications. Imaging these radiations enables us to look inside a human body for medical diagnosis, screen luggage at the airports, and observe black holes and supernova in the cosmos. However, X-rays and gamma rays are not as easy to manipulate with refractive optics as visible light. Therefore, the methods for imaging high-energy radiations primarily rely on reflection or diffraction optics.

The classical imaging architectures for X-rays and gamma rays use a collimator in front of a sensor. A collimator typically consists of a thick sheet of lead or other material opaque to the incoming rays with multiple holes. Every sensor pixel behind a hole has a narrow field of view, since only a small cone of light in a particular direction can travel through each hole. Thus, a collimator localizes the directions of the rays

SWIFT/BAT Imager with Coded Aperture Mask

FIGURE S1. SWIFT is a multiwavelength space observatory dedicated to the study of gamma-ray bursts. Its burst alert telescope (BAT) uses a coded mask to detect gamma ray burst events and compute their coordinates in the sky. The D-shaped coded aperture mask is made of nearly 54,000 lead tiles [36]. (Images courtesy of NASA.)

reaching the sensor. Light from multiple locations and angles can be recorded by moving the collimator and the detector accordingly. The two primary drawbacks of collimator-based imaging are 1) light throughput is extremely low, since the collimator allows only a fraction of incoming light to reach the sensor, and 2) the recorded image has a low angular resolution, because every sensor pixel records the average intensity of light over its entire field of view.

A coded aperture-based imaging system offers better light efficiency and angular resolution as compared to either a pinhole- or collimator-based system. A coded aperture camera consists of a mask with transparent and opaque features placed in front of a sensor. Light from any particular location in the scene casts its unique shadow of the mask on the sensor plane. Therefore, each sensor pixel records a coded multiplexing of light from multiple scene locations. The relationship between the sensor measurements and the scene intensities can be described as a linear system, which can be solved using a computational algorithm.

and complementary metal-oxide semiconductor (CMOS) sensor arrays], image reconstruction models and algorithms, and computing resources have made lensless imaging a burgeoning field. Here, we briefly review some of the recent research in this area.

Lensless imaging using programmable apertures

Programmable mask-based lensless imaging designs have recently been proposed in [10]–[12]. The camera proposed in

[10] consists of a sensor and layers of programmable spatial light modulators (SLMs) whose transmittances are controllable in space and time. By applying different patterns in each layer, the incoming light can be manipulated in a number of ways. For example, the camera can track a moving object by shifting a pinhole in one of the layers, select and capture disjoint regions in the scene, or perform computations on the scene and record the results directly on the sensor.
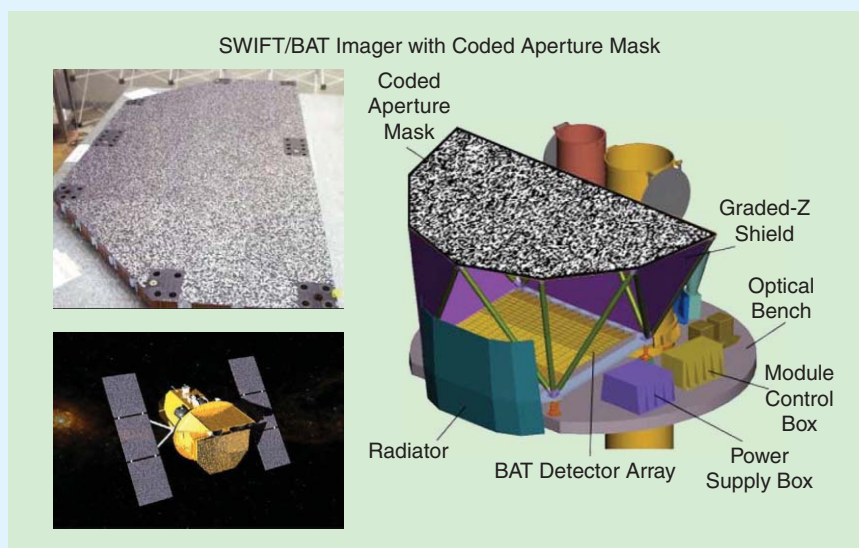
The lensless camera in [11] (the first example of lensless in Figure 1) uses compressive sensing principles to capture and recover images. It consists of a single programmable SLM and a single pixel detector. It captures multiple measurements of the scene by changing the mask pattern. The scene is then reconstructed by solving a sparse recovery program. Using multiple pixel detectors, this design can reconstruct a higher resolution image for a planar or a sufficiently distant scene [13].

The camera in [12] consists of a sensor array and an SLM implementing a separable mask pattern. This camera can reconstruct the scene using a single sensor image, but the reconstruction quality improves using multiple sensor images with different mask patterns. In the development of this camera, the authors showed that traditional techniques [3] of using URA and modified URA (MURA) aperture patterns fail due to significant diffraction effects in the visible spectrum.

Ultra-miniature lensless imaging with diffraction gratings

Ultra-miniature cameras (approximately $100\,\mu$m width and thickness) have been implemented in [14]–[17] using integrated diffraction gratings and CMOS image sensors. The pixels in [14] use diffraction gratings over a photodiode in order to be sensitive to the angle of incident light. The angle selectivity is achieved due to a phenomenon called the Talbot effect [18] and enables the camera to perform lensless 3-D imaging in the near field. The gratings were fabricated as metal wiring layers over the photodiodes.

The phase gratings in [16] are designed such that they impose spiral-shaped diffraction patterns (the second example for lensless in Figure 1) on the sensor array. The diffraction pattern is etched on a refractive medium placed above the sensor. The spiral pattern can also be viewed as the point spread function of these imaging systems. Similar to a coded aperture system, the image formed on the sensor is a superposition of shifted and scaled spiral patterns. However, in contrast to an amplitude mask, a phase grating-based mask has improved light efficiency, since it blocks much less light. While an image of the scene can be recovered using a computational algorithm, the primary purpose of these small-size and low-cost designs is distributed monitoring and inspection (for example, in the Internet of Things).

Lensless microscopy via shadow and diffraction imaging

Lensless cameras have also been successfully demonstrated for several microscopy and lab-on-chip applications. We can divide the lensless microscopes into two broad categories: contact-mode shadow imaging-based microscopes [19]–[21] and diffraction-based lensless microscopes [22]–[27]. In a shadow imaging-based microscope, a microscopic sample is placed extremely close to a sensor array (ideally within $1\,\mu$m) so that diffraction is minimized. Light from an illumination source passes through the sample and casts a shadow on the sensor with unit magnification. The shadow image represents the image of the microscopic sample under observation. It is also possible to capture multiple images of a sample with subpixel shifts for the purpose of digital superresolution. The on-chip
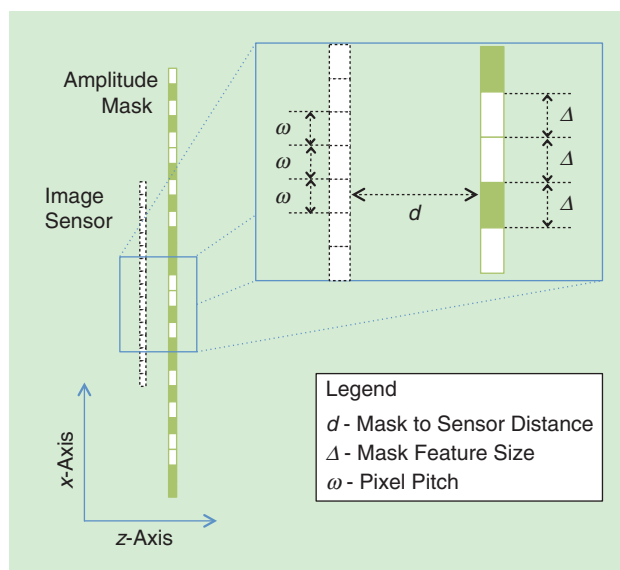
FIGURE 2. A schematic of a lensless imager using a single amplitude mask.

microscope in [20] demonstrated imaging of red blood cells at a resolution of 600 nm by combining multiple low-resolution shadow images of blood flowing in a microfluidic channel.

Diffraction-based lensless microscopes allow a significant distance between the sample and the sensor plane. Light scattered by the sample interferes with itself and creates an interference pattern on the sensor (the third example of lensless in Figure 1). These interference patterns can be digitally processed to reconstruct an image of the sample [24], [25]. The on-chip microscope in [25] demonstrated imaging of red blood cells at a resolution less than $7\,\mu$m with a field-of-view of 20.5 mm^2. Since the optical sensor records only the intensity of the interference patterns and loses the phase information, image reconstruction relies on computational methods for phase retrieval [28], [29].

A mathematical model for lensless imaging

A simple mathematical model can be used to explain, characterize, and analyze the operation of a variety of lensless imagers.

Lensless imaging architecture

Consider the imaging architecture in Figure 2, which consists of an amplitude mask placed in front of an image sensor. Both the sensor and the mask are assumed to be planar and parallel to each other. The mask is placed a distance d (typically measured in microns) in front of the sensor; hence, we can assume the sensor is placed on the plane $z = 0$ and the mask on the plane $z = d$. Assume, without loss of generality, that the mask is binary-valued and consists of opaque and transparent elements that either block or transmit light. An important variable is the smallest feature size on the mask, Δ; intuitively, the binary mask is constructed using opaque or transparent building blocks of size $\Delta \times \Delta$. Denote the pixel pitch, or the size of individual sensor pixels, by w. Given this basic setup, we can characterize the spot size produced by a mask element and

characterize when the spot can be well-approximated using geometric (ray) optics.

Image formation

We characterize image formation using the geometric optics model. While this approach largely ignores diffraction, the resulting model is useful for the design and analysis of well-conditioned imaging architectures. Furthermore, the calibration procedure that we detail in subsequent sections can account for unmodeled diffraction effects. For the simplicity of notation, we assume a simplified 2-D world imaged by a one-dimensional (1-D) mask and sensor. The extension to a 3-D world imaged by a 2-D mask and sensor is straightforward except where stated otherwise.

For a suitably defined scene irradiance vector $\mathbf{x} \in \mathbb{R}^N$, the scene-to-sensor mapping can be described using the linear set of equations

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\Phi \in \mathbb{R}^{M \times N}$ is the measurement matrix, $\mathbf{y} \in \mathbb{R}^M$ is the image formed on the sensor, and e is measurement noise. This model can be interpreted in two different ways:

1) Each sensor measures a weighted, linear combination of light from multiple scene locations, and each row in Φ encodes the weights for the respective sensor. For a scene at infinity, the weights for two different sensor pixels simply differ by a translation of the mask pattern. As a consequence, the matrix Φ has a Toeplitz structure.

2) Every light source in the scene casts a shadow of the mask on the sensor. Thus, the image formed on the sensor is a superposition of shifted and scaled versions of the mask. The shift and the scaling of the mask pattern encodes the angle and distance of the light source onto the sensor.

These properties are invaluable in the design of masks that provide near-optimal recovery under noise. Given the image formation model in 1), our tasks are to formulate an inversion algorithm that recovers the scene x from the sensed image y and design mask patterns that achieve optimal recovery performance. We study both problems in the subsequent sections.

Image reconstruction

Given the sensor measurements $\mathbf{y} \in \mathbb{R}^M$ and the measurement matrix Φ, recovering $\mathbf{x} \in \mathbb{R}^N$ depends mainly on the rank of the matrix Φ and its condition number. When $\text{rank}(\Phi) = N$ and the matrix is well-conditioned, we can obtain an estimate of $\mathbf{x}$ by solving the least-squares problem

$$\min_{\mathbf{x}} \| \Phi\mathbf{x} - \mathbf{y} \|_2^2, \tag{2}$$

which has the closed-form solution $\hat{\mathbf{x}}_{\text{LS}} = \Phi^+ \mathbf{y} = \mathbf{x} + \Phi^+ \mathbf{e}$, where $\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T$ is the pseudoinverse of Φ. When Φ is not well-conditioned, the least squares estimate $\hat{\mathbf{x}}_{\text{LS}}$ suffers from noise amplification. When Φ is rank-deficient, the matrix becomes singular and an estimate cannot be achieved.

In the ill-conditioned and rank-deficient cases, we can use an image prior to regularize the inverse problem. Specifically, instead of solving (2), we solve

$$\min_{\mathbf{x}} \| \mathbf{y} - \Phi\mathbf{x} \|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \tag{3}$$

where $\| \mathbf{y} - \Phi x \|^2$ quantifies the data fidelity, $\mathcal{R}(\mathbf{x})$ is a regularization term that enforces an image prior, and $\lambda > 0$ controls the tradeoff between fidelity and regularization. A popular choice for the regularizer that is useful for noise-suppression is Tikhonov regularization (also known as *ridge regression*) via $\mathcal{R}(\mathbf{x}) = \| \mathbf{x} \|_2^2$.

Natural signals, such as images and videos, exhibit a host of geometric properties including sparse gradients and sparse coefficients in certain transform domains (e.g., Fourier or wavelets). By enforcing these geometric properties, we can suppress noise amplification as well as obtain unique solutions even when Φ is rank-deficient (i.e., $M < N$). A pertinent example for image reconstruction is the total-variation (TV) model, where the regularizer $\mathcal{R}(\mathbf{x}) = \| \mathbf{x} \|_{TV}$ corresponds to the TV of the image, which is computed from its gradients. Writing the scene $\mathbf{x}$ as the 2-D image $\mathbf{x}(u, v)$ and defining $\mathbf{g}_u = D_u \mathbf{x}$ and $\mathbf{g}_v = D_v \mathbf{x}$ as the u- and v-components, respectively, of the spatial gradient of the image, the TV of the image is given by

$$\mathcal{R}(\mathbf{x}) = \| \mathbf{x} \|_{\text{TV}} = \sum_{u,v} \sqrt{\mathbf{g}_u(u,v)^2 + \mathbf{g}_v(u,v)^2}.$$

The minimization (3) with a TV prior is convex and produces images with sparse gradients. A host of efficient techniques have been developed to obtain the solution. A range of even more realistic image models have been developed (e.g., [30]), but the resulting optimization might not be convex.

FlatCam: A lensless imaging case study

To illustrate the design tradeoffs involved in a practical lensless camera design, we review the FlatCam [31], which was inspired by the coded aperture imaging principles pioneered in astronomical X-ray and gamma-ray imaging [2]–[4], [32] (see "Coded Aperture in X-ray and Gamma-Ray Imaging").

Architectural overview

The FlatCam design achieves a large photosensitive area with a thin form factor by replacing the lens with a coded, binary mask. The thickness of the camera is minimized by placing the mask almost immediately on top of a bare conventional sensor array. The image formed on the sensor can be viewed as a superposition of many pinhole images. An illustration of the FlatCam design is presented in Figure 3. Light from all points in the scene passes through a coded mask and forms a multiplexed image on the sensor. A computational algorithm is used to recover the original light distribution of the scene from the sensor measurements.

The FlatCam design has many attractive properties besides its slim profile. First, since it reduces the thickness of the camera but not the area of the sensor, it can collect more light than a miniature, lens-based camera of the same thickness. The light
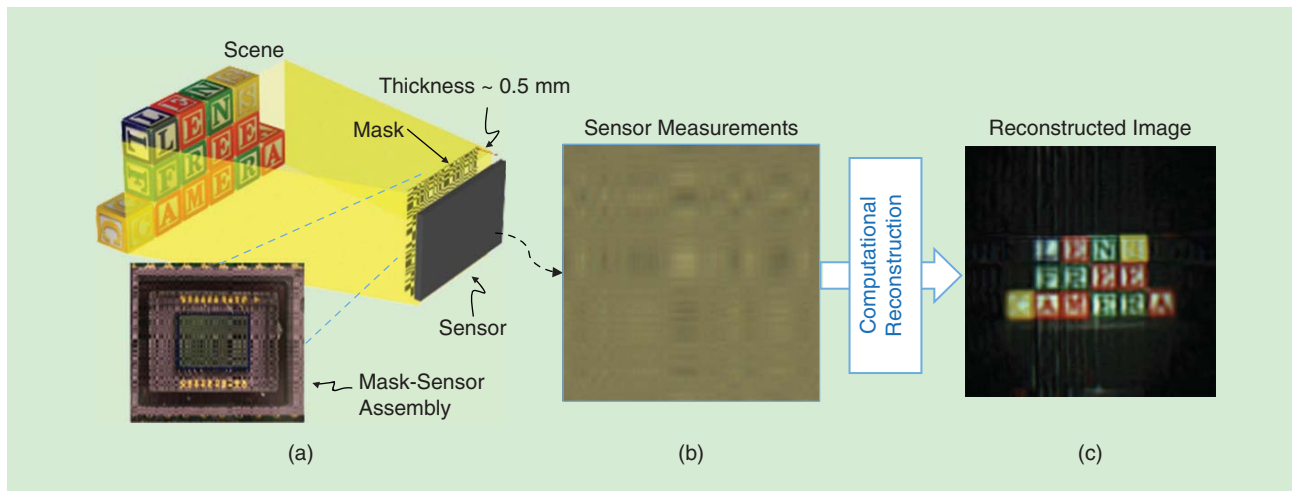
FIGURE 3. The FlatCam architecture. Every light source within the camera field of view contributes to every pixel in the multiplexed image formed on the sensor. A computational algorithm reconstructs the image of the scene. (a) The inset shows the mask-sensor assembly of our prototype, in which a binary, coded mask is placed 1.2-mm away from an off-the-shelf digital image sensor. (b) An example of sensor measurements. (c) An image reconstructed by solving a computational inverse problem of the form (3). (Figure modified from [31] and used with permission.)

collection ability of FlatCam is proportional to the size of the sensor and the transparent regions (pinholes) in the mask. In contrast, the light collection ability of a lens-based camera is limited by the lens aperture size, which is restricted by the requirements on the device thickness. Second, the mask can be created from inexpensive materials that operate over a broad range of wavelengths. Third, the mask can be fabricated simultaneously with the sensor array, creating new manufacturing efficiencies.

Mask design and calibration

Separable masks

The FlatCam uses a separable mask pattern, i.e., the 2-D mask pattern is the outer product of two 1-D patterns. Such a pattern drastically reduces the storage and computational footprint of the measurement matrix Φ. When the mask pattern is separable, the imaging equation (1) can be rewritten as

$$Y = \Phi_L X \Phi_R^T + E, \tag{4}$$

where X is an $N \times N$ matrix containing the scene radiance; Y in an $M \times M$ matrix containing the sensor measurements; Φ_L and Φ_R are matrices representing 1D convolution along the rows and columns of the scene, respectively; and E denotes the sensor noise and model mismatch. For a megapixel scene/image and a megapixel sensor, Φ_L and Φ_R each have only 10^6 elements, as opposed to 10^{12} elements in Φ. A similar idea has been recently proposed in [12] with the design of doubly Toeplitz mask.

Mask design

The mask pattern should be chosen to make the matrices Φ_L and Φ_R as numerically stable as possible, which ensures a stable recovery of the image X from the sensor measurements Y. In the context of image reconstruction using signal priors (for example,

the aforementioned TV prior), random matrices enjoy stable recovery guarantees. Hence, we construct the separable mask pattern as the outer-product of two 1-D pseudo-random sequences.

Calibration

The low-dimensionality of Φ_L and Φ_R in (4) support a simple and efficient calibration scheme. Instead of modeling the convolution shifts and diffraction effects for a particular mask-sensor arrangement, we directly estimate the system matrices from training data. To align the mask and sensor, we adjust their relative orientation such that a separable scene in front of the camera yields a separable image on the sensor. For a perfectly aligned system, displaying a horizontal/vertical line on a screen in front of the camera results in an image containing a set of sharp horizontal/vertical stripes. We first achieve sharpness by rotating the mask relative to the screen. Then, we align the sensor and mask so that the stripes on the sensor image are parallel to the image axis. To calibrate a system that can recover an image X with dimensions $N \times N$, we estimate the left and right matrices Φ_L and Φ_R using the sensor measurements of $2N$ known calibration patterns projected on a screen as depicted in Figure 4. Our calibration procedure relies on an important observation: If the scene X is separable, i.e., $X = \mathbf{ab}^T$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$, then, for an ideal system,

$$Y = \Phi_L \mathbf{ab}^T \Phi_R^T = (\Phi_L \mathbf{a})(\Phi_R \mathbf{b})^T.$$

In essence, the image formed on the sensor is a rank-1 matrix, and using a truncated singular value decomposition (SVD), we can obtain estimates of $\Phi_L \mathbf{a}$ and $\Phi_R \mathbf{b}$ up to a signed, scalar constant. We take N separable pattern measurements for calibrating each of Φ_L and Φ_R. In practice, we average several measurements of each calibration pattern to reduce the effects of sensor noise.
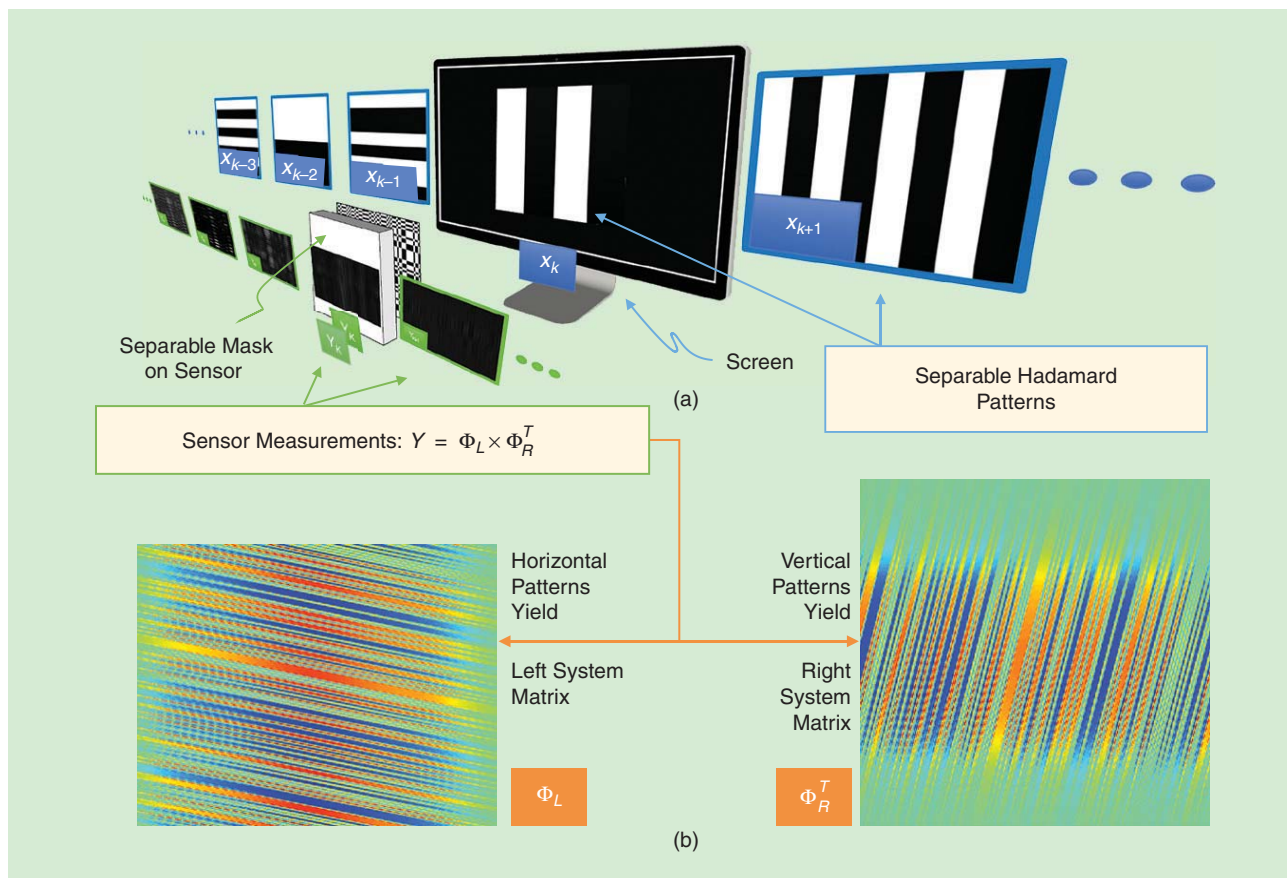
FIGURE 4. Calibration for measuring the left and right matrices Φ_L and Φ_R corresponding to a separable mask. (a) Separable patterns displayed on a screen in front of the camera. The patterns are orthogonal, one-dimensional Hadamard codes that are repeated along either the horizontal or vertical direction. (b) Estimated left and right matrices. (Figure modified from [31] and used with permission.)

Prototypes

We have built two different FlatCam prototypes. The first prototype consists of a Point Grey Flea3 with a Sapphire EV76C560 CMOS sensor, which has a 5.3-μm pixel size and measured Chief Ray Angle (CRA) of 25°. (The CRA of a sensor determines the cone of light that can enter a pixel.) The diffractive mask is chrome on quartz glass placed adjacent to the infrared filter of the sensor (mask-to-sensor distance: 1.2 mm). The pattern on the mask is an outer product of two length-1024 pseudorandom sequences of smallest feature size 25 μm. Sample reconstructions using this prototype are shown in Figure 5(a). Reconstructions of a dynamic scene are shown in Figure 5(b); here, we operated the camera with a 3 ms exposure and recovered videos of 60 frames per second with each frame of the video recovered independently as a stand-alone image.

The second prototype was assembled with a diffractive mask and spacer attached directly to the surface of an Omnivision OV5647 CMOS sensor (fourth example of lensless in Figure 1). The Omnivision sensor has pixels of size 1.4 μm and measured CRA of 28°. The diffractive mask was fabricated by depositing a thin-film of chrome on fused silica that was then patterned with photoresist and etched to leave the desired pattern. The mask was then diced, aligned to the CMOS pixel array, and attached with optical epoxy (mask to sensor distance 500 μm). The pattern on the mask is the outer product of length-1296 and 972 pseudorandom sequences of smallest feature size of 2.8 μm. The smaller feature size and pixel pitch of this prototype enable reconstructions at a higher resolution. However, a drawback of the smaller pixel pitch is a sensor with poorer SNR performance, which results in noisier measurements and reconstruction compared to our first prototype (see Figure 6).

The remainder of the examples below were obtained using the higher-quality Flea3 sensor prototype.

Programmable masks

In many applications, a camera has the opportunity to acquire several images of a scene, and both folk wisdom and theory tell us that averaging the acquisitions should suppress noise. In contrast to a lens-based camera, a lensless camera could be equipped with a programmable mask that changes for each acquisition to provide diversity in the acquisitions. Presumably, such a camera should not only suppress noise but also average out imperfections in the measurement operator Φ (recall that Φ is never perfectly conditioned in a coded aperture system).

To demonstrate the potential of a programmable mask FlatCam, we simulated multi-image capture using the sequences of separable measurement matrices $\{\Phi_{L_i}\}_{i=1,...,L}$ and
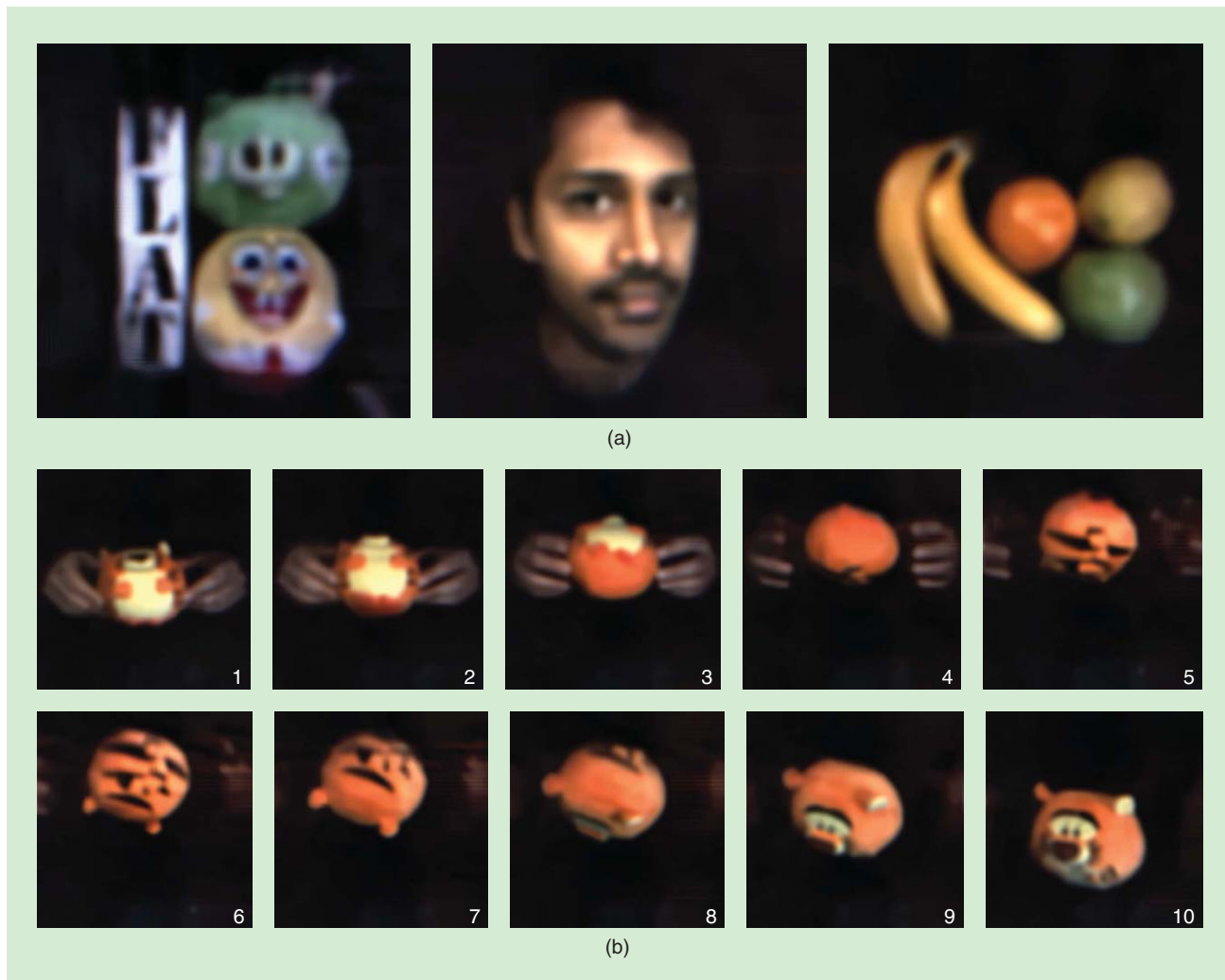
FIGURE 5. (a) The reconstruction of three static scenes using the Flea FlatCam prototype. (b) Sample frames from the video reconstruction of a toy performing a backflip aided by human hands. The video was recorded at 60 frames per second.

$\{\Phi_{R_j}\}_{j=1,...,R}$. The measurements of the scene X using each pair of measurement matrices are given by

$$Y_{(i,j)} = \Phi_{L_i} X \Phi_{R_j}^T + E_{(i,j)}, \tag{5}$$

which can be stacked into the larger system

$$Y = \begin{bmatrix} Y_{(1,1)} & . & . & . & . \\ . & . & . & . & . \\ . & . & \{Y_{(i,j)}\} & . & . \\ . & . & . & . & . \\ . & . & . & . & Y_{(L,R)} \end{bmatrix}, \Phi_L = \begin{bmatrix} \Phi_{L_1} \\ . \\ \{\Phi_{L_i}\} \\ . \\ \Phi_{L_L} \end{bmatrix}, \Phi_R = \begin{bmatrix} \Phi_{R_1} \\ . \\ \{\Phi_{R_j}\} \\ . \\ \Phi_{R_R} \end{bmatrix} \tag{6}$$

to estimate X by solving the least-squares problem

$$\hat{X}_{LS} = \underset{X}{\arg\min} ||\Phi_L X \Phi_R^T - Y||_F^2, \tag{7}$$

where $||\cdot||_F$ denotes the Frobenius norm. A regularization term can also be added as in the single-image-capture case.

Figure 7(a) illustrates the results of a simulation of this approach. We generated a virtual high-resolution scene X, acquired a number of noisy acquisitions according to the model (5), and recovered the image estimate according to (7) with an additional Tikhonov regularizer. The horizontal axis corresponds to the number of acquisitions, each of which used a new mask. The vertical axis corresponds to peak-signal-to-noise ratio (PSNR), which measures the mean-squared error between the image estimate and X. Each acquisition had the same, fixed exposure time, and so we expect the quality of the image estimate to improve as we fuse more acquisitions. The blue curve demonstrates not only this improvement, but also an additional improvement due to changing the mask for each acquisition rather than reusing the same mask repeatedly. In particular, taking nine acquisitions using nine different masks attains a PNSR of 13.7 dB, while taking nine acquisitions with the same mask attains a PSNR of only 11.8 dB [the green dot in Figure 7(a)] This 2-dB gain is testimony to the power of programmable masks.

The careful reader will note that in practice the scene might change during acquisition, which would invalidate the
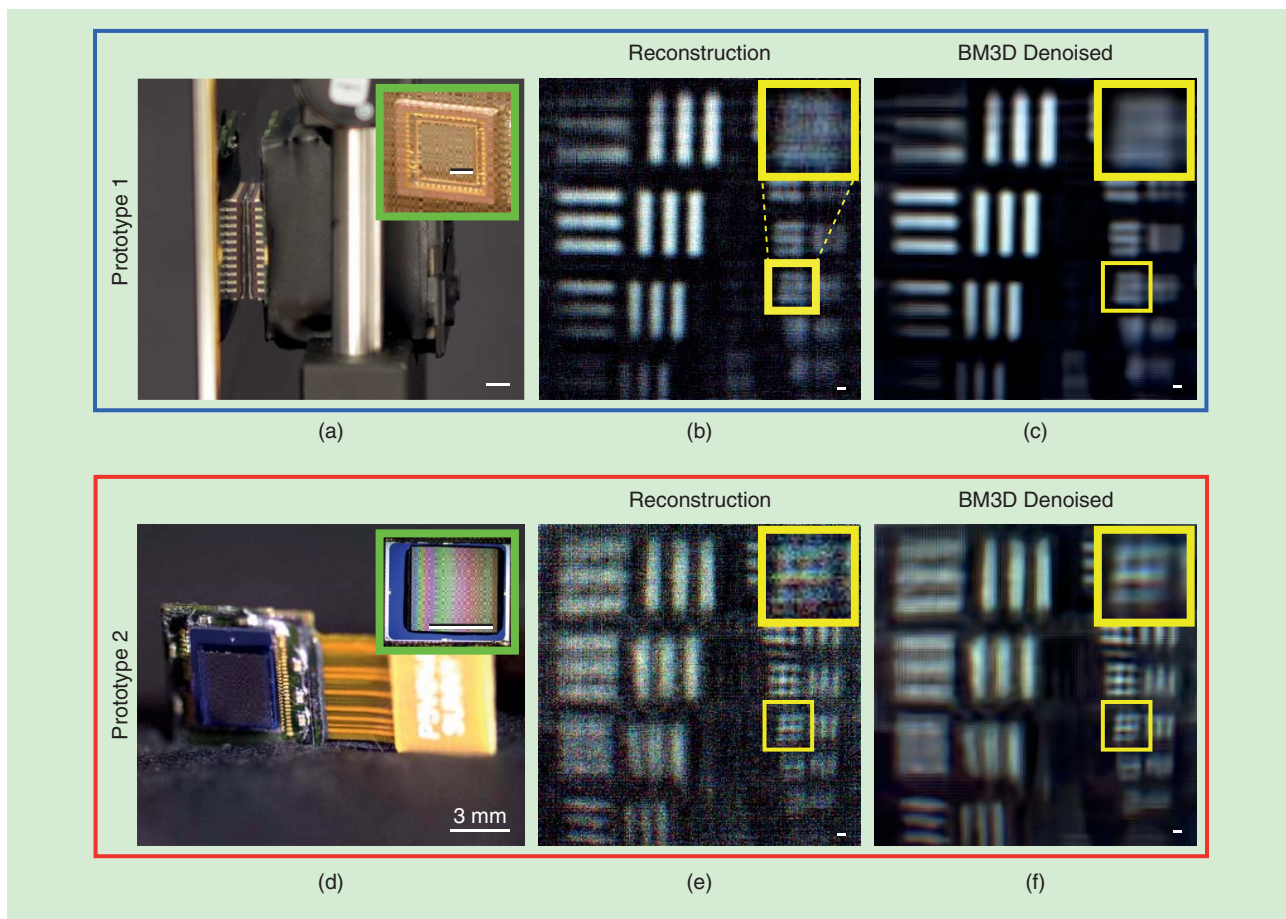
FIGURE 6. A FlatCam prototype comparison. (a) Our first prototype with the chrome mask placed directly in front of the Flea3 sensor; (d) is our second prototype with the Omnivision sensor directly epoxied to mask (insets show close-up of the sensors and masks). (b) and (e) are reconstructions and (c) and (f) are BM3D denoised reconstructions for first and second prototype, respectively. The smaller feature size and pixel pitch of the Omnivision prototype provide superior resolution at the cost of more noisier image reconstruction.
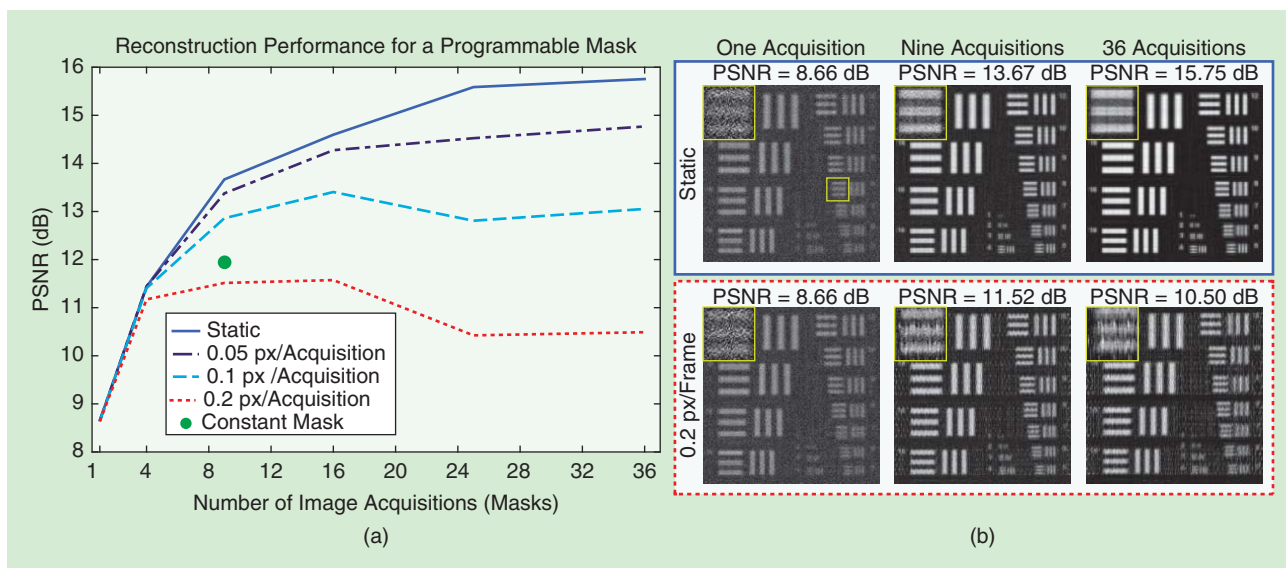
FIGURE 7. A simulation experiment of a FlatCam with a programmable mask. (a) Reconstruction performance (in terms of peak-signal-to-noise ratio, PSNR) as we increase the number of image acquisitions (masks). A different mask pattern is used for each acqusition. The PSNR increases consistently for static and slow-moving scenes, but after peaking early, deteriorates for faster moving scenes due to model mismatch. The green dot indicates the performance when all nine image acquisitions of a static scene are done using the same (constant) mask. Though the performance is better than a single acquisition, it is still outshined by the programmable mask. (b) Reconstructed images using one, nine, and 36 acquisitions for the static scene and fast-moving scene.
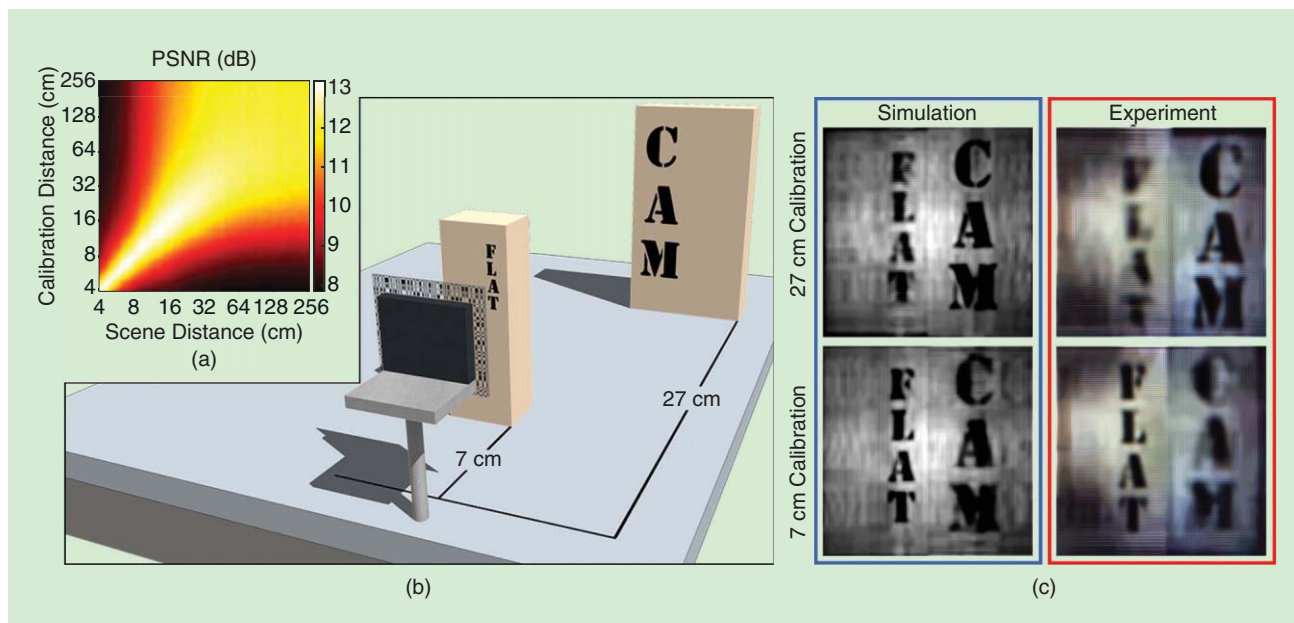
FIGURE 8. The 3-D imaging with FlatCam. (a) A heat map of the reconstruction PSNR of a simulation of the scene as a function of the scene distance and the calibration distance of the measurement matrices. At closer scene distances, the reconstruction is sensitive to the choice of multiplexing matrix at the correct calibration depth; at further scene distances, the sensitivity decreases. (b) An experimental setup showing "FLAT" and "CAM" at different distances from the camera. (c) Reconstruction at 7 cm and 27 cm through simulation and our prototype FlatCam. The word "FLAT" placed at 7 cm is in focus when reconstructed using measurement matrices calibrated to depth of 7cm. The word "CAM" placed at 27 cm is in focus when reconstructed using measurement matrices calibrated to depth of 27 cm.

model (5) and commensurately reduce performance. To investigate the effect of scene motion, we repeated the above experiment with three different dynamic scenes that diagonally translate the virtual high-resolution scene X by 0.05, 0.1, and 0.2 pixels per acquisition. For the slow-moving scene, the reconstruction performance improves with the number of acquisitions, just as with a static scene. But for the faster-moving scenes, after peaking early on, the reconstruction performance deteriorates with the number of acquisitions, due to the increased deviation from the model. Figure 7(c) shows the reconstructed images using one, nine, and 36 measurements for the static scene and the scene moving at 0.2 pixels per acquisition.

The trade-off between spatial and temporal resolution could be improved by estimating the motion between frames and registering the measurements before reconstructing the image (a difficult, but solvable problem; see [33]). Adaptive measurement schemes also hold promise for balancing this tradeoff.

3-D imaging
FlatCam can computationally change its focus to new depths in a scene from a single acquisition. The key is that, for a given mask design, we can calibrate a set of separable measurement matrices $\{\Phi_{L_i}\}_{i=1,..,L}$ and $\{\Phi_{R_j}\}_{j=1,..,R}$, each obtained using a screen at a different depth (recall Figure 4).

> The lensless imaging approach promises to challenge the traditional barriers of size, weight, cost, and performance in a broad range of applications spanning consumer, medical, scientific imaging, machine vision, and remote sensing.

Figure 8(a) shows a heat map of reconstruction PSNR of a simulated 2-D scene as a function of the scene distance and the calibration distance of the measurement matrices. We see that the reconstruction quality improves as the calibration depth of the camera approaches the actual scene depth. Moreover, the sensitivity of the reconstruction due to the discrepancy in these depths decreases with increasing scene distance.

Figure 8(c), shows the reconstruction of a 3-D scene at two different depth planes. For a particular fixed mask, we calibrated the measurement matrices with a screen at the distances of 7 cm and 27 cm. We accounted for the field of view of the sensor by adjusting the size of the calibration patterns in accordance with the CRA of the sensor. The resulting two sets of matrices were then used to create focused images at 7 cm and 27 cm from a single acquisition with FlatCam. (The line artifacts in the experimental reconstruction are due to scene illumination leaking into the sensor from the sides that was not accounted for in the calibration procedure. We can reduce the unaccounted light in future prototypes by introducing baffles.)

Limitations and challenges facing lensless imaging
The very first cameras were lensless (pinhole cameras), but the advent of lenses and other advanced optics relegated such

systems to niche applications like X-ray and gamma-ray imaging. The resurgence of lensless imaging can be attributed to the convergence of four factors: the development of digital CMOS and CCD sensor arrays, efficient and realistic image models and recovery algorithms, powerful computing, and new mask designs (such as the separable mask in the FlatCam).

The further development of lensless imaging, however, will face challenges. As the mask is moved closer to the sensor in any pinhole or coded aperture camera, the angular resolution decreases, resulting in a trade-off between minimal thickness and spatial resolution [29]. Additionally, computationally recovering a scene from less-than-perfectly conditioned sensor measurements results in noise amplification. Although noise amplification cannot be eliminated, careful design of mask patterns and regularization models can minimize this effect. The necessity for a computational algorithm also results in a time-lag between image acquisition and reconstruction (~100 ms for FlatCam). Such a delay may be acceptable in certain applications but unacceptable in others such as augmented or virtual reality. There are a number of avenues for continued research and development that could lead to significantly improved lensless imaging performance, including new architectures for improving spatial resolution, new image models to reduce the demultiplexing noise, and new computational algorithms to support high-speed sensing. Sometimes, size matters. The lensless imaging approach promises to challenge the traditional barriers of size, weight, cost, and performance in a broad range of applications spanning consumer, medical, scientific imaging, machine vision, and remote sensing. Indeed, the future of lensless imaging research and development looks very bright.

Acknowledgments

Authors

Vivek Boominathan (vivekb@rice.edu) is a Ph.D. candidate in the Electrical and Computer Engineering Department at Rice University, Houston, Texas. His research interests are in the areas of computer vision, signal processing, and computational imaging. He received his B.Tech. degree in electrical engineering from the Indian Institute of Technology, Hyderabad, in 2012. He is a three-time recipient of Todai (Japan)-IIT Undergraduate Student Scholarships for academic excellence in 2009, 2010, and 2011.

Jesse K. Adams (jka4@rice.edu) is a Ph.D. candidate in the Applied Physics/Electrical and Computer Engineering Department at Rice University, Houston, Texas. His research interests lie in the areas of nanofabrication, nano-optics, neuroengineering, and computational imaging. He received his B.Sc. degree in physics from the University of North Florida in 2014.

M. Salman Asif (sasif@ece.ucr.edu) is an assistant professor in the Department of Electrical and Computer Engineering at the University of California, Riverside. He received his B.Sc. degree in 2004 from the University of Engineering and Technology, Lahore, Pakistan, and his M.S.E.E. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, in 2008 and 2013, respectively. He worked as a research intern at Mitsubishi Electric Research Laboratories in Cambridge, Massachusetts, in the summer of 2009 and at Samsung Standards Research Laboratory in Richardson, Texas, in the summer of 2010. He worked as a senior research engineer at Samsung Research America, Dallas, Texas, from August 2012 to January 2014 and as a postdoctoral researcher at Rice University from February 2014 to June 2016. His research interests include compressive sensing, computational and medical imaging, and machine learning.

Benjamin W. Avants (bwa1@rice.edu) is currently a research engineer in the Robinson Lab at Rice University, Texas, who specializes in the design, fabrication, development, programming, and automation of experimental tools and apparatus, including micro- and nano-fabrication. His role also includes system modeling and the development of data processing tools and techniques. He received his bachelor's degree in electrical engineering and computer engineering from the University of Memphis in 2012 with second majors in physics and math. He began working with the Robinson Lab that same year. His personal interests also include alternative power generation and storage methods as well as robotics.

Jacob T. Robinson (jtrobinson@rice.edu) is an assistant professor in the Department of Electrical and Computer Engineering with joint appointments in the Bioengineering Department at Rice University and the Neuroscience Department at Baylor College of Medicine. He earned his B.S. degree in physics from the University of California, Los Angeles, in 2003, followed by his M.S. and Ph.D. degrees in applied physics from Cornell University in 2008. His research is focused on developing scalable nanotechnologies to manipulate and measure brain activity. From 2008 to 2012, he was a postdoctoral researcher in the Department of Chemistry and Chemical Biology at Harvard University, after which he joined Rice University as an assistant professor. He is a recipient of the John S. Dunn Foundation Collaborative Research Award, the Hamill Innovations Award, and the DARPA Young Faculty Award. His areas of interest include nanowire electrodes, nanophotonic probes, and nanoscale magnetic materials.

Richard G. Baraniuk (richb@rice.edu) is the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University and the founder and director of OpenStax (openstax.org). His research interests lie in new theory, algorithms, and hardware for sensing, signal processing, and machine learning. He is a Fellow of the IEEE and the American Association for the Advancement of Science and has received national young investigator awards from the National Science Foundation and Office of Naval Research; the Rosenbaum Fellowship from the Isaac Newton Institute of Cambridge University; the ECE Young Alumni Achievement Award from the University of Illinois; the IEEE Signal Processing Society Best Paper, Best Column, Education, and Technical Achievement Awards; and the IEEE James H. Mulligan, Jr. Medal.

Aswin C. Sankaranarayanan (saswin@andrew.cmu.edu) is an assistant professor in the Electrical and Computer Engineering (ECE) Department at Carnegie Mellon University. He received his Ph.D. degree from the University of Maryland, College Park, where he was awarded the distinguished dissertation fellowship by the ECE Department in 2009. He was a postdoctoral researcher in the digital signal processing group at Rice University. His research encompasses problems in compressive sensing and computational imaging. He has received best paper awards at the Computer Vision and Pattern Recognition Workshops on Computational Cameras and Displays (2015) and Analysis and Modeling of Faces and Gestures (2010).

Ashok Veeraraghavan (vashok@rice.edu) is an assistant professor of electrical and computer engineering at Rice University, Texas, where he directs the Computational Imaging and Vision Lab. He received his bachelor's degree in electrical engineering from the Indian Institute of Technology, Madras, in 2002 and his M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering at the University of Maryland, College Park, in 2004 and 2008, respectively. His research interests are broadly in the areas of computational imaging, computer vision, and robotics. Prior to joining Rice University, he spent three years as a research scientist at Mitsubishi Electric Research Labs in Cambridge, Massachusetts. His thesis received the doctoral dissertation award from the Department of Electrical and Computer Engineering at the University of Maryland.

References

[1] V. Dragoi, A. Filbert, and S. Zhu, and G. Mittendorfer, "CMOS wafer bonding for back-side illuminated image sensors fabrication," in *Proc. 11th Int. Conf. Electronic Packaging Technology & High Density Packaging*, 2010, pp. 27–30.

[2] R. Dicke, "Scatter-hole cameras for X-rays and gamma rays," *Astrophys. J.*, vol. 153, pp. L101, 1968.

[3] E. Fenimore and T. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, 1978.

[4] T. Cannon and E. Fenimore, "Coded aperture imaging: Many holes make light work," *Optic. Eng.*, vol. 19, no. 3, pp. 193–283, 1980.

[5] A. Busboom, H. Elders-Boll, and H. Schotten, "Uniformly redundant arrays," *Exp. Astronomy*, vol. 8, no. 2, pp. 97–123, 1998.

[6] J. W. Goodman, *Introduction to Fourier Optics*. Englewood, CO: Roberts and Company, 2005.

[7] H. H. Barrett, "Fresnel zone plate imaging in nuclear medicine," *J. Nuc. Med.*, vol. 13, no. 6, pp. 382–385, 1972.

[8] J. Kirz, "Phase zone plates for X-rays and the extreme UV," *J. Opt. Soc. Am.*, vol. 64, no. 3, pp. 301–309, 1974.

[9] Y. S. Chu, J. M. Yi, F. De Carlo, Q. Shen, Wah-Keat Lee, H. J. Wu, C. L. Wang, J. Y. Wang, C. J. Liu, C. H. Wang, S. R. Wu, C. C. Chien, Y. Hwu, A. Tkachuk, W. Yun, M. Feser, K. S. Liang, C. S. Yang, J. H. Je, and G. Margaritondo "Hard-X-ray microscopy with Fresnel zone plates reaches 40 nm rayleigh resolution," *Appl. Phys. Lett.*, vol. 92, no. 10, pp. 103119, 2008.

[10] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2006, vol. 1, pp. 339–346.

[11] G. Huang, H. Jiang, and K. Matthews, and P. Wilford, "Lensless imaging by compressive sensing," in *Proc. 20th IEEE Int. Conf. Image Processing*, 2013, pp. 2101–2105.

[12] M. J. DeWeert and B. P. Farm, "Lensless coded-aperture imaging with separable doubly-toeplitz masks," *Opt. Eng.*, vol. 54, no. 2, pp. 023102, 2015.

[13] H. Jiang, G. Huang, and P. Wilford, "Multi-view in lensless compressive imaging," in *Proc. Picture Coding Symp.*, 2013, pp. 41–44.

[14] A. Wang, P. Gill, and A. Molnar, "Angle sensitive pixels in CMOS for lensless 3D imaging," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2009, pp. 371–374.

[15] P. R. Gill, C. Lee, D. G. Lee, A. Wang, and A. Molnar, "A microscale camera using direct Fourier-domain scene capture," *Opt. Lett.*, vol. 36, no. 15, pp. 2949–2951, 2011.

[16] P. R. Gill and D. G. Stork, "Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings," in *Computational Optical Sensing and Imaging*. Washington DC: Optical Society of America, 2013, pp. CW4C–3.

[17] D. Stork and P. Gill, "Lensless ultra-miniature CMOS computational imagers and sensors," in *Proc. Int. Conf. Sensor Technologies and Applications*, 2013, pp. 186–190.

[18] H. F. Tablot Esq. F.R.S, "LXXVI. facts relating to optical science. no. IV," *Phil. Mag. Series 3*, vol. 9, no. 56, pp. 401–407, 1836.

[19] X. Cui, L. M. Lee, X. Heng, W. Zhong, P. W. Sternberg, D. Psaltis, and C. Yang. (2008). "Lensless high-resolution on-chip optofluidic microscopes for caenorhabditis elegans and cell imaging, presented at National Academy of Sciences. [Online]. Available: http://www.pnas.org/content/early/2008/07/25/0804612105.abstract

[20] S. A. Lee, R. Leitao, G. Zheng, S. Yang, A. Rodriguez, and C. Yang, "Color capable sub-pixel resolving optofluidic microscope and its application to blood cell imaging for malaria diagnosis," *PLoS One*, vol. 6, no. 10, pp. 1–6, Oct. 2011.

[21] G. Zheng, S. A. Lee, Y. Antebi, M. B. Elowitz, and C. Yang, "The epetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (spsm)," *Proc. Natl. Acad. Sci.*, vol. 108, no. 41, pp. 16 889–16 894, 2011.

[22] J. Spence, U. Weierstall, and M. Howells, "Phase recovery and lensless imaging by iterative methods in optical, X-ray and electron diffraction," *Phil. Trans. R. Soc. London A*, vol. 360, no. 1794, pp. 875–895, 2002.

[23] H. Faulkner and J. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," *Phys. Rev. Lett.*, vol. 93, no. 2, pp. 023903, 2004.

[24] A. Greenbaum, W. Luo, T. W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: Achievements and remaining challenges of wide-field on-chip microscopy," *Nat. Methods*, vol. 9, no. 9, pp. 889–895, 2012.

[25] A. Greenbaum, Y. Zhang, A. Feizi, P. L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," *Sci. Transl. Med.*, vol. 6, no. 267, pp. 267ra175–267ra175, 2014.

[26] J. Rodenburg, A. Hurst, A. Cullis, B. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson, "Hard-X-ray lensless imaging of extended objects," *Phys. Rev. Lett.*, vol. 98, no. 3, pp. 034801, 2007.

[27] M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer, "Ptychographic X-ray computed tomography at the nanoscale," *Nature*, vol. 467, no. 7314, pp. 436–439, 2010.

[28] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[29] J. Miao, "Coherent diffraction imaging," *Microsc. Microanal.*, vol. 20, no. S3, pp. 368–369, 2014.

[30] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, 1999.

[31] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk. (2015). Flatcam: Thin, bare-sensor cameras using coded aperture and computation. [Online]. Available: arXiv preprint arXiv:1509.00116

[32] P. Durrant, M. Dallimore, I. Jupp, and D. Ramsden, "The application of pinhole and coded aperture imaging in the nuclear environment," *Nucl. Inst. Methods Phys. Res. Sec. A*, vol. 422, no. 1, pp. 667–671, 1999.

[33] A. C. Sankaranarayanan, X. Lu, C. Studer, Y. Li, K. F. Kelly, and R. G. Baraniuk. (2015). Video compressive sensing for spatial multiplexing cameras using motion-flow models. *CoRR*, vol. abs/1503.02727. [Online]. Available: http://arxiv.org/abs/1503.02727

[34] E. Caroli, J. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino, Coded aperture imaging in X-and gamma-ray astronomy," *Space Sci. Rev.*, vol. 45, no. 3–4, pp. 349–403, 1987.

[35] D. J. Brady, *Optical Imaging and Spectroscopy*. New York: Wiley, 2009.

[36] C. Markwardt, S. Barthelmy, J. Cummings, D. Hullinger, H. Krimm, and A. Parsons, "The swift bat software guide," NASA/GSFC, Greenbelt, MD, vol. 6, 2007.

SP

**SIGNAL PROCESSING FOR
COMPUTATIONAL PHOTOGRAPHY AND DISPLAYS**

Pradeep Sen and Cecilia Aguerrebere

Practical High Dynamic Range Imaging of Everyday Scenes

Photographing the world as we see it with our own eyes
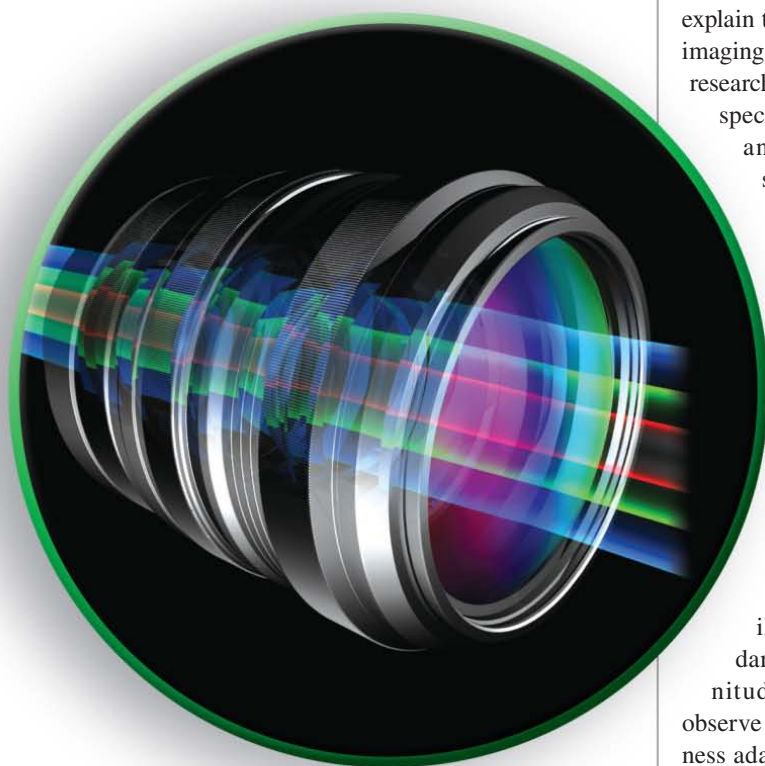
©ISTOCKPHOTO.COM/YAKOBCHUK

High dynamic range (HDR) imaging enables the capture of an extremely wide range of the illumination present in a scene and so produces images that more closely resemble what we see with our own eyes. In this article, we explain the problem of limited dynamic range in the standard imaging pipeline and then present a survey of state-of-the-art research in HDR imaging, including the technology's history, specialized cameras that capture HDR images directly, and algorithms for capturing HDR images using sequential stacks of differently exposed images. Because this last is among the most common methods for capturing HDR images using conventional digital cameras, we also discuss algorithms to address artifacts that occur when using with this method for dynamic scenes. Finally, we consider systems for the capture of HDR video and conclude by reviewing open problems and challenges in HDR imaging.

Overview of HDR imaging

The world around us is visually rich and complex. Some of this richness comes from the wide range of illumination present in daily scenes—the illumination intensity between the brightest and the darkest parts of a scene can vary by many orders magnitude. Fortunately, the human visual system can observe very wide ranges of luminosity by means of brightness adaptation, which allows us, for example, to easily see the bright scene outside a window as well as the darkened interior. A digital camera, on the other hand, has a sensor that responds linearly to illumination; coupled with the sensor pixels' limited capacity to store energy and the noise present in the acquisition process, this fundamentally limits the sensor's measurable dynamic range. The low dynamic range (LDR) of modern digital cameras is a major factor preventing them from capturing images as humans see them (Figure 1). For this reason, an entire research community, both in academia and industry, is engaged in

(a) Images Taken at Two Different Exposures　　　(b) Full Stack of Images and HDR Result

FIGURE 1. (a) Images captured by standard digital cameras cannot reproduce the wide range of illumination we see in everyday scenes, even after adjusting the exposure, as illustrated by these two images taken at different exposures. (b) HDR imaging allows for the capture of a wider range of illumination; here, a stack of images was captured at different exposures (left) and merged with the algorithm described in [1] to reduce motion artifacts and produce the result shown on the right.

developing HDR imaging algorithms and systems to allow better photographs to be captured.

In this article, we describe research within the computational photography community on HDR imaging that enables the capture of a wider range of illumination than is normally captured and produces images closer to what we see with our own eyes. In a way, HDR imaging represents the epitome of computational photography: many of the solutions involved require novel optics, new acquisition processes, and clever algorithms in the back end to produce better images. As such, this article will focus only on the acquisition of HDR images and will not discuss related topics that have been extensively studied such as HDR image representation (how to compress and store HDR images) or tone mapping (turning an HDR image into an LDR image suitable for standard display) [2]. Further, because of this tutorial's strict space limitations, we cannot cover in depth the large body of existing work on HDR imaging and refer interested readers instead to textbooks and papers that survey the subject [1]–[6].

Historical background

As early as the mid-1800s—soon after the invention of photography itself—early photography pioneers were already struggling with the limited dynamic range of film and began to develop techniques that provided the basis of what we now know as HDR imaging. The French photographer Hippolyte Bayard was the first to propose that two negatives, each one properly exposed for different content, could be combined to create a well-balanced photograph. His compatriot Gustave Le Gray captured many beautiful seascape photographs with his *ciel rapporté* technique, where one negative was used for the dark sea and the other for the bright sky. Others, such as Oscar Rejlander, combined many well-exposed negatives to produce photographs that emulated contemporary paintings in which everything was properly "exposed" (Figure 2).

This idea of combining images acquired with different exposures to produce an HDR result was reintroduced for

digital photography in the 1990s (almost 150 years later) by Madden [7] and Mann and Picard [8]. However, HDR imaging received relatively little attention until the seminal paper by Debevec and Malik [9] placed it at the forefront of the burgeoning computational photography community. Since then, there has been almost 20 years of research on HDR imaging. Before we delve into this research, however, we must first review the standard imaging pipeline and understand the reasons for its limited dynamic range. In addition, we need to formalize colloquial terms such as *brightness* by introducing the appropriate radiometric units that characterize light.

The standard imaging pipeline and its limited dynamic range

The standard imaging pipeline (Figure 3) starts with a set of rays leaving the scene in the direction of the camera, with each ray carrying some amount of radiant power called *radiance* (L; units: $W/m^2 sr$). The rays entering the lens aperture and striking the sensor at a point are integrated over the solid angle subtended by the aperture (thereby integrating away the steradian sr term), resulting in a radiant power density at the
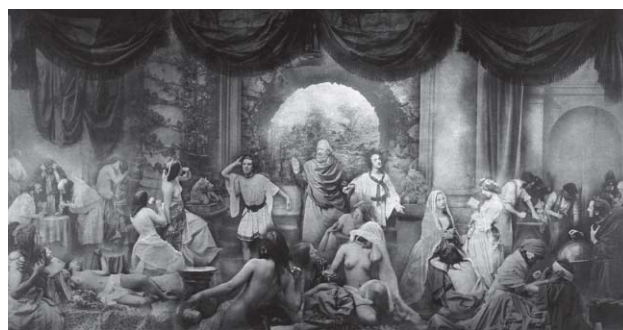

FIGURE 2. *Two Ways of Life*, Oscar Gustave Rejlander, 1857. This is one of the earliest examples of combination printing, in which differently exposed negatives are combined to extend the dynamic range of the final result. In this case, 32 negatives were combined to complete the final image. (Image in the public domain.)
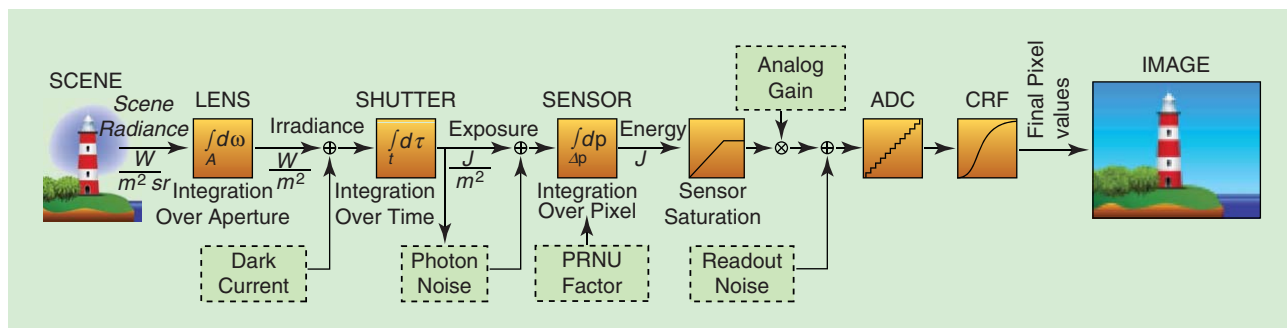
FIGURE 3. The standard imaging pipeline in modern digital cameras, inspired by diagrams in [9] and [10]. The radiance from scene rays captured by the camera are first integrated over the angle subtended by the lens aperture, over the time the shutter is open, and over the pixel's footprint area. This energy can then be cut off by the saturation of the photon well at that pixel sensor, which limits the camera's dynamic range. The result is then quantized by an ADC, and the CRF is applied to get the final digital pixel values. Different kinds of noise or error are injected at various stages in the pipeline, as described in the article text. (Lighthouse image designed by Freepik.com.)

sensor called *irradiance* (E; units: W/m^2). This irradiance is then integrated over the time the shutter is open to produce an energy density, commonly referred to as *exposure* (X; units: J/m^2). If the scene is static during this integration, the exposure can be written simply as $X(p) = E(p) \cdot t$, where p is the point on the sensor and t is the length of the exposure (integration time).

The exposure can then be integrated over the pixel's footprint (integrating away the m^2 term) to result in the total energy (units: J) accumulated in each pixel's photon well. The measured energy is then read out by an analog-to-digital converter (ADC), often with an analog gain factor applied to amplify the energy before it is converted. For non-raw images, the digital value is then mapped through a nonlinear camera response function (CRF) to emulate the logarithmic response of the human eye and make the final image look better. This produces the final pixel values that are output in the image file.

Two aspects of the pipeline limit the sensor's dynamic range of measurable light. First, the pixels' photon wells are of finite size and will saturate if too much energy is accumulated, creating an upper limit for the amount of light energy that can be measured at each pixel. Second, the minimum amount of detectable light is limited by the sources of noise in the imaging pipeline. The first is dark current, which is caused by thermal generation and induces a signal even if no photons arrive at the sensor (i.e., it is dark). Next is photon shot noise, which is caused by the discrete nature of light and is the variance of the number of photons arriving at the sensor during exposure time t. Like many arrival processes, this count is modeled by a Poisson random variable, the expected value (as well as the variance) of which is based on the true irradiance $E(p)$. The spatial nonuniformity of the sensor also causes different pixels to respond differently to the same amount of incident photons, which is modeled by the photo-response nonuniformity (PRNU) factor. Finally, there is readout noise caused by thermal generation of electrons when the signal is being read from the sensor.

Given all of these noise sources (excepting dark current), the actual measured exposure value $\hat{X}(p)$ for well-exposed regions can be modeled as a Gaussian random variable with mean and variance [4]

$$\mu_{\hat{X}(p)} = ga(p)E(p) \cdot t + \mu_R$$
$$\sigma^2_{\hat{X}(p)} = g^2 a(p)E(p) \cdot t + \sigma^2_R, \qquad (1)$$

where g is the camera gain, $a(p)$ is the PRNU factor for the pixel, and μ_R and σ^2_R are the readout mean and variance, respectively. The Poisson nature of the photon shot noise is responsible for the pixel variance's dependence on the irradiance. Without loss of generality, we can think of this measured exposure $\hat{X}(p)$ at each point p in the sensor as being mapped to a final digital pixel value $Z(p)$ with a function f that effectively combines the CRF with the quantization and saturation steps: $Z(p) = f(\hat{X}(p))$.

The challenge of HDR imaging, therefore, is to recover the original HDR irradiance $E(p)$ from noisy LDR images such as $Z(p)$. To do this, two main approaches have been proposed: 1) specialized HDR camera systems that measure a larger dynamic range directly and 2) capturing a stack of differently exposed LDR images that are merged together to produce an HDR result, as described in the following two sections, respectively.

Specialized HDR camera systems

Previous work on specialized HDR camera systems can be divided into two main categories: 1) those that modify the measurement properties of a single sensor to capture a larger dynamic range and 2) those that use prisms, beamsplitters, or mirrors in the optical path to image a number of sensors at different exposures simultaneously.

In the first category, researchers have proposed HDR sensors that measure light in alternate ways, such as measuring the pixel saturation time [11], counting the number of times each pixel reaches a threshold charge level [12], or incorporating a logarithmic response like that of the human eye [13]. Others, such as Nayar and Mitsunaga [14], have proposed to fit different neutral-density filters over individual pixels in the sensor to vary the amount of light absorbed at each pixel.

The main advantage of this spatially varying pixel exposures (SVE) approach is that it allows HDR imaging from a single exposure, thus avoiding the need for alignment and motion estimation. Later, Nayar et al. [15] proposed using a digital micromirror device in front of the sensor for modulating the amount of light that arrives at each pixel to acquire HDR images. Hirakawa and Simon [16] proposed another SVE system that exploits the different sensitivities already present in a regular Bayer pattern, while Schöberl et al. [17] improved this idea further, introducing a nonregular filter pattern to avoid aliasing problems. In addition, a patch-based approach to single-image HDR with SVE acquisition [18] uses a piece-wise linear estimation strategy to reconstruct an irradiance image by simultaneously estimating over- and underexposed pixels as well as denoising the well-exposed ones. Finally, there has been related work that uses a spatial light modulator displaying a random mask pattern to modulate the light before it arrives at the sensor and then uses compressed sensing or sparse reconstruction to recover the HDR image [19].

In the second category, approaches include those that do not use a single sensor but rather split the light onto a set of sensors with different absorptive filters to produce simultaneous images with varying exposures. These exposures can then be merged to form the final HDR result using the stack-based approaches described in the following section. Some systems use pyramid-shaped mirrors, refracting prisms, or beamsplitters to do this [21], although each such approach suffers from parallax errors (because each "looks" through the camera lens from a slightly different angle) as well as wasted light (because of the absorptive filters in front of the sensors). Tocci et al. [20] addressed these problems with a novel beamsplitter design that efficiently reflects the light onto three different sensors to produce high-quality HDR images (Figure 4).

However, despite promising results, all of these specialized HDR systems require the manufacture of new camera hardware, and so they are not widely available today. Nevertheless, this could change as HDR imaging becomes more mainstream.

HDR imaging using image stacks

With conventional cameras, the most practical approach for HDR imaging is to capture a sequence of LDR images at different exposures and combine them into a final HDR result [7]–[9]. Specifically, if we acquire a stack of N different exposures $Z_1, \ldots, Z_N$, we can merge them and estimate the irradiance map $\tilde{E}$ using a simple weighting scheme that takes into account the measured irradiance $\hat{E}_i = \hat{X}_i(p)/t_i$ from each image:

$$\tilde{E}(p) = \frac{\sum_{i=1}^{N} w_i(p) \cdot \hat{X}_i(p)/t_i}{\sum_{i=1}^{N} w_i(p)}. \tag{2}$$

Here, the measured exposure $\hat{X}_i$ can be recovered from well-exposed pixel values using the inverse of the camera response function: $\hat{X}_i(p) = f^{-1}(Z_i(p))$. Of course, this requires the CRF to be known, but methods have been proposed to estimate it from the image stack [9], even for highly dynamic scenes [22].

Because poorly exposed pixels do not have a good estimate for the irradiance map, the weight $w_i(p)$ should be adjusted at each pixel based on how well-exposed it is. For example, Debevec and Malik [9] proposed a simple triangle function for this weight that gives priority to pixels in the middle of the pixel range and reduces the influence of poorly exposed pixels: $w_i(p) = \min(Z_i(p), 255 - Z_i(p))$, where we assume the pixel values range from 0 to 255. Once the stack of images has been merged in this way, the resulting irradiance map $\tilde{E}$ is output as the final HDR result. This method is commonly implemented on modern smartphones to extend their camera's dynamic range (i.e., "HDR mode").

Fundamental limits on irradiance estimation performance

It is interesting to understand the fundamental limits of irradiance estimation performance for stack-based algorithms such as these. To study this, the problem of irradiance estimation from an image stack can be posed as a parameter estimation
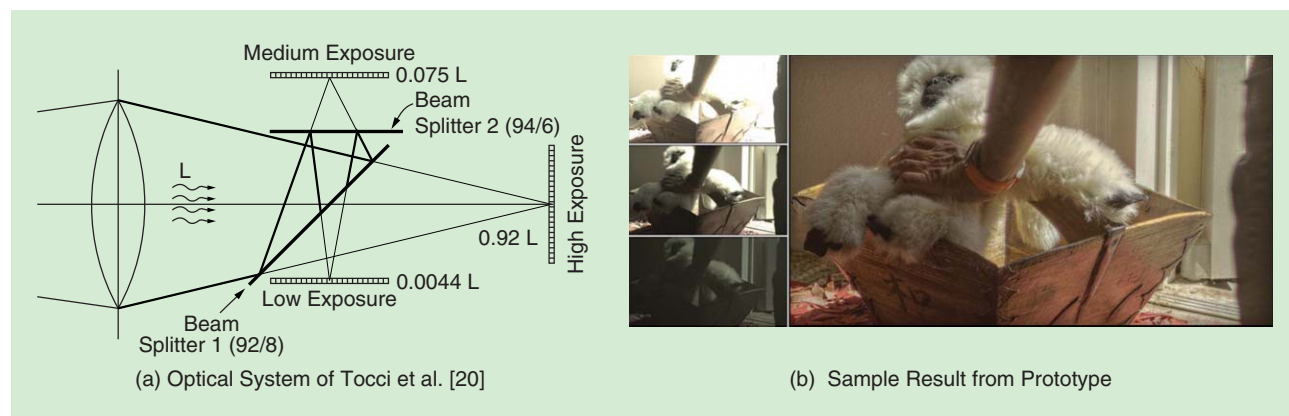
FIGURE 4. In the optical system of Tocci et al. [20], (a) two beamsplitters reflect the light so that the three sensors capture images with 92%, 7.52%, and 0.44% of the total light gathered by the camera lens (increasing the dynamic range by a factor of over 200 ×), and only 0.04% of it is wasted. (b) shows the sample HDR result captured by the camera (the three captured LDR images are on left); note that the detail in both the white fur and dark regions is captured faithfully, even though it does not appear simultaneously in any of the input images. (Figure courtesy of [20].)

problem from a set of noisy samples. In the case of static scenes, N independent samples $\hat{X}_1(p),\ldots,\hat{X}_N(p)$ following the random model in (1) are given per pixel, corresponding to exposure times $t_1,\ldots,t_N$. Assuming the camera parameters are known from a calibration stage, the only unknown parameter in (1) is the irradiance $E(p)$ reaching each pixel p.

In this statistical framework, the Cramér–Rao lower bound (CRLB) gives a lower bound on the variance of any unbiased estimator of $E(p)$ computed from those samples. Aguerrebere et al. [4] introduced the CRLB for this problem and showed that, because the bound cannot be attained, no efficient estimator exists for $E(p)$ under the considered hypotheses. Nevertheless, it was shown experimentally that the approximation of the maximum-likelihood estimator (MLE) proposed by Granados et al. [23] not only outperforms the other evaluated estimators but also has nearly optimal behavior. Theoretically, the MLE is efficient for a large number of samples (asymptotically efficient), which is not the case in HDR imaging, where very few samples are usually available (normally $N = 2$ to 4 exposures). Therefore, it is remarkable that, under the considered hypotheses, the MLE is still experimentally the best possible estimator for the pixel-wise irradiance estimation for static scenes. Improvements, however, may be possible by combining information from different pixel positions with similar irradiance values, such as in recent patch-based denoising approaches [24], or even by considering information from saturated samples [4].

Handling dynamic scenes

The stack-based HDR capture algorithms described in the previous section work very well when the scene is static and the camera is tripod-mounted. However, when the scenes are dynamic or the camera moves while the different pictures are being captured, the images in the stack will not line up properly with one another. This misalignment results in ghost-like artifacts in the final HDR image, which are often more objectionable than the limited dynamic range that is being compensated for (see Figure 5). Because this is the most common scenario in imaging, there has been almost 20 years of research into HDR deghosting algorithms that seek to eliminate these artifacts from motion. Specifically, three different

kinds of methods have been proposed to deal with motion, each of which we discuss in the three sections that follow, using a taxonomy similar to those in two previous publications by the first author [1], [10]. Because of space limitations, we limit the discussion here to a couple of key algorithms in each category.

Algorithms that align the different exposures

The first kind are algorithms that attempt to deghost the HDR reconstruction by warping the individual images in the stack to match a reference image and so eliminate misalignment artifacts. Unlike the rejection methods discussed in the "Algorithms That Reject Misaligned Information" and "Patched-Based Optimization Algorithms" sections, these algorithms can actually move content around in each image and can, therefore, potentially handle dynamic HDR objects.

The simplest methods in this category assume the images can be aligned with rigid transformations. For example, a common method is to compute scale-invariant feature transform (commonly called *SIFT*) features in the image and use them to estimate a homography that warps the images to match [25]. Of course, these simple rigid-alignment algorithms cannot handle artifacts caused by parallax due to camera translation or from significant motion in the scene, although they can serve as a preprocess for more complex algorithms, such as those described later in the article.

One of the first algorithms of this kind was proposed by Bogoni [26]. This method first uses an affine motion estimation step to globally align the images and then estimates motion using optical flow to further align the images. To make the optical flow more robust, some have proposed acquisition schemes to make the different exposures more similar. The Fibonacci exposure bracketing work of Gupta et al. [27], for example, cleverly adjusts the exposure times in the sequence so that the longer exposure times are equal to the sum of the shorter exposure times. Because of this, optical flow can be computed between a longer exposure and the sum of the shorter exposures, thereby ensuring that the two images will have similar exposure times and, therefore, comparable motion blur.

The state-of-the-art HDR alignment algorithm is perhaps the work of Zimmer et al. [28], which aligns the images using

(a) Input LDR Images   (b) Result from Standard Merge   (c) Result from Sen et al. [1]

FIGURE 5. Ghosting artifacts can occur when stack-based HDR algorithms are applied to dynamic scenes. (a) Stack of input LDR images. Note, how some images capture the details in the dark sweater, while others capture the detail in the bright exterior. (b) HDR results from the standard HDR merging algorithm produces ghosting artifacts because of the motion. (c) HDR results from the patch-based optimization algorithm of Sen et al. [1] contains detail in all regions of the image without artifacts.

an energy-based optical flow optimization robust to changes in exposure. Specifically, their energy function has a data term that encourages the image to align to the reference and a regularizer that enforces smooth flow wherever the reference is poorly exposed. However, these alignment algorithms all suffer from the problem of finding good correspondences, which is extremely difficult, in particular for highly dynamic scenes with deformable motion (e.g., a person moving). Furthermore, scenes with occlusion and/or parallax do not even have valid correspondences between the images in these regions, making it impossible to align the images in the stack correctly. Therefore, the HDR results from alignment algorithms often still contain objectionable ghosting artifacts for scenes with complex motion.

Algorithms that reject misaligned information

A second set of algorithms for HDR reconstruction assume that the camera is static (or that the images have been preregistered using a rigid alignment process, such as those described in the "Algorithms That Align the Different Exposures" section) and that the scene motion is localized, meaning that the majority of pixels contain no motion artifacts. The basic goal of these methods is to identify those pixels that are affected by motion and those that are not. The pixels that do not contain motion artifacts can be merged using the standard HDR merging algorithms described in the "HDR Imaging Using Image Stacks" section. For the pixels that are affected by motion, however, only a subset of the images deemed to be static at these pixels will be merged to suppress artifacts from moving objects.

To accomplish this, two different kinds of rejection methods are possible: 1) those in which a reference image is specified by the user and 2) those that do not use a reference image. For algorithms in the first category, the user first selects an image from the stack as the reference. These algorithms then simply revert back to this reference for any pixels where motion is detected so that the main difference between them is in how they detect motion. For example, the method of Grosch [29] assumes two images in the stack and predicts values in the second image by multiplying the values in the reference by the ratio of the exposure times, taking into account the nonlinear camera response curves. With this approach, a pixel is deemed to be affected by motion if the actual color is beyond a given threshold from the predicted value. In such cases, the algorithm simply reverts back to using the values in the reference image for these pixels.

Gallo et al. [30] improved on this work by using the log-irradiance domain to do the threshold comparisons. Further, for robustness they compare patches instead of individual pixels, so that a patch from an image in the stack would be merged with the corresponding patch from the reference only if a certain number of pixels meet the threshold constraint. To reduce visible seams between different patches, the authors apply Poisson blending to the final results.

In the second category are rejection algorithms without a reference image, which must select a "static" subset of images at every pixel to merge to produce HDR values. These methods have a fundamental advantage over those that utilize a single reference image because motion may occur in areas where the reference might be poorly exposed. At these pixels, an HDR value cannot be properly computed solely from the reference image. However, rejection algorithms that do not use a reference must ensure that subsets are selected for neighboring pixels in a way that does not introduce artifacts.

Reinhard et al. [3] proposed one of the earliest methods in this category. For every pixel that is deemed to be affected by motion, the authors try to use the longest exposure that is not saturated (effectively, a single-image subset). To determine which pixels are affected by motion, they first compute the variance of the irradiance values at each pixel p, weighted to exclude poorly exposed pixels. This estimated variance is then thresholded, and the result is smeared out with a 3×3 kernel to reduce edge and noise effects. Adjacent regions are then joined together to form the "ghosted" regions for which a single image from the stack will be used. To select which image they will use for each region, the authors find the biggest irradiance value in the region that is not in the top 2% (deemed to be outliers). They then select the longest exposure that includes this value within its valid range to fill in this ghosted region, because the longest exposure will contain least noise. To further suppress artifacts, Reinhard et al. linearly interpolate this exposure with the original HDR result, using the per-pixel variance as a blending parameter.

An alternative approach is proposed by Khan et al. [31]; here, instead of detecting and handling differently the pixels affected by motion, the authors propose to iteratively weight the contribution of each pixel depending on the probability of its being static (i.e., belonging to the background of the scene). To do this, they assume that most of the pixels are of the static background and so determine the probability of a pixel being static by measuring its similarity to the neighborhood around it.

Finally, some recent methods cleverly use rank minimization to deghost HDR images [32], [33]. These methods are based on the observation that if the scene is static, the different exposure images $X(p)$ would simply be linear scalings of one another. Therefore, they use the different exposure images to construct a matrix and essentially minimize its rank to solve for the motion-free image.

The biggest problem with these and other rejection algorithms is that they cannot handle dynamic HDR content because they do not move information between pixels but rather only merge information from corresponding pixels across the image stack. Therefore, if different parts of a moving HDR

> **The biggest problem with rejection algorithms is that they cannot handle dynamic HDR content because they do not move information between pixels but rather only merge information from corresponding pixels across the image stack.**

object are well exposed in disjoint regions of the different images, these parts cannot be brought together to produce an acceptable result.

Patch-based optimization algorithms

Recently, Sen et al. [1] proposed a new alternative for HDR deghosting that uses patch-based optimization, which addresses the problems of both the rejection and alignment methods. Specifically, a formulated equation codifies the objective of most reference-based HDR reconstruction algorithms: 1) to produce an HDR result that resembles the reference image in the parts where the reference is well exposed and 2) to leverage well-exposed information from other images in the stack wherever the reference is poorly exposed. This HDR synthesis equation can be written as

$$
\text{Energy}(E) = \sum_{p \in \text{pixels}} [\alpha_{\text{ref}}(p) \cdot (f^{-1}(Z_{\text{ref}}(p))/t_{\text{ref}} - E(p))^2 \\
+ (1 - \alpha_{\text{ref}}(p)) \cdot E_{\text{BDS}}(E \mid Z_1, \ldots, Z_N)]. \tag{3}
$$

The first term states that the desired HDR image E should be close in an L_2 sense to the LDR reference Z_{ref} mapped to the linear irradiance domain by applying the inverse camera response function f^{-1} and dividing by the exposure time t_{ref}. This is only to be done for the pixels where the reference is

properly exposed, as given by the α_{ref} term, which is a trapezoidal function in the pixel value domain [similar to the weighting function in (2)] that favors intensities near the middle of the pixel value range.

In the regions where the reference image Z_{ref} is poorly exposed (indicated by $1 - \alpha_{\text{ref}}$), the algorithm draws information from the other images in the stack using a bidirectional similarity metric, given by the E_{BDS} term. This energy term enforces that for every pixel patch in the image stack (given by $Z_1, \ldots, Z_N$), there must be a similar patch in the final result E, and vice versa. The first similarity ensures that as much well-exposed content from the image stack is included in the final HDR result, while the second ensures that the final result does not contain objectionable artifacts, as these artifacts would not be found anywhere in the stack. This energy equation is optimized with an iterative method that solves for the aligned LDR images and the HDR image simultaneously, producing high-quality results (Figure 6).

Patch-based optimization algorithms like this are fundamentally different from those discussed in the "Algorithms that Align the Different Exposures" section, which warp the images to match based on correspondences. As was pointed out earlier, alignment methods fail in cases of occlusion or parallax (which happen commonly in dynamic scenes) because

FIGURE 6. (a) and (b) show sample HDR results (right) from the input LDR images (left) using the patch-based optimization of Sen et al. [1].

they do not have valid correspondences in these regions and so the images cannot be aligned in these parts. Patch-based HDR reconstruction, on the other hand, is related to patch-based image synthesis methods (e.g., for single-image hole filling) because they both use a patch-based similiarity optimization to resynthesize content in the final reconstruction without an underlying correspondence. Because of this advantage, these methods have proved to be the most successful HDR deghosting algorithms proposed to date.

For example, a recent state-of-the-art report by Tursun et al. [6] testing many deghosting algorithms found that the algorithm of Sen et al. [1] and the later, related method of Hu et al. [34] ranked first and second over other deghosting techniques by a fairly large margin. The success of patch-based optimization for HDR reconstruction has led others to explore ways to further improve the quality of these approaches. For example, Aguerrebere et al. [24] focused on reducing the noise of the estimated irradiance. First, this method synthesizes a "reference" containing well-exposed, de-ghosted information in all parts of the image using Poisson image editing (although the method in Sen et al. [1] could also be used). Noise is then reduced through a patch-based denoising method that finds all patches in the image stack within a threshold to each patch in the reference, where the L_2 distance between patches is normalized by the variance from (1). The MLE of the patch-centers at each pixel is then computed to significantly reduce the noise in the final result.

> **Patch-based HDR reconstruction is related to patch-based image synthesis methods (e.g., for single-image hole filling) because they both use a patch-based similiarity optimization to resynthesize content in the final reconstruction without an underlying correspondence.**

HDR video

Up to now, we have focused exclusively on the HDR acquisition of still images. However, the problem of capturing HDR video sequences is of considerable interest as well. For example, filmmaking companies incur a significant cost to light sets, a cost that would be largely eliminated by high-quality, HDR video systems. For this reason, professional movie camera system suppliers such as RED have been pushing the dynamic range of standard sensors. Moreover, specialized HDR camera systems such as that of Tocci et al. [20] have been proved capable of capturing high-quality HDR video, although they are not yet widely available.

For conventional digital cameras, the only way to capture HDR video is to alternate exposures through the entire sequence. This problem was first tackled by Kang et al. [35], who use gradient-based optical flow to compute a bidirectional flow from the current frame to neighboring frames and unidirectional flows from neighboring frames to the current frame (four flows total). Once computed, the flows can be used to produce four warped images by deforming each of the two neighboring frames. The resulting images can be merged with the reference to produce an HDR image at every frame of the sequence, while rejecting the pixels that are still misaligned, to avoid artifacts.

The state of the art in HDR video reconstruction is the work of Kalantari et al. [5], which extended the patch-based optimization work of Sen et al. [1] to produce coherent HDR video streams. Specifically, they modify the HDR image synthesis equation (3) to enforce temporal coherence by performing a bidirectional similarity between adjacent frames. In addition, they use optical flow during the optimization to constrain the patch-based search, which produces a stream of high-quality HDR frames.

Open problems and challenges

Despite the tremendous progress of the computational photography community on HDR imaging over the last 20 years, many challenges remain. For example, the capture of high-quality HDR images of highly dynamic scenes with conventional digital cameras is still a challenging problem. Although state-of-the-art deghosting algorithms like the patch-based optimization of Sen et al. [1] can suppress many of the ghosting artifacts that would normally occur in these scenes, these methods cannot recover scene content that is poorly exposed in the reference image and is not visible in any of the other images in the stack. Moreover, the patch-based optimization in these algorithms is computationally expensive and can take several minutes to compute an image. This limits the applicability of these methods to long video sequences or for real-time, on-board computation in current smart phones, for example.

It is entirely possible that new sensor technologies, such as Fuji Film's recent Super CCD EXR sensor, will bypass the problems inherent in stack-based methods by capturing a single image with extended dynamic range. However, even these new technologies will likely raise interesting questions, such as how users will employ and interact with HDR images. Furthermore, as HDR imaging becomes more mainstream, we expect that new applications for HDR imaging (such as for medical imaging or manufacturing) will be proposed and explored.

Conclusions

In this article, we first summarized the main aspects of HDR imaging, starting with an overview of the problem of limited dynamic range in standard digital cameras and the physical constraints responsible for this limitation. We then surveyed state-of-the-art approaches developed to tackle the HDR imaging problem, focusing on both specialized HDR camera systems and stack-based approaches captured with standard cameras. For the latter, we discussed algorithms to address ghosting artifacts that can occur when capturing dynamic scenes. Finally, we discussed algorithms for capturing HDR video and concluded with a review of open problems in HDR imaging. We hope that

this article encourages researchers from areas such as signal processing, solid-state devices, and image processing to continue to pursue this interesting set of problems.

Acknowledgments

Authors

Pradeep Sen (psen@ece.ucsb.edu) received his B.S. degree from Purdue University, West Lafayette, Indiana, and his M.S. and Ph.D. degrees from Stanford University, California. He is currently an associate professor with the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. He has coauthored more than 30 technical publications, including eight in *ACM Transactions on Graphics*. His research interests include algorithms for image synthesis, computational image processing, and computational photography. He received the 2009 National Science Foundation CAREER Award and more than US$2.2 million in funding.

Cecilia Aguerrebere (cecilia.aguerrebere@duke.edu) received her B.S., M.S., and Ph.D. degrees in electrical engineering from the Universidad de la República, Montevideo, Uruguay, in 2006, 2011, and 2014, respectively; an M.S. degree in applied mathematics from the École normale supérieure de Cachan, France, in 2011; and a Ph.D. degree in signal and image processing from Télécom ParisTech, France, in 2014 (as part of a joint Ph.D. program with Universidad de la República). Since August 2015, she has been with the Department of Electrical and Computer Engineering at Duke University, Durham, North Carolina, where she holds a postdoctoral research associate position.

References

[1] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203:1–203:11, 2012.

[2] R. K. Mantiuk, K. Myszkowski, and H.-P. Seidel, *High Dynamic Range Imaging.* Hoboken, NJ: Wiley, 2015.

[3] E. Reinhard, G. Ward, S. N. Pattanaik, and P. E. Debevec, *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting.* San Mateo, CA: Morgan Kaufmann, 2005.

[4] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Best algorithms for HDR image generation. A study of performance bounds," *SIAM J. Imaging Sci.*, vol. 7, no. 1, pp. 1–34, 2014.

[5] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen, "Patch-based high dynamic range video," *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, vol. 32, no. 6, pp. 202:1–202:8, Nov. 2013.

[6] O. T. Tursun, A. O. Akyz, A. Erdem, and E. Erdem, "The state of the art in HDR deghosting: A survey and evaluation," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 683–707, 2015.

[7] B. C. Madden, "Extended intensity range imaging," Univ. of Pennsylvania, Tech. Rep. MS-CIS-93-96, 1993.

[8] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IS&T*, 1995, pp. 442–448.

[9] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.

[10] O. Gallo and P. Sen, "Stack-based algorithms for HDR capture and reconstruction," in *High Dynamic Range Video, From Acquisition to Display and

Applications*, 2nd ed., F. Dufaux, P. L. Callet, R. Mantiuk, and M. Mrak, Eds. Amsterdam, The Netherlands: Elsevier, 2016, ch. 3.

[11] V. Brajovic and T. Kanade, "A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors," in *Proc. ICRA,* vol. 2, pp. 1638–1643, Apr. 1996.

[12] H. Zhao, B. Shi, C. Fernandez-Cull, S.-K. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, Apr. 2015, pp. 1–10.

[13] U. Seger, U. Apel, and B. Höfflinger, "HDRC-Imagers for natural visual perception," in *Handbook of Computer Vision and Application*, B. Jähne, H. Haußecker, and P. Geißler, Eds. New York: Academic, 1999, vol. 1, pp. 223–235.

[14] S. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2000, vol. 1, pp. 472–479.

[15] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable imaging: Towards a flexible camera," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 7–22, 2006.

[16] K. Hirakawa and P. Simon, "Single-shot high dynamic range imaging with conventional camera hardware," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 1339–1346.

[17] M. Schöberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup, "High dynamic range video by spatially non-regular optical filtering," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2012, pp. 2757–2760.

[18] C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau, and P. Musé, "Single shot high dynamic range imaging using piecewise linear estimators," in *Proc. IEEE Int. Conf. Computational Photography (ICCP)*, 2014, pp. 1–10.

[19] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum*, vol. 35, no. 2, 2016.

[20] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 41:1–41:10, July 2011.

[21] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *Int. J. Comput. Vis.*, vol. 58, no. 1, pp. 7–17, June 2004.

[22] A. Badki, N. K. Kalantari, and P. Sen, "Robust radiometric calibration for dynamic scenes in the wild," in *Proc. IEEE Int. Conf. Computational Photography (ICCP)*, Apr. 2015, pp. 1–10.

[23] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 215–222.

[24] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Simultaneous HDR image reconstruction and denoising for dynamic scenes," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2013, pp. 1–11.

[25] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proc. Int. Conf. Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 2007, pp. 49–56.

[26] L. Bogoni, "Extending dynamic range of monochrome and color images through fusion," in *Proc. IEEE Intl. Conf. Pattern Recognition (ICPR)*, 2000, pp. 3007–3016.

[27] M. Gupta, D. Iso, and S. Nayar, "Fibonacci exposure bracketing for high dynamic range imaging," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2013, pp. 1473–1480.

[28] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 405–414, Apr. 2011.

[29] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," in *Proc. Int. Symp. Vision, Modeling and Visualization*, 2006, pp. 277–284.

[30] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2009, pp. 1–7.

[31] E. Khan, A. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2006, pp. 2005–2008.

[32] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Processing Lett.*, vol. 21, no. 9, pp. 1045–1049, Sept. 2014.

[33] T.-H. Oh, J.-Y. Lee, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Machine Intelli.*, vol. 37, no. 6, pp. 1219–1232, June 2015.

[34] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Computer Vision and Pattern Recog. (CVPR)*, June 2013, pp. 1163–1170.

[35] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, July 2003.

SP

Ayush Bhandari and Ramesh Raskar

Signal Processing for Time-of-Flight Imaging Sensors

An introduction to inverse problems in computational 3-D imaging
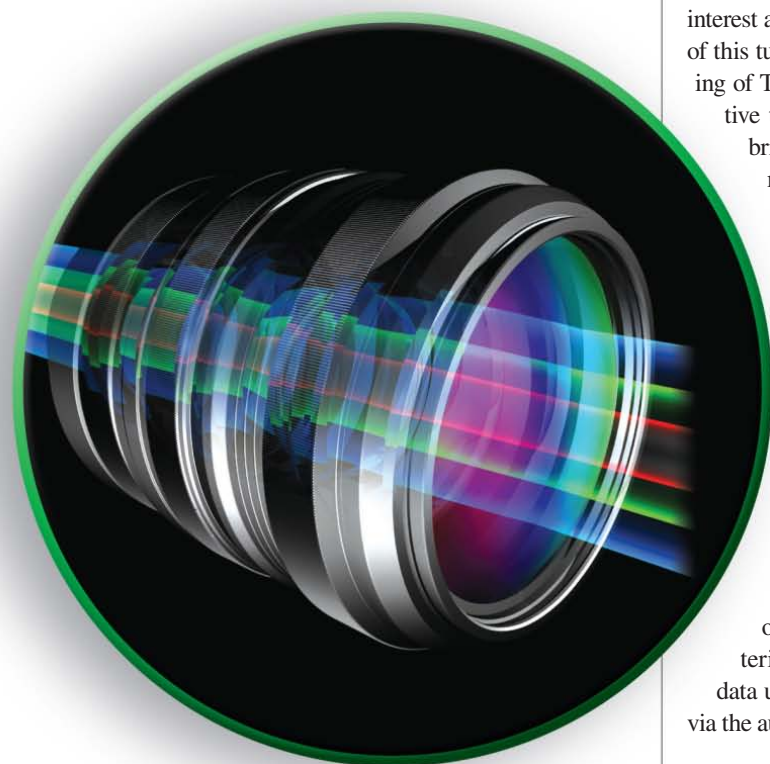
©ISTOCKPHOTO.COM/YAKOBCHUK

Time-of-flight (ToF) sensors offer a cost-effective and real-time solution to the problem of three-dimensional (3-D) imaging—a theme that has revolutionized our scene-understanding capabilities and is a topic of contemporary interest across many areas of science and engineering. The goal of this tutorial-style article is to provide a thorough understanding of ToF imaging systems from a signal processing perspective that is useful to all application areas. Starting with a brief history of the ToF principle, we describe the mathematical basics of the ToF image-formation process, for both time- and frequency-domain, present an overview of important results within the topic, and discuss contemporary challenges where this emerging area can benefit from the signal processing community. In particular, we examine case studies where inverse problems in ToF imaging are coupled with signal processing theory and methods, such as sampling theory, system identification, and spectral estimation, among others. Through this exposition, we hope to establish that ToF sensors are more than just depth sensors; depth information may be used to encode other forms of physical parameters, such as, the fluorescence lifetime of a biosample or the diffusion coefficient of turbid/scattering medium. The MATLAB scripts and ToF sensor data used for reproducing figures in this article are available via the author's webpage: http://www.mit.edu/~ayush/Code.

Introduction

A brief history of computational imaging

"A picture is worth a thousand words." Throughout the past several centuries, this immemorial phrase has pushed the development of photography from analog to digital, accelerated by breakthrough advancements in sensor technology. The digital revolution created a necessity for sophisticated signal processing algorithms tailored for image enhancement, storage, and compression. Shannon's sampling theory was the pathway for

analog-to-digital image conversion. The classical Weiner filter was used for image enhancement. Transform-domain coding utilizing discrete cosine and wavelet transforms played a pivotal role in JPEG compression. Over the last few decades, with the advent of the wavelet transform and compressed sensing theory, the field of signal processing has undergone a philosophical reformation. A field that was once heavily reliant on smoothness assumptions now uses principles inspired by the notion of sparsity. Until the last decade or so, the fields of image sensor technology and signal processing ran in parallel to one another with minimal interaction or exchange of know-how. However, there has recently been a growing trend toward the coherent codesign of sensors and algorithms: this is the theme of the emerging area of computational imaging or computational photography [1].

Practitioners of this computational imaging/photography ideology have devised many solutions that were previously not possible when individually adding functionality to the sensor architecture or using a more sophisticated processing algorithm. For example, consider the problem of motion deblurring that arises in conventional imaging. Whenever an object moves during sensor exposure, it causes pixels to smear across the frame, resulting in a blurred image. In the context of signal processing, this is an ill-posed problem that has been well explored within the theme of deconvolution. The key problem is that the exposure time defines a temporal filter, which is essentially a box filter that annihilates any high-pass, spatial information. Consequently, algorithmic sophistication alone is not enough. Blurring may be avoided by a shorter exposure time, but this comes at the expense of low signal-to-noise ratio (SNR). It is clear that neither deconvolution nor the sensor level adaptation in itself suffice for a solution to the deblurring problem. The distinct role of computational imaging emerges when one considers the so-called flutter-shutter approach [2]. This approach involves a codesign of sensor and algorithms: in contrast to traditional imaging methods, which require the shutter to be fully open during the exposure time, the flutter-shutter method flutters the shutter on and off in a binary, pseudo-random sequence. This sequence converts the noninvertible box-filter into an invertible one and, based on the choice of pseudo-random sequence, the corresponding deconvolution filter may be devised. Beyond deblurring in consumer imaging, the flutter-shutter approach is also useful in bioimaging [3], where the imaging sensor may not be fast enough to capture flowing structures, such as blood cells. Other notable examples of the computational imaging philosophy are high-dynamic-range imaging [4], light-field imaging [5], [6], single-pixel imaging [7], and Fourier ptychographic microscopy [8].

For the most part, image sensor design, signal/image processing, and computational imaging have largely been restricted to two-dimensional (2-D) scenes. However, a true and richer representation of the environment around us lives in a 3-D space. Capturing 3-D information of a scene offers unparalleled benefits in accuracy and capabilities and is surely the future in many areas. This necessitates development of imaging modalities capable of recording 3-D images.

A number of methods have been developed for the purpose of 3-D imaging. An overview of the main techniques is presented in

"An Overview of 3-D Imaging Techniques." Of all the 3-D capture techniques, the ToF method has arguably attracted the most commercial and scientific interest in the last couple of years; there has been a surge of research toward improving both the sensor design as well as the algorithms used for processing 3-D images. ToF imaging is the theme of this article, and we take a deep dive into the topic in the following section.

The time-of-flight revolution

The ToF principle exploits the idea that distance and time are proportional quantities. As the name suggests, ToF is the round-trip time between the source and the destination taken by a particle or a wave. Hence, knowing one entity is equivalent to knowing the other. Nature is replete with examples that rely on the ToF principle. Bats, dolphins, and visually impaired human beings use the ToF principle for navigational purposes.

Chronologically, the use of sound waves superseded the use of electromagnetic waves. Humans have used stones to estimate the depth of wells for millennia. The earliest work on using light waves for measuring ToF dates back to an experiment conducted by Galileo, who was interested in estimating the speed of light. Unfortunately, his choice of distance (the separation between two hills) did not lead to a conclusive result. The Danish astronomer Ole Rømer overcame this difficulty by using planetary distances. About 200 years later, the French physicist Hippolyte Fizeau was the first to precisely estimate the speed of light. Through the discovery of the law of the photoelectric effect by Albert Einstein in the 1900s and the development of the electronic imaging sensors [charge couple device (CCD)/complementary metal–oxide–semiconductor (CMOS)], we are now at a point where the accumulated research efforts in the area of photonics and electronics have culminated in mass-producible optical ToF sensors.

Contrary to conventional imaging sensors such as digital cameras that produce 2-D images $I(x, y)$, ToF sensors capture 3-D images, $I(x, y, z)$. The unique ToF sensor produces two images per exposure: an amplitude image and a depth image. The amplitude image is the standard 2-D photograph, $I(x, y)$. Each pixel on the depth image represents the corresponding distance in the scene. The combination of the amplitude and the depth image produces the 3-D image. Using our custom-designed ToF sensor, we show the amplitude, depth, and resulting 3-D images in Figure 1.

ToF-based 3-D imaging allows for applications that were previously unexplored. One of the first results demonstrated non-line-of-sight imaging capability [10]. This result—in parallel to "Doc" Edgerton's iconic *Bullet Through Apple* image (see Figure 2)—led to ultrafast imaging of light packets at an exorbitant frame rate of one trillion frames per second. A flurry of follow-up work lead to results that allowed imaging through scattered media [11], light-in-flight imaging [12], [13], and 3-D imaging in extremely low light [14].

With the advent of 3-D sensing technology (most notably, the Microsoft XBox One's Kinect) we can now replace a room-sized apparatus [10], moving sensors, and raster scan systems [14] by miniaturized, cost-effective, real-time, and full-frame
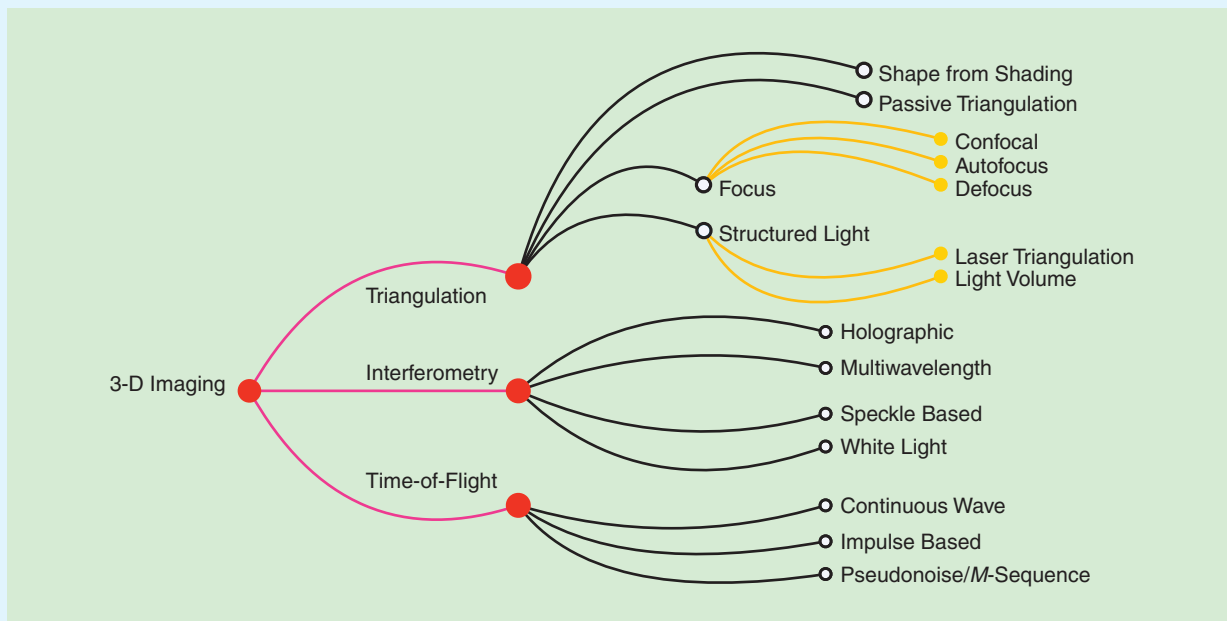
An Overview of 3-D Imaging Techniques

FIGURE S1. Taxonomic classification of major 3-D imaging modalities.

As shown in Figure S1, optical 3-D scene capture can be broadly categorized into the following three approaches:

1) The triangulation [9] method relies on the trigonometric principles. The distance to an unknown point is measured by computing the respective angles to the point from two edges of a triangle.

2) The ToF method, as the name suggests, relies on the time it takes for light to backscatter from an object at an unknown distance. Since time delays are linearly proportional to the distance, measuring the ToF amounts to measuring the range of the object.

3) Interferometry is similar to ToF with the main distinction that it requires the light waves to be coherent.

Each of these approaches can be further classified based on the application as well as the specialized principle linked with the approach.
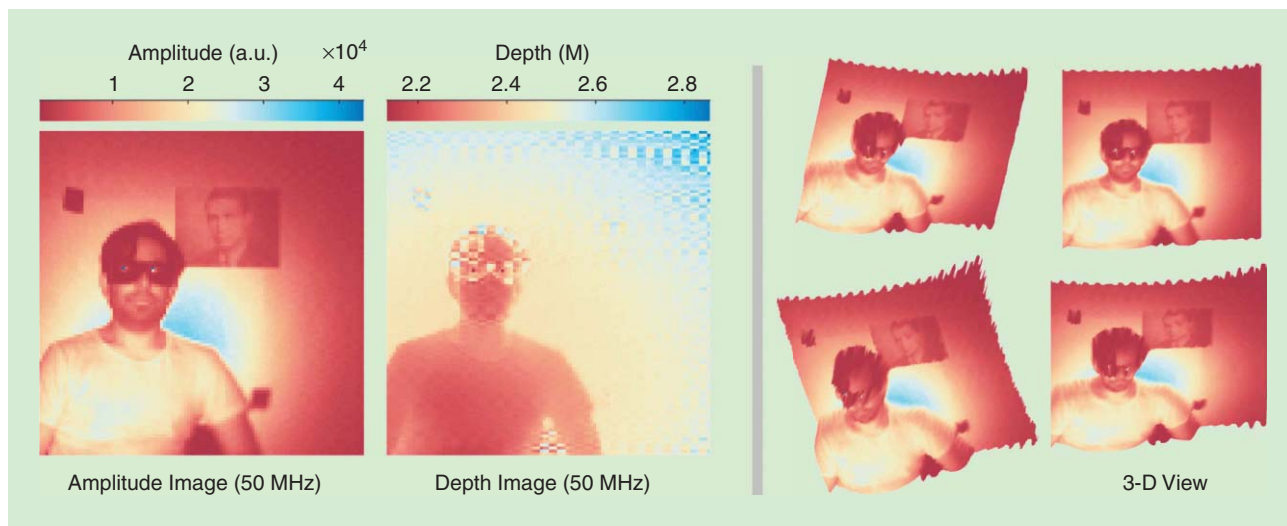
FIGURE 1. Three-dimensional images captured via a ToF sensor. We show an amplitude image (or the conventional digital image), a depth image, and 3-D images seen from multiple viewpoints.
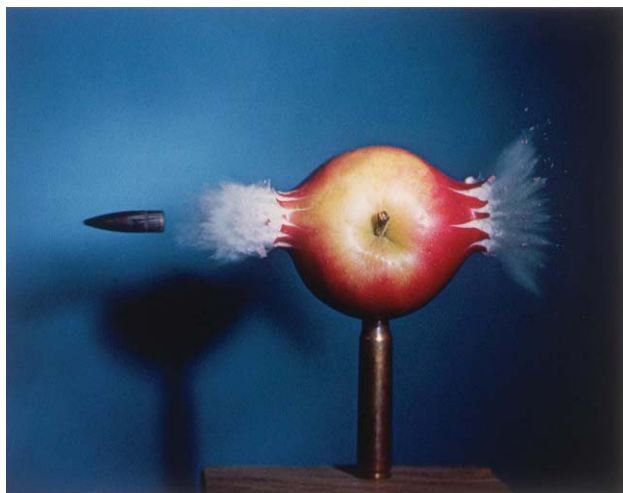
FIGURE 2. Harold Edgerton's iconic *Bullet Through Apple* image (1964) was one of his many early demonstrations of high-speed imaging. His technique involved the use of stroboscopic light to capture high-speed motion. (© 2010 MIT. Courtesy MIT Museum.)

ToF systems. Computational ToF imaging has already found a plethora of applications in, for example, ultrafast imaging [15], [16], non-line-of-sight imaging [17], imaging through scattering media [18], and colored ToF imaging [19]. Outside of the computational imaging and human-computer interaction communities, an important application area is health-care technology [20] and bioimaging [21].

ToF sensors motivate a demand for forward models and algorithms that can handle this new wave of data. A handful of algorithms inspired by the signal processing community have been used to tackle inverse problems in optical ToF, but significant and numerous challenges still remain. Current efforts are directed at establishing empirical results with a rare discussion on the design of efficient algorithms, fundamental limits, or performance bounds.

While optical ToF sensors are a recent phenomenon, other ToF systems such as ultrasound, seismic, and radar technology have been around for decades. The knowledge transfer between optical and other ToF systems is far from reality. Each ToF modality has its own idiosyncratic constraints that stem from the physics of the problem. However, there are commonalities that are shared by all of these systems. By discussing specific case studies, we take a first step toward the goal of bridging this gap.

A road map of this article

The main goal of our tutorial-style article is to introduce the audience to the rapidly emerging field of ToF imaging from a signal processing perspective. We start this journey with a primer on the ToF image formation model. To reconcile any confusion stemming from taxonomic classification of ToF modalities—time domain (TD–ToF), frequency domain (FD–ToF) or amplitude-modulated continuous wave (AMCW–ToF)—we discuss a unifying image formation model that is applicable to all known optical ToF systems. Furthermore, this model is backward compatible with other nonoptical ToF modalities, such as terahertz,

ultrasound, and microwave. Our first milestone covers a discussion on acquiring a single-depth image: how consumer-grade ToF sensors capture 3-D images? From there, the discussion turns to computational ToF imaging. Within the scope of computational ToF imaging, we enumerate case studies that map ToF imaging problems to signal processing subfields such as 1) line spectrum estimation theory, 2) sampling theory of sparse signals, and 3) system identification. Finally, we discuss current challenges and future directions in computational ToF imaging where signal processing theory and methods can contribute significantly.

3-D imaging with ToF sensors

ToF sensors are active illumination devices consisting of an illumination unit capable of probing a scene with an amplitude-modulated light that is not necessarily coherent. We call this amplitude-modulated waveform the *probing function* or $p(x, y, t)$, where (x, y) are the spatial coordinates and t is the continuous time variable. For simplicity of exposition, we will consider per-pixel processing and simply write $p(t)$. The probing function interacts with the scene response function (SRF). This interaction results in the reflected signal $r(t)$. The reflected signal is observed at the ToF sensor, which is characterized by its transfer function that we refer to as the instrument response function (IRF). The IRF models the sensor's electro-optical assembly. For example, the IRF for a digital camera is the point spread function of the lens. The interaction between the reflected signal and the ToF sensor results in the measured signal $m(t)$, which is converted to a digital signal via sampling. The precise mathematical description of the ToF imaging pipeline is discussed in "ToF Image Formation Model." Based on the ToF image formation model parameters, $\{p, h, \varphi\}$, one may now define the specific inverse problem at hand. This is also true of other wave-based ToF modalities—radar, sonar, ultrasound, terahertz, and so on. A distinct feature of the consumer ToF sensors is their use of the lock-in principle [22], which implements the cross-correlation operation. From a mathematical standpoint, and in the absence of noise and distortion, this translates to the fact that,

$$\varphi(t, \tau) = p(\tau + t). \tag{1}$$

Next, we discuss how consumer ToF sensors capture 3-D images. Depending on the choice of probing function, the ToF imaging setup may be categorized into time-domain or frequency-domain modes. In either case, the SRF for the case of single-depth imaging is modeled as a shift-invariant kernel,

$$h(t, \tau) = \Gamma_0 \delta(t - \tau - t_0), \quad t_0 = 2\frac{d_0}{c}, \tag{2}$$

where δ denotes Dirac distribution and the goal is to estimate the amplitude and delay (or depth), $\{\Gamma_0, t_0\}$, respectively, at each pixel.

Time-domain 3-D imaging (TD–ToF)

TD–ToF systems probe the scene with a time-localized, periodic signal of form $p(t) = p(t + \Delta), \Delta > 0$. While

ToF Image Formation Model

As shown in Figure S2, we next elaborate on the elements of the ToF imaging pipeline.

1) The probing function, denoted by $p(t)$, represents the waveform emitted by the ToF sensor's illumination unit. The probing function may be a time-localized pulse or a continuous wave. In either case, it is chosen to be a periodic function of form $p(t) = p(t + \Delta), \Delta > 0$.

2) The scene response function (SRF), denoted by $h(t, \tau)$, models the transfer function of the scene. For example, for an object with reflection coefficient Γ and at a depth d meters away from the sensor, the SRF takes form of, $h(t, \tau) = \Gamma \delta(t - \tau - 2d/c)$. Here, c is the speed of light. The SRF may also be characterized as the Green's function of a differential equation that models some physical phenomenon such as fluorescence emission, diffusion or scattering.

3) The reflected function, denoted by $r(t)$, is the result of interaction between the probing signal and the SRF. The reflected signal is modeled as a Volterra/Fredholm integral,

$$r(t) = \int_{\Omega_1} p(\tau) h(t, \tau) d\tau. \tag{S1}$$

Whenever the SRF is a shift-invariant kernel, i.e., $h_{SI}(t, \tau) = h(t - \tau)$, the reflected signal is simply a convolu-

tion/filtering operation between the probing function and the SRF, $r(t) = (p * h)(t)$.

4) The instrument response function (IRF), denoted by $\varphi(t, \tau)$, models the transfer function of the electro-optical elements of the ToF sensor. For example, in conventional digital cameras, the spatial IRF is the point spread function of the lens.

5) Measurements denoted by $m(t)$ are a result of sensing the reflected signal via the electro-optical elements of the ToF sensor. Continuous-time measurements are modeled as

$$m(t) = \int_{\Omega_2} r(\tau) \varphi(t, \tau) d\tau. \tag{S2}$$

The ToF sensor stores discrete measurements by sampling continuous-time signal $m(t)$ and this results in the discrete sequence $m[k] = m(kT)|_{t = kT, k \in \mathbb{Z}}$, where $T > 0$ is the sampling interval.

In many practical cases of interest, both the SRF and the IRF are shift-invariant. In that case, the measurements can be written as a convolution product, $m(t) = (p * h * \varphi)(t)$. Whenever the IRF is a function of form $\varphi(t, \tau) = \varphi(t + \tau)$, the measurements amount to $m(t) = (r \otimes \varphi)(t)$ where $\otimes$ denotes cross-correlation operation. "Lock-in" sensors operate on this principle.
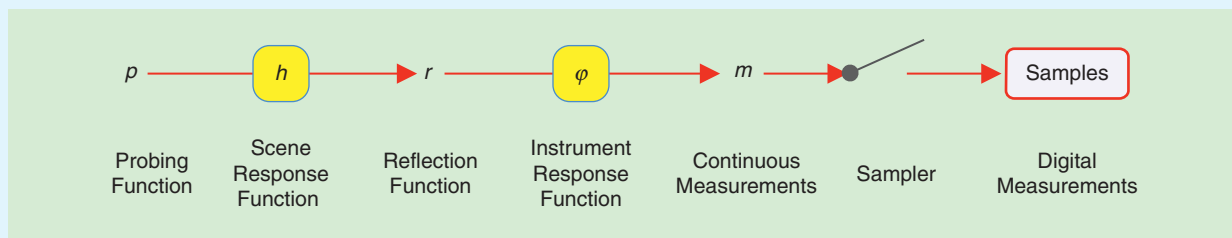

FIGURE S2. A block diagram for the ToF imaging pipeline.

specialized scientific instruments, such as the streak tube, may be able to produce a pulse that mimics the Dirac's Delta distribution δ [10], this form of precision is impractical for consumer-grade instruments. In practice, a maximum-length sequence (MLS) [23] is an optimal choice of probing function in regard to time-localization. In this case, given (S1) and (2), the reflected signal reads $r(t) = \Gamma_0 p(t - t_0)$, with delay $t_0 = 2d_0/c$. Due to the lock-in sensor architecture [22], which constraints the IRF in (1), the measurements simplify to $m(t) = (r \otimes p)(t) \equiv (\bar{r} * p)(t)$, where $\otimes$ denotes the cross-correlation operation and $\bar{r}(t) = r(-t)$. A closer look at m reveals an underlying autocorrelation operation involving p, that is, $\bar{m}(t) = \Gamma_0 (p * \bar{p})(t - t_0)$. Consequently, we may write $\phi = p * \bar{p}$. The ToF is then estimated by solving for $\tilde{t}_0 = \arg\max_{t_0} \bar{m}(t) = \arg\max_{t_0} \phi(t - t_0)$.

Whenever $p(t)$ is modeled to be some parametric waveform, such as a Gaussian function, B-spline, or a combination of parametric pulses, parameter estimation techniques may be used to estimate the ToF t_0 and the reflection coefficient Γ_0. However, this may not be the case in practice because of model mismatch or the physical aspects of light propagation. In such a setting, it is effective to use the property of band-limited approximation: approximate p, and hence ϕ, with the first few Fourier components,

$$\tilde{p}(t) = \sum_{|m| \leq M_0} \hat{p}_m e^{Jm\omega_0 t} \text{ with } \hat{p}_m = \frac{1}{\Delta} \int_0^\Delta p(t) e^{-Jm\omega_0 t} dt, \tag{3}$$

where $\omega_0 = 2\pi/\Delta$ is the fundamental frequency and Δ is the maximum operating range of the ToF sensor. This choice is aptly justified if one considers the fact that

- Most electronic/optical instrumentation is approximately bandlimited due to physical constraints [24].
- The probing function may not admit a parametric representation. Even if the probing function assumes a parametric representation, bandlimited approximation via Fourier series coefficients circumvents the estimation of parameters of the probing function.

The utility of bandlimited approximation property is demonstrated via experiments shown in Figure 3(a). Starting with an MLS, we design a probing function. We plot $\phi = p * \bar{p}$ together with its bandlimited approximation

$\tilde{\phi}(t)$ obtained by retaining first M_0 Fourier series coefficients $\{\hat{\phi}_m = |\hat{p}_m|^2\}_{|m| \leq M_0}$. We are thus able to rewrite measurements as $\tilde{m}(t) = \Gamma\phi(t - t_0) \equiv C_0 \sum_{|m| \leqslant M_0} \hat{\phi}_m e^{jm\omega_0 t}$, where the complex-valued constant $C_0 = \Gamma e^{-j\omega_0 t_0}$ is the unknown to be estimated.

Frequency-domain 3-D imaging (FD–ToF)

ToF sensors, such as the Microsoft XBox One's Kinect, use a continuous wave-based probing function $p(t) = 1 + p_0\cos(\omega t)$, $p_0 < 1$, where ω is the modulation frequency and p_0 is the modulation amplitude. With the SRF defined
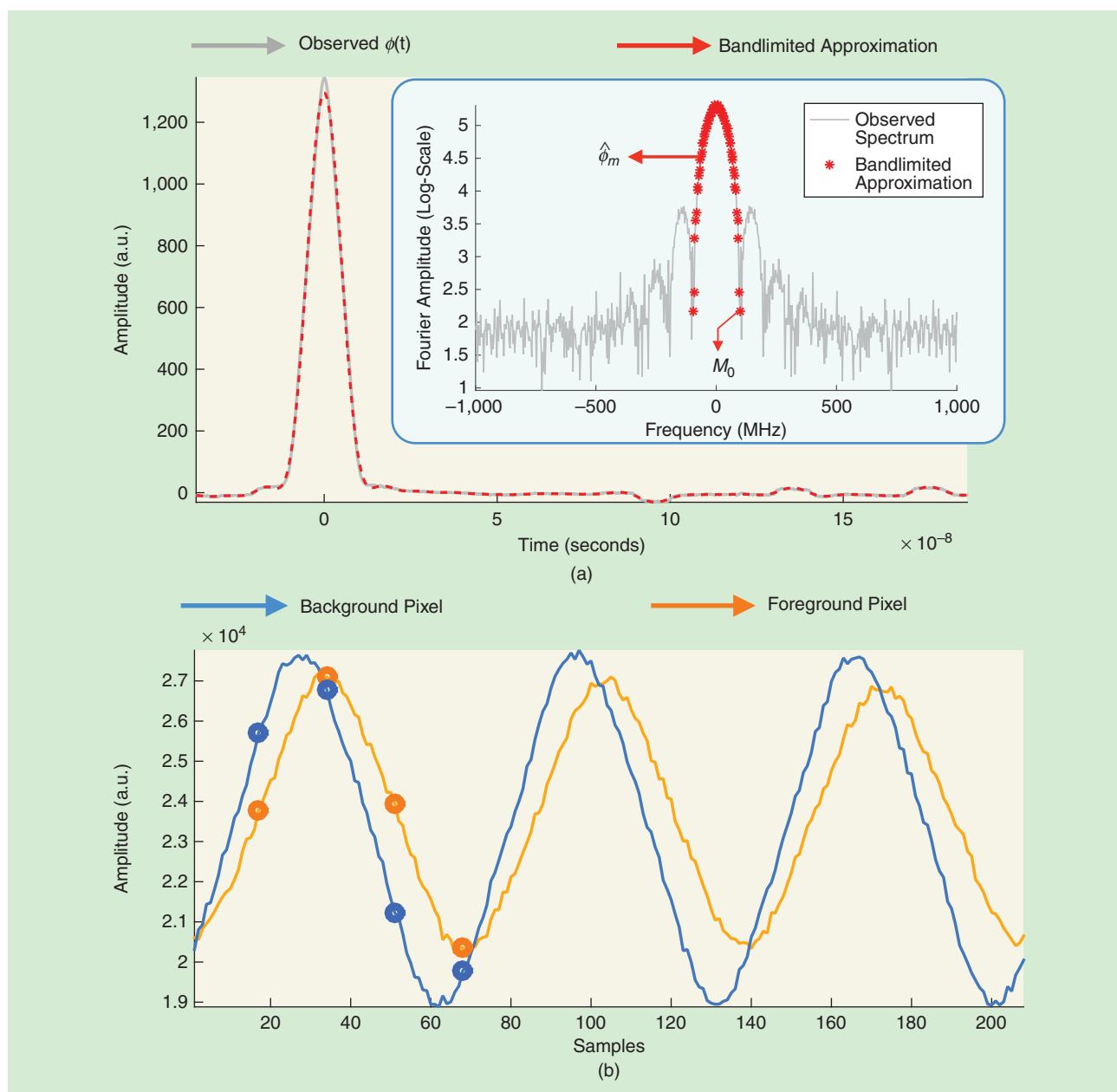
FIGURE 3. Time- and frequency-domain ToF signals. (a) Bandlimited approximation of autocorrelated probing signal ($\phi = p * \bar{p}$) in a time-domain ToF setup. The low-pass property is evident from its Fourier spectrum. This is a result of an experiment with $\Delta = 310$ ns and $M_0 = 30$. (b) Samples of measured signal in (4) that were used to create the depth map in Figure 1. We plot 208 samples for an experiment with $f = \omega/2\pi = 50$ MHz. With reference to Figure 1, the foreground and background pixels map to 2.2 and 2.5 m, respectively.

in (2), the reflected signal reads $r(t) = \Gamma p(t - t_0)$, $t_0 = 2d_0/c$. Again, the lock-in sensor acts as an electronic homodyne detector such that (1) holds, and measurements results in [25],

$$m(t) \overset{(S2)}{=} \Gamma_0 \left(1 + \frac{p_0^2}{2} \cos(\omega t + \theta_\omega) \right), \quad \theta_\omega \in [0, 2\pi), \qquad (4)$$

where $\theta_\omega = \omega t_0 = 2d_0 \omega/c$ is the frequency dependent phase. The ToF sensor records discrete measurements of form, $m_k = m(kT)$ with $T = \pi/2\omega$ and uses a phase estimation algorithm commonly known as the "Four Bucket Method" [22], [21] to estimate parameters d_0 and Γ_0. For a given modulation frequency, this method works with four discrete measurements $\{m_k\}_{k=0}^3$ that are used to form a complex number $z_\omega \in \mathbb{C}$,

$$z_\omega = (m_0 - m_2) + \jmath(m_3 - m_1) \text{ with}$$

$$\begin{bmatrix} m_0 & m_1 \\ m_2 & m_3 \end{bmatrix} = \frac{\Gamma_0}{2} \begin{bmatrix} 2 + p_0^2 \cos(\omega t_0) & 2 - p_0^2 \sin(\omega t_0) \\ 2 - p_0^2 \cos(\omega t_0) & 2 + p_0^2 \sin(\omega t_0) \end{bmatrix}. \qquad (5)$$

The scene parameters are then estimated by $\tilde{\Gamma}_0 = |z_\omega|/p_0^2$ and $\tilde{d}_0 = c\angle z_\omega/2\omega$. We use this method to create the depth image in Figure 1. The raw data samples corresponding to the experiment are plotted in Figure 3(b). This completes our discussion on depth imaging with time- and frequency-domain ToF sensors.

Key takeaways of this section
- The ToF imaging pipeline consists of a probing function, scene response function, and an instrument response function.
- For the case of 3-D imaging, the scene response function is the shift-invariant, time-delay operator in (2).
- In TD–ToF, the probing function is a time-localized pulse [cf. Figure 3(a)], and the object's distance from the sensor is encoded as the round-trip time-delay $t_0 = 2d_0/c$.
- In FD–ToF, the probing function is a sinusoidal waveform with known modulation frequency [cf. Figure 3(b)]. The object's distance from the sensor is encoded as a frequency dependent phase.

Landscape of inverse problems in ToF imaging
Having covered the mathematical basics of time- and frequency-domain ToF imaging systems, we will now discuss case studies where signal processing theory may be used in combination with ToF sensors to solve inverse problems.

Optical multipath interference and spectral estimation
Multipath interference (MPI) is a problem that naturally occurs in communications, acoustics, and array signal processing. Consumer-grade FD–ToF sensors are designed to work under the premise that each point in the scene maps to a single pixel on the sensor. Whenever this hypothesis does not hold, the ToF sensor measurements are erroneous due to MPI. This results in corrupted 3-D images. For example, this can happen when imaging through a semitransparent object, such as a glass window.

A consequence of multiple optical paths combining at a given pixel is that the shift-invariant SRF for single depth (2) now takes the form of a K-sparse filter,

$$h_K(t) = \sum_{k=0}^{K-1} \Gamma_k \delta(t - t_k), \quad t_k = 2\frac{d_k}{c}, \qquad (6)$$

where K is the number of optical paths, and $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ are the scene parameters corresponding to each optical path. For simplicity, we assume that $p_0 = 1$ in the definition of the probing function, $p(t) = 1 + p_0 \cos(\omega t)$. The reflected signal in this case reads

$$r_K(t) = \sum_{k=0}^{K-1} \Gamma_k \cos(\omega t - \omega t_k) = \sum_{k=0}^{K-1} \Gamma_k \cos(\omega t - \theta_{k,\omega}) \qquad (7)$$

which indicates an addition of sinusoids with varying phases. Let us resort to complex-valued representation and let z^* denote the complex-conjugate of z. The measurements take the form

$$m_K(t) = \frac{1}{2} e^{\jmath\omega t} \underbrace{\sum_{k=0}^{K-1} \Gamma_k e^{+\jmath\omega t_k}}_{\hat{h}_K^*(\omega)} = \frac{1}{2} e^{\jmath\omega t} \hat{h}_K^*(\omega)$$

$$\text{with } \underbrace{\hat{h}_K(\omega) = \sum_{k=0}^{K-1} \Gamma_k e^{-\jmath\omega t_k}}_{\text{Fourier Transform}}, \qquad (8)$$

where $\hat{h}_K(\omega)$ is the Fourier transform of the shift-invariant SRF in (6)—a sum of K complex exponentials or phasors. Since the probing function is a sinusoid—an eigenfunction of a linear system—the measurements amount to observing Fourier transform at modulation frequency ω.

Whenever $K = 1$—the case of single-depth 3-D imaging with no MPI—the unknown constant $\hat{h}_K(\omega)$ is directly estimated by implementing the four-bucket method that maps $\{m_K(\pi k/2\omega)\}_{k=0}^3 \to z_\omega = \hat{h}_K^*(\omega)$. However, in the presence of MPI, the FD–ToF sensor estimates the scalar $z_\omega = \hat{h}_K^*(\omega)$ rather than $\{\Gamma_k, d_k\}_{k=0}^{K-1}$—the pixel-wise scene parameters. For $K = 2$, we plot the SRF in time-domain, frequency-domain, and phasor-domain in Figure 4(a). The fact that $z_\omega = \hat{h}_K^*(\omega)$ is an argument of modulation frequency motivates the necessity for frequency diversity [25]. Given N harmonic measurements of form $\{z_{n\omega_0}\}_{n=0}^{N-1}$, the task of estimating $2K$ unknowns $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ is a classical problem in signal processing that is studied under the theme of spectral estimation theory [26]. Due to the pervasiveness and wide applicability of this problem, a self-contained review of Prony's method that seeks to estimate $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ given $\{z_{n\omega_0}\}_{n=0}^{N-1}$ is provided in "Spectral Estimation (Prony's Method)." This approach may be used for the correction of MPI in ToF sensors. In Figure 4(b), we show multifrequency data acquired using a ToF sensor with $K = 3$. Our demonstration, which is based on the matrix pencil method [26], shows the constituent sinusoidal components. Other methods such as the orthogonal matching pursuit [25], EPIRIT/MUSIC [26], or atomic norm thresholding [34] may also be used.

While MPI suggests that interfering optical paths are a nuisance, it may be exploited for 3-D scanning and imaging of translucent objects. We demonstrate multiple depth imaging
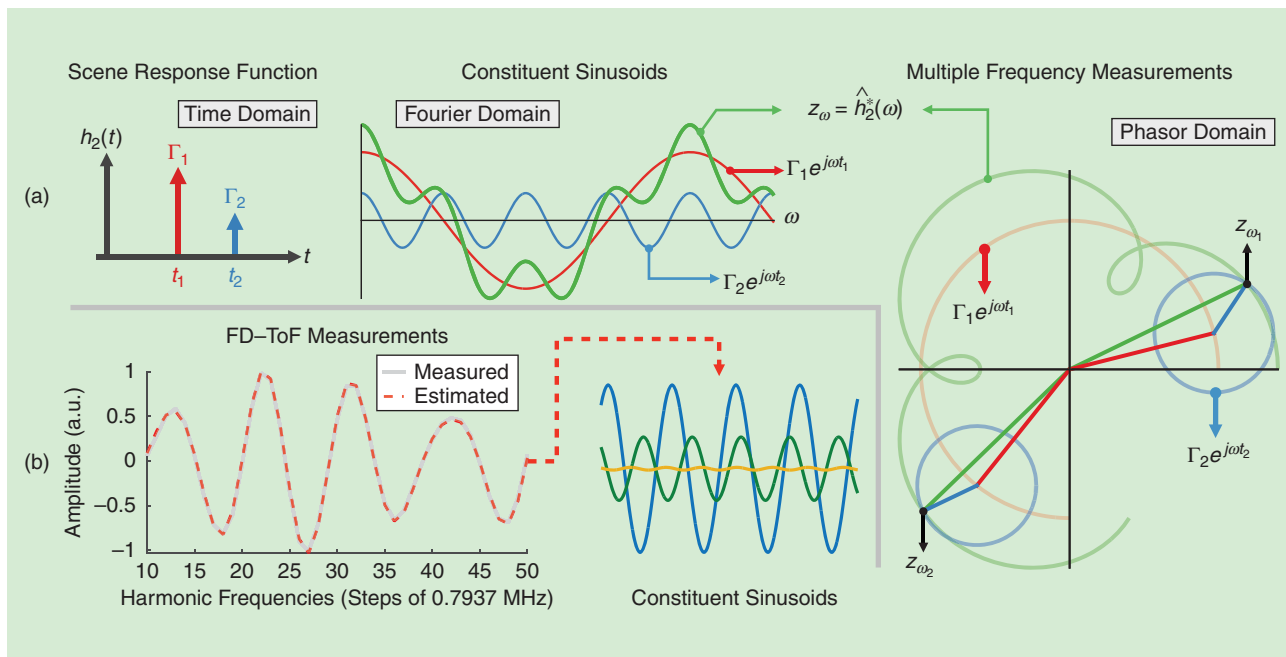
FIGURE 4. Multipath interference in ToF sensors. (a) The scene response function for $K = 2$ in time-domain [cf. (6)], frequency-domain, and complex phasor-domain. At a given modulation frequency ω, the ToF sensor measures the complex number z_ω. However, the correct values to be estimated are $\{\Gamma_k, d_k\}_{k=0}^{K-1}$. This model mismatch results in erroneous depth reconstruction. (b) Experimentally measured multifrequency data and its decomposition into sinusoids via spectral estimation.

capability of ToF sensors in Figure 5, where up to $K = 3$ optical paths interfere at the sensor. As seen in Figure 5, the resulting 3-D images for various modulation frequencies are erroneous. Although Prony's method for spectrum estimation is highly unstable in the presence of noise and model mismatch, oversampling—that is, using much more than $2K$ measurements—is a reasonable solution to counteract perturbations in data. Again, we use the matrix pencil method to estimate the multiple depth related scene parameters. With 46 multiple frequency measurements [cf. Figure 4(b)], the estimation procedure shows that it is possible to reconstruct objects at multiple depths even in presence of MPI. The recovered amplitude and phase images, $\{\Gamma_k\}_{k=0}^2$ and $\{t_k\}_{k=0}^2$, respectively, are shown in Figure 5.

Transient imaging and sampling sparse signals

Understanding light propagation through physical medium has interesting theoretical and practical consequences. For example, a recent demonstration by Velten and coworkers [10] showed that the information contained in multiple echoes of light can be used for non-line-of-sight imaging. From a signal processing perspective, the properties of light/wave propagation can be best understood as a time-dependent transfer function of the scene. The three main ingredients of time-resolved light transport are direct reflections, inter-reflections and subsurface scattering:

$$\underbrace{\Gamma_0 \delta(t - t_0)}_{\text{Direct Reflection}} \quad \underbrace{\sum_{k=1}^K \Gamma_k \delta(t - t_k)}_{\text{Interreflections}} \quad \underbrace{\Gamma_s \delta(t - t_s) * e^{-\mu t} u(t)}_{\text{Subsurface Scattering}},$$

where $\{\Gamma_k, t_k\}_{k=0}^{K-1}$ and $\{\Gamma_s, t_s\}$ are the usual multipath components, μ is the absorption coefficient, and $u(t)$ is the Heaviside function. If it were possible to create a highly time-localized probing function of form $p \sim \delta$, identification of scene related parameters, $\{\{\Gamma_k, t_k\}_{k=0}^{K-1}, \Gamma_s, t_s, \mu\}$ would be a relatively easy task. However, the time-resolution of probing functions programmed on the state-of-the-art ToF sensors is, at best, about 11 ns [cf. Figure 3(a)]. This is orders of magnitude longer than the sophisticated apparatus used by Velten et al. [10], which attains a pulse width of approximately 50 ps. That said, TD–ToF sensors may still be used for imaging transient phenomenon linked with direct and indirect inter-reflections. In fact, this problem is closely tied with the problem of sampling of continuous-time sparse signals [27].

The SRF for the transient imaging problem with direct and indirect components of the light propagation is a continuous-time K sparse signal,

$$h_K(t) \overset{(6)}{=} \underbrace{\Gamma_0 \delta(t - t_0)}_{\text{Direct Reflection}} + \underbrace{\sum_{k=1}^K \Gamma_k \delta(t - t_k)}_{\text{Indirect Reflections}}.$$

With the IRF defined in (1), the ToF sensor measurements are $m(t) = (r \otimes p)(t) = (p * \bar{p} * \bar{h}_K)(t)$ (cf. "ToF Image Formation Model."). With the commutativity property of the convolutions, the sampled TD–ToF measurements may be written as $m_n = (\phi * \bar{h}_K)(t)|_{t=nT}, n = 0, \ldots, N-1$, where $\phi(t) = (p * \bar{p})(t)$. Since the probing function admits a bandlimited approximation [see (3) and Figure 3(a)], the time-reversed TD–ToF measurements may be reinterpreted as samples of a
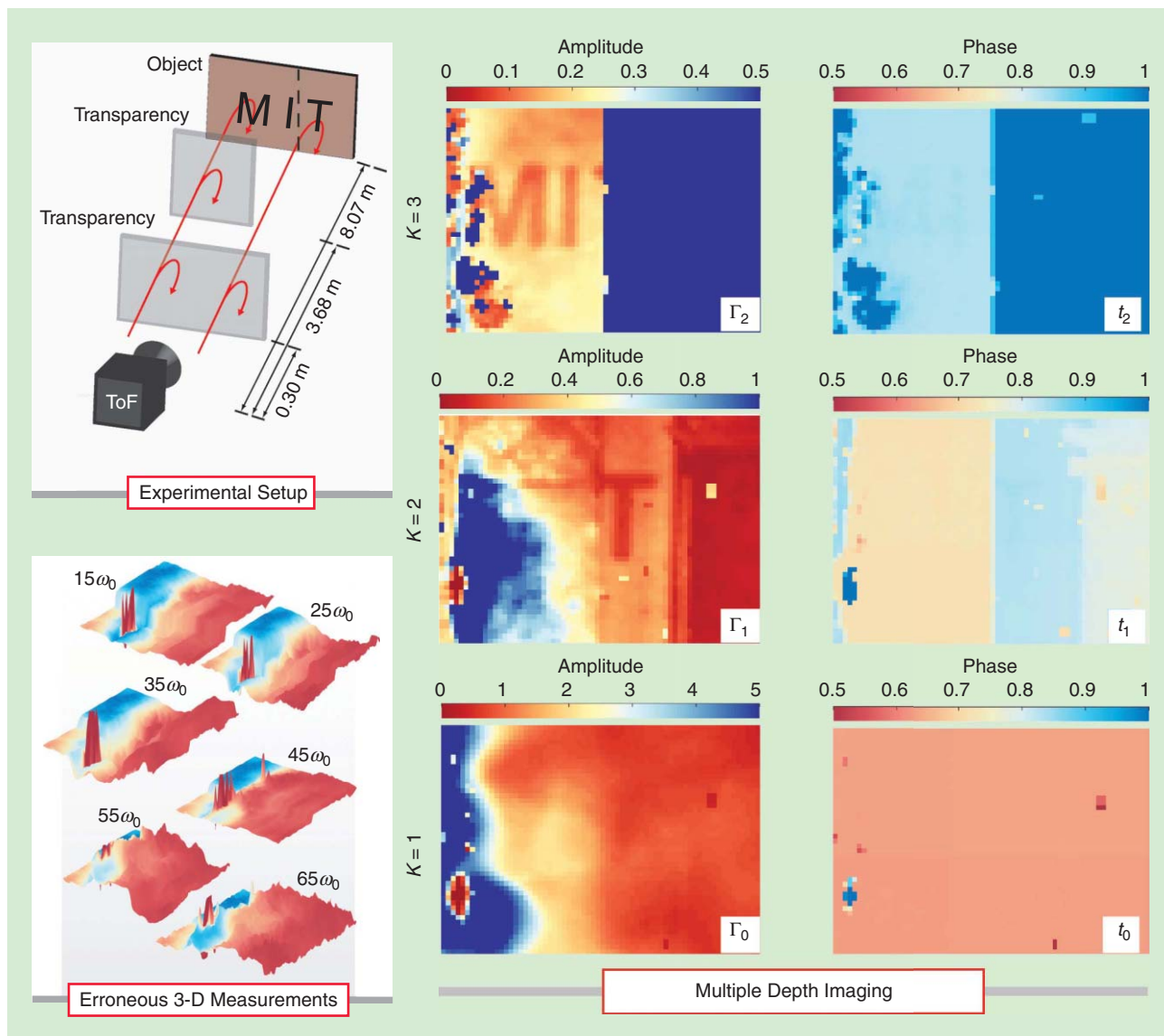
FIGURE 5. Multidepth imaging with ToF sensors. The experimental setup describes a scene with up to $K = 3$ optical paths arising from three different depths. With $f_0 = \omega_0/2\pi = 0.8$ MHz, the ToF sensor captures raw 3-D images using the four-bucket method described in the section "Frequency-Domain 3-D Imaging (FD–ToF)." The resulting 3-D maps are erroneous. We show 3-D images for modulation frequencies $15\omega_0$–$65\omega_0$ in steps of $10\omega_0$. We estimate $\{\Gamma_k, t_k\}_{k=0}^2$ using the spectrum estimation method implemented via the matrix pencil algorithm. The results are consistent with the experimental setup. For example, the letters "MI" and "T" associated with Γ_1 and Γ_2 correspond to their respective depths.

sparse signal that has been low-pass filtered with a calibrated sampling kernel $\phi(t)$. Furthermore, the bandlimited approximation property allows us to write,

$$\bar{m}_n = (\phi * h_K)(t)|_{t=nT} \approx \sum_{\ell=-M_0}^{\ell=+M_0} \hat{\phi}_\ell \hat{h}_K(\ell\omega_0) e^{j\ell\omega_0 nT},$$
$$n = 0, \ldots, N-1,$$

where $\hat{\phi}_\ell = |\hat{p}_\ell|^2$, since the sampling kernel ϕ is obtained by autocorrelating the probing function, and $\hat{h}_K(\ell\omega_0)$ is the sampled Fourier transform of the SRF. Note that the discrete Fourier transform of the sampled measurements $\{\bar{m}_n\}_{n=0}^{N-1}$ results in the $2M_0 + 1$ point-wise samples of the weighted Fourier transform of the SRF, $\{\hat{\phi}_\ell \hat{h}_K(\ell\omega_0)\}_{|\ell| \le M_0}$. Hence, given

sampled measurements we can deconvolve the SRF in Fourier domain, which is a sum of K complex exponential functions. The only requirement being that $\hat{\phi}_\ell$ should not vanish in the interval $-M_0 \le \ell \le M_0$. In a noiseless setting, upon deconvolving the SRF $\{\hat{h}_K(\ell\omega_0)\}_{|\ell| \le M_0}$, we can use spectral estimation [cf. "Spectral Estimation (Prony's Method)"] to estimate the parameters $\{\Gamma_k, t_k\}_{k=0}^{K-1}$, provided that $M_0 \ge K$. Several approaches have been proposed in literature to stabilize the spectral estimation method in presence of noise [26]. As before, we use the matrix pencil method for our experimental demonstrations.

Figure 6 illustrates the reinterpretation of the transient imaging problem as a sparse sampling problem. Time-domain measurements are acquired for the case $K = 2$ and $K = 3$, respectively,

Spectral Estimation (Prony's Method)

Given $2K$ contiguous measurements $z_{n\omega_0} = \sum_{k=0}^{K-1} \Gamma_k e^{jn\omega_0 t_k}$, $n \in [m, 2K + m]$, Prony's method aims to estimate unknowns $\{\Gamma_k, d_k\}_{k=0}^{K-1}$ from a system of equations that is linear in reflection coefficients Γ_k but nonlinear in depths $t_k = 2d_k/c$. The solution to this problem relies on an underlying annihilation equation. Let us define a polynomial of degree K,

$$Q(z) = \prod_{k=1}^{K} (z - e^{j\omega_0 t_k}) = z^K \sum_{m=0}^{K} q_m z^{-m}$$

whose roots are $\{e^{j\omega_0 t_k}\}_{k=0}^{K-1}$. Observe that filtering sequence $\{q_m\}_{m=0}^{K}$ with measurements $\{z_{m\omega_0}\}_{m=0}^{K-1}$ results in nullity,

$$\begin{aligned}(q_m * z_{m\omega_0}) &= \sum_{l=0}^{K} q_l z_{\omega_0(m-l)} = \sum_{l=0}^{K} q_l \sum_{k=0}^{K-1} \Gamma_k e^{j(m-l)\omega_0 t_k} \\ &= \sum_{k=0}^{K-1} \Gamma_k \underbrace{\left(\sum_{l=0}^{K} q_l e^{-jl\omega_0 t_k} \right)}_{Q(e^{jm\omega_0 t_k}) = 0} e^{jm\omega_0 t_k} = 0.\end{aligned}$$

Given $\{z_{n\omega_0}\}_{m \le n}^{n < m+2K}$, in vector-matrix notation, the solution to the previous equation is achieved by solving for

$$m \le n < m + 2K,$$

$$\underbrace{\begin{bmatrix} z_{(m+K)\omega_0} & z_{(m+K-1)\omega_0} & \cdots & z_{m\omega_0} \\ z_{(m+K+1)\omega_0} & z_{(m+K)\omega_0} & \cdots & z_{(m+1)\omega_0} \\ \vdots & \vdots & \ddots & \vdots \\ z_{(m+2K-1)\omega_0} & z_{(m+2K-2)\omega_0} & \cdots & z_{(m+K-1)\omega_0} \end{bmatrix}}_{\text{Toeplitz Matrix}} \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_K \end{bmatrix} = 0.$$

With filter $\{q_m\}_{m=0}^{m=K}$ known, we construct $Q(z)$, and its roots lead to the K estimates, $\{t_k\}_{k=0}^{K-1}$. With the depths known, estimating $\{\Gamma_k\}_{k=0}^{K-1}$ is the usual linear least-squares problem since Γ_k's are linearly constrained in the definition of $z_{n\omega_0}$.

with sampling rate $T = 70$ ps. In each case, we recover the K-sparse SRF, h_K in the following steps:

1) Obtain the discrete Fourier transform of sampled measurements $\bar{m}_n$, that is $\hat{m}_\ell^*$.
2) Deconvolve $\hat{m}_\ell^*$ with the sampling kernel $\hat{\phi}_\ell$ to obtain $\hat{h}_K(\ell\omega_0) = \hat{m}_\ell^*/\hat{\phi}_\ell$, $\ell \in [-M_0, M_0]$, $M_0 \ge K$.
3) Use Prony's method [cf. "Spectral Estimation (Prony's Method)"] to estimate the sparse SRF $\{\Gamma_k, t_k\}_{k=0}^{K-1}$ given $\hat{h}_K(\ell\omega_0)$.

We plot the sampled, time-reversed measurements $\bar{m}_n$ together with the real part of $\hat{h}_K(\ell\omega_0)$ in Figure 6. The resulting

SRF parameters $\{\Gamma_k, t_k\}_{k=0}^{K-1}$ are consistent with the experimental setup.

Imaging transient phenomenon may be used for several applications. For example, consider the experimental setup in Figure 7. A diffusive, semitranslucent sheet covers a placard that reads "Time of Flight." A conventional digital camera is not able to image through the diffuse/scattering object. However, in the context of a TD–ToF imaging setup, this scene has a sparse SRF with $K = 2$. In fact, the case of $K = 2$ described in Figure 6 uses measurements directly from the example in Figure 7 under consideration. By applying the sparse signal
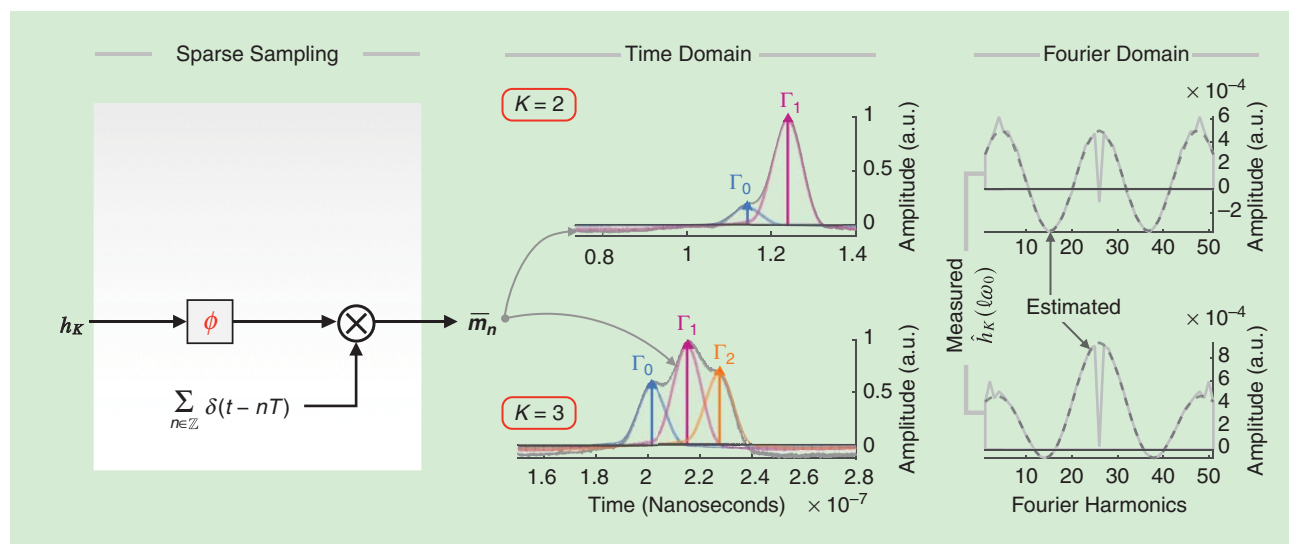

FIGURE 6. Time-domain ToF can be reinterpreted as a sparse sampling problem in the case of transient imaging. The goal is to recover the sparse scene response function given its low-pass-filtered samples with bandwidth $M_0 = 25$. In the context of sampling theory, the low-pass kernel ϕ is an autocorrelation of the probing function. We discuss the recovery of sparse signal for the case $K = 2$ and $K = 3$. We also show the real part of the measured and estimated Fourier transform of the scene response function $\hat{h}_K(\ell\omega_0)$.
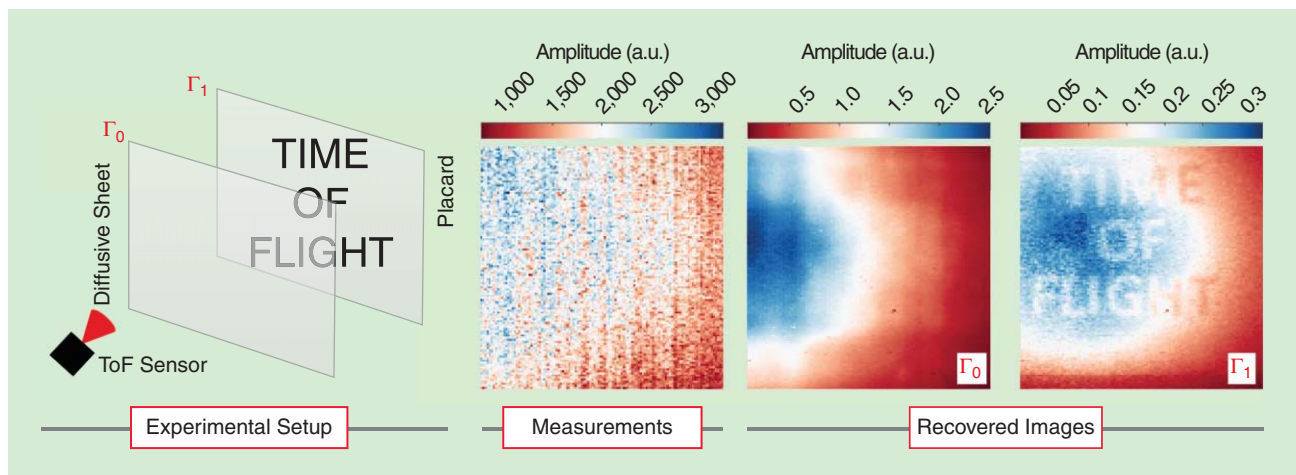
FIGURE 7. The experimental setup for K = 2, TD–ToF measurements and transient image components Γ_0 and Γ_1.

recovery method discussed in this section, it is possible to estimate the SRF h_2. We discuss the results of this experiment in Figure 7. As can be seen, the amplitude image corresponding to Γ_1 matches the experimental setup, and we are thus able to image through a diffusive medium.

Fluorescence lifetime imaging and system identification

Fluorescence lifetime imaging (FLI) is a problem of interest across several disciplines of science and engineering, with the notable example of fluorescence lifetime microscopy. This imaging modality finds applications in a number of fundamental problems including DNA sequencing, tumor detection, fluorescence tomography, and high-resolution microscopy. Like the ToF sensors, FLI can be categorized into time-domain and frequency-domain modes. Time-domain FLI (TD-FLI) utilizes an impulse-like excitation pulse that is used to probe a fluorescent sample. The resulting time-resolved, reflected signal is then used to calculate lifetimes. Given a fluorescent sample with lifetime parameter λ_0, this interaction is modeled as $m(t) = \delta(t) * \rho_0 e^{-\frac{t}{\lambda_0}} u(t)$, where m is the measured signal, δ is the Dirac distribution, and u is the usual Heaviside function. On the other hand, in the case of frequency-domain FLI (FD-FLI), the sample is excited with a sinusoidal probing function of form $\cos(\omega t)$. The phase of the reflected signal ($\propto \cos(\omega t - \tan^{-1}(\omega \lambda_0))$) encodes the lifetime parameter λ_0. Let $h(t)$ be a linear, shift-invariant system. Recall that

$$r(t) = \cos(\omega t) * h(t) = |\hat{h}(\omega)| \cos(\omega t + \angle \hat{h}(\omega)), \quad (9)$$

where $\hat{h}$ is the Fourier transform of h. More precisely, with $h(t) = \rho_0 e^{-\frac{t}{\lambda_0}} u(t)$, its Fourier transform is $\hat{h}(\omega) = (\rho_0 \lambda_0 / 1 + \jmath \omega \lambda_0)$, then $\angle \hat{h}(\omega) = -\tan^{-1}(\omega \lambda_0)$. Consequently, the FD-FLI problem boils down to phase estimation. This is a nonlinear problem and several methods have been proposed in literature. Noting that the measurements (9) are proportional to the so-called modulation depth, that is, $|\hat{h}(\omega)|$, the intensity of the reflected signal may also be used for lifetime estimation due to its dependence on λ_0.

In either case, time- or frequency-domain FLI, the equipment is prohibitively expensive due to strict system constraints and precise electro-optical components. Furthermore, measurements must be calibrated to account for path delays attributed to sample's placement relative to the imaging sensor. Alternatively, ToF sensors are a consumer-grade commodity and are available at an affordable price. In the context of ToF imaging, the SRF for FLI is a shift-invariant function that takes form of [21],

$$h_{\text{FLI}}(t) = \underbrace{\Gamma_0 \delta(t - t_0)}_{\text{Direct Reflection}} + \underbrace{\rho_0 e^{-\frac{t - t_0}{\lambda_0}} u(t - t_0)}_{\text{Fluorescent Sample}}, \quad t_0 = 2\frac{d_0}{c}, \quad (10)$$

where d_0 is the distance of the sample from the sensor—a quantity that is often calibrated in FLI setup, which we assume to be an unknown. In parallel to the frequency-domain FLI, the probing function in the case of FD–ToF is defined by $p(t) = 1 + p_0 \cos(\omega t)$. With the IRF defined in (1), the measurements read

$$m(t) \overset{(4)}{=} |\hat{h}_{\text{FLI}}(0)| + |\hat{h}_{\text{FLI}}(\omega)|\frac{p_0^2}{2}\cos(\omega t - \angle \hat{h}_{\text{FLI}}(\omega)), \quad (11)$$

where $\hat{h}_{\text{FLI}}(\omega)$ is the Fourier transform of the SRF. As described in the section "Frequency-Domain 3-D Imaging (FD–ToF)," the ToF sensor uses the four-bucket method (5) to record the sampled SRF spectrum at modulation frequency ω,

$$z_\omega = \rho_0^2 \hat{h}_{\text{FLI}}^*(\omega) \quad \text{where} \quad \hat{h}_{\text{FLI}}(\omega) = \left(\Gamma_0 + \frac{\rho_0 \lambda_0}{1 + \jmath \omega \lambda_0}\right)e^{-\jmath \omega t_0}. \quad (12)$$

We show the amplitude and phase images, $|z_\omega|$ and $\angle z_\omega$, respectively in Figure 8(a) with $f = \omega/2\pi = 20$ MHz. As shown in [21], the reflection coefficients are dependent on the optical frequency of light and the direct component may be filtered ($\Gamma_0 = 0$). Since the ToF sensor samples the Fourier spectrum of the SRF at each modulation frequency, the task of estimating $\{t_0, \lambda_0\}$ can only be accomplished if multiple frequency observations $z_{n\omega_0} = p_0^2 \hat{h}_{\text{FLI}}(n\omega_0), n = 0, \ldots, N-1$ are
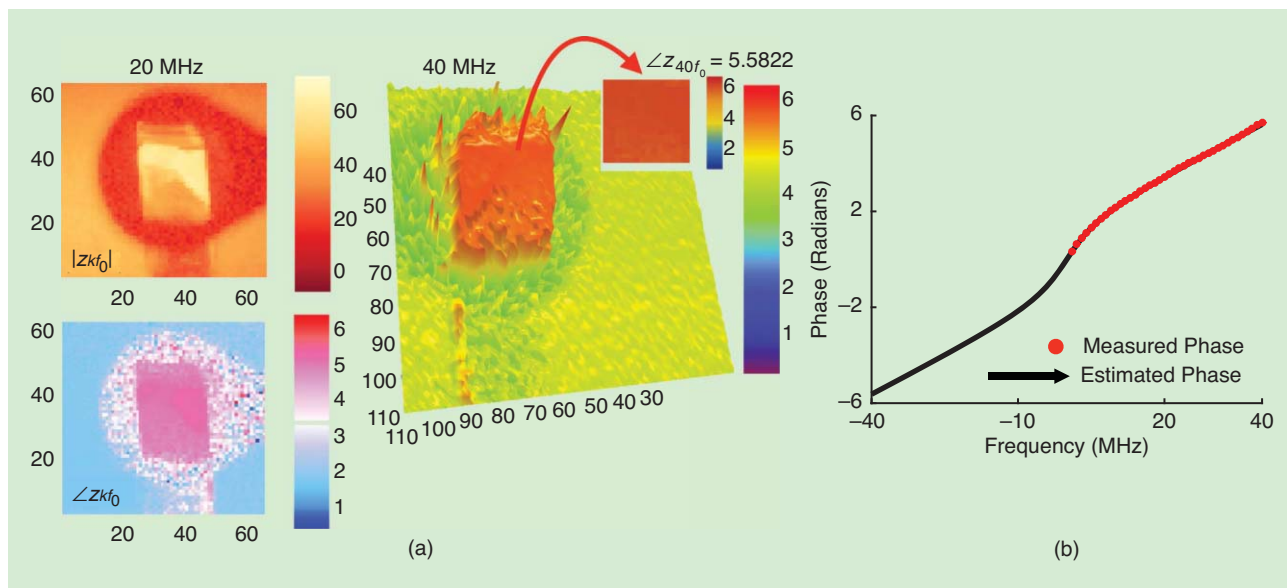
FIGURE 8. Frequency-domain ToF imaging. (a) Multifrequency measurements of a scene containing a fluorescent sample with a lifetime $\tau = 32$ ns. We show amplitude image $|z_{20f_0}|$ (in dB) as well as the phase image $\angle z_{20f_0}$ (in radians) with $f_0 = 1$ MHz. We also show a 3-D plot of the phase image. This shows the effect of multipath interference. The phase at the background pixel is 4.1625 rads, which relates to a depth of 2.4842 m, while the actual depth is 2.5 m. At this depth, the phase recorded as the sample location is 5.5822. The extra 3.3316 is attributed to the fluorescent lifetime decay phenomenon $\tan^{-1}(\omega\lambda_0)$. (b) At the sample location, we plot phase $\angle z_{kf_0} = -\angle \hat{h}(kf_0), k = 1, \ldots, 40$ (12) recorded by the ToF sensor. We also plot the phase estimated by our method.

available. With $t_0 = 0$, estimation of λ_0 is a classical problem in system identification [28]. The presence of nonzero t_0 motivates development of new algorithms. In contrast to [21], where the authors use nonlinear least squares fitting, here, we develop a closed form solution to the problem. In view of (12), let us set $y_n = z_{n\omega_0}/p_0^2$ and $\omega_n = n\omega_0$, and we have,

$$\frac{y_{n+1}}{y_n} = \frac{1 + \jmath\omega_n\lambda_0}{1 + \jmath\omega_{n+1}\lambda_0} e^{-\jmath\omega_0 t_0}$$

$$\Updownarrow$$

$$y_{n+1}(1 + \jmath\omega_{n+1}\lambda_0) = y_n(1 + \jmath\omega_n\lambda_0)e^{-\jmath\omega_0 t_0},$$

which is a linear system of equations in λ_0 and $e^{-\jmath\omega_0 t_0}$ and can be solved with any four contiguous values such that $\ell \leq n < \ell + 4, \ell \in \mathbb{Z}$. We solve this linear system of equations for $\{y_n\}_{n=2}^{n=20}$ since the modulation frequencies in range $f_0 = 2 - 20$ MHz are highly stable. While system stability and signal integrity is one part that affects the SNR, variation of sample spread on the slide may also lead to weak emission. For this purpose, we use a confidence threshold for pixel amplitudes.

We compare our system identification approach with previously used nonlinear least squares fitting [21], which uses $N = 40$ samples and favorable initialization conditions to obtain

(nanoseconds)	$\tilde{\lambda}_0$	32.16	32.32	31.46
	$\tilde{\lambda}_{0,\text{NNLS}}$	30.33	30.79	31.51
(meters)	$\tilde{d}$	2.54	2.49	2.54
	$\tilde{d}_{\text{NNLS}}$	2.494	2.496	2.496.

For comparable distance estimates, the identification method provides a more robust estimate of lifetimes. In comparison

to NNLS fitting (based on MATLAB'S curve-fit toolbox), our relatively modest and noniterative method is orders of magnitude faster ($\sim 10^3$). We close this subsection by highlighting that the phase/depth images in ToF sensors may be used for encoding interesting physical/material properties.

Key takeaways of this section

- Multipath interference (MPI) in FD–ToF mode can be reinterpreted as a spectrum estimation problem. To recover MPI components, one must acquire multiple frequency measurements. Our demonstrations were based on the matrix pencil method [26] but any other method is an option. For example, in [25], the authors use orthogonal matching pursuit.
- Transient imaging in TD–ToF mode can be recast as a sparse sampling problem [27], where the sampling kernel ϕ is the autocorrelation of the probing function. Whenever the probing function admits a bandlimited approximation, one may use spectral estimation to recover the transient image components.
- We demonstrated that fluorescence lifetime imaging can be performed with ToF sensors. We used system identification methods to estimate an unknown, parametric transfer function linked with lifetime imaging.

Conclusions and future directions

We hope that we have convinced you that ToF sensors are more than just depth sensors! In what follows, we present our concluding remarks with hints on possible future research directions.

ToF imaging pipeline

We started with an image formation model that allows for studying different ToF modalities under one common

framework. Almost all ToF systems can be characterized by the model parameters $\{p, h, \varphi\}$. Depending on the problem at hand, the role of $\{p, h, \varphi\}$ and the associated algorithms needs to be adapted. For example, the SRF discussed in context of MPI appears naturally in other problems, such as ultrasound tomography [29], single photo imaging [31] and light detection and ranging (LiDAR) [30]. However, the probing function and the IRF are very different for each case. In particular, consider the case of LiDAR. The probing function is modeled as $p = \delta$ and the IRF is a parametric function of form $\varphi(t) = \alpha e^{(a_k - T_0)t + b_k}$, where $\{a_k, b_k\}_{k=1}^4$ take four different values with continuous transitions, depending whether $t \in \{I_k\}_{k=1}^4$, where I_k is an instrument or sensor dependent quantity. This gives rise to a new form of sampling kernel $\phi = \varphi$ [32], as opposed to $\phi = p * \bar{p}$ (TD–ToF case). Hence, we believe that by systematically studying the role of $\{p, h, \varphi\}$ across various ToF problems—optical and nonoptical—better insights may developed.

Another interesting direction may be to consider the case when the SRF is modeled by a differential equation. For example, in the case of fluorescence lifetime imaging, the associated differential equation is $\mathcal{L}_\lambda = \partial_t + (1/\lambda)$, and the resulting reflected signal is the solution to $\Gamma \mathcal{L}_\lambda[r(t)] = p(t), t \geq 0$. The SRF in this case is the Green's function. Similar ideas may be used to develop algorithms for imaging through scattering/diffusive media where $\mathcal{L}$ is some differential equation that models diffusion. The parameters of $\mathcal{L}$ encode physical properties such as lifetime or scattering coefficient.

Probing function

Since the probing function is the only available degree-of-freedom in ToF imaging pipeline, it is important to understand what mathematical principles should be used for designing probing functions. Waveform design is a known art in radar and wireless communications. However, such options are rarely considered in optical ToF systems. Maximum length sequences for TD–ToF and sinusoids for the FD–ToF are the de facto examples. On the hand, it may not always be feasible to calibrate the probing function. In that case, it may be worthwhile to use blind deconvolution algorithms for image reconstruction.

Algorithms and fundamental limits

MPI is a significant problem in ToF imaging and a number of papers have attempted to address this issue—both in time- and in frequency domain. However, to date, most of the results remain empirical and rarely discuss any details on performance guarantees. In a recent work [24], we used Cramér–Rao bounds in context of TD–ToF-based multiple depth imaging. For instance, for the TD–ToF case, one may write the probing function as a Fourier series with Fourier coefficients $\{\phi_m\}_{m \in \mathbb{Z}}$. In this case, two optical paths, Δd apart, are resolvable provided that

$$\frac{\Delta d}{T} \geq \frac{c}{4\pi} \frac{1}{\sqrt{\text{PSNR}}} \sqrt{\frac{S}{N}},$$

where $T = 2\pi/\omega_0$ is the fundamental time period, PSNR is the peak-signal-to–noise–ratio, $S^{-1} = \sum_{m \in \mathbb{Z}} m^2 |\phi_m|^2$ and N

is the number of measurements. For parametric SRFs, we believe that more such efforts could lead to hints on interesting applications of ToF sensors and motivate new problems in sampling theory. This could be the key to questions such as: when can two lifetimes in fluorescence lifetime imaging setup be super-resolved?

Modeling nonideal reflections

In our experience, the SRF of form $\Gamma_k \delta(t - t_k)$ [c. f. (6)] only approximately models a reflection. In a practical setting, $\Gamma_k \psi_k(t - t_k)$ may serve as a good starting point for modeling reflections. Here, ψ_k is a filter that models the interaction of the probing function with the material property or accounts for distortion, system nonlinearities and dispersive media. In seismic engineering, terahertz spectroscopy, and ultrasound systems, this behavior is much more pronounced as material properties play an important role when the probing function undergoes a reflection.

FD-ToF

This is an interesting mode of operation since most consumer ToF systems are based on FD-ToF, which uses phase estimation. As seen in (4), if $\theta_\omega > 2\pi$, the depth estimates suffer with ambiguity or phase-wrapping problem. Previous solutions use coprime frequencies [33], however, there is room for improvement. For example, in theory, phase is a linear function of frequency, $\theta_\omega = 2\omega d/c$. This is not the case in practice and leads to erroneous depth estimates with multiple frequency measurements. Hence, a desirable phase estimation algorithm should jointly correct for any distortions and phase-wrapping.

Calling $m(t, \omega) \stackrel{(4)}{=} \Gamma_0(1 + (p_0^2/2)\cos(\omega t + \theta_\omega))$, θ_ω is estimated by sampling in time-domain, that is, $m_k = m(kT, \omega_0)$ given a fixed modulation frequency ω_0. Alternatively, one may use multifrequency sampling using $m_k = m(t_0, k\omega_0)$ for the estimation of θ_ω. This gives rise to a broader question of when can time-frequency sampling be used, that, $m_{\ell,k} = m(\ell t_0, k\omega_0)$ in context of solving inverse-problems linked with ToF imaging. Specifically, when $\omega = t$ for depth imaging, the problem boils down to parameter estimation of chirp signals. Finally, in view of (8), phase retrieval algorithms may be designed when only intensity measurements are available, that is, $|m_k|^2 = |m(t_0, k\omega_0)|^2$.

Sensor design for higher modulation frequencies

Most consumer-grade ToF sensors are based on continuous wave probing functions. Currently, such sensors work with high fidelity up to a modulation frequency of about 80 MHz. We believe that much of the interesting physical phenomenon may only be observed as higher frequencies. For example, higher modulation frequencies will certainly enhance depth resolution and MPI correction capabilities. In context of fluorescence lifetime imaging [21], shorter lifetimes may be resolved with higher modulation frequencies. Similarly, subsurface scattering properties can be studies with streak tubes [10]. This hints that higher modulation frequencies are the

pathway to scattered imaging. Such examples motivate the necessity of hardware or computational imaging solutions that can over come the current technological limits.

Acknowledgments

Authors

Ayush Bhandari (ayush@mit.edu) is a researcher at the Massachusetts Institute of Technology (MIT). Prior to working at MIT, he held various research positions starting with INRIA–Rennes, France, followed by Nanyang Technological University, Singapore, and the Chinese University of Hong Kong, with the most recent appointment at École Polytechnique Fédérale de Lausanne, Switzerland. His research interests include sampling and approximation theory and inverse problems linked with computational imaging. He was recipient of the Best Student Paper Award at the IEEE International Symposium on Signal Processing and Its Applications (2007) and the Best Paper Award at the IEEE Systems and Information Engineering Design Symposium (2007).

Ramesh Raskar (raskar@mit.edu) heads the Camera Culture Research Group at the Massachusetts Institute of Technology Media Lab. His research interests span the fields of computational photography, inverse problems in imaging, and human-computer interaction. He has been the recipient of numerous awards including the TR100 Award from *Technology Review*, the Global Indus Technovator Award, Sloan Research Fellowship (2009), and the DARPA Young Faculty Award (2010). Other awards include Marr Prize honorable mention (2009), LAUNCH Health Innovation Award, and the 2011 Vodafone Wireless Innovation Project Award (first place). He holds more than 40 U.S. patents and is currently coauthoring a book on computational photography.

References

[1] O. Cossairt, M. Gupta, and S. K. Nayar, "When does computational imaging improve performance?" *IEEE Trans. Image Processing*, vol. 22, no. 2, pp. 447–458, 2013.

[2] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: Motion deblurring using fluttered shutter," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 795–804, 2006.

[3] S. S. Gorthi, D. Schaak, and E. Schonbrun, "Fluorescence imaging of flowing cells using a temporally coded excitation," *Opt. Express*, vol. 21, no. 4, pp. 5164–5170, 2013.

[4] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annu. Conf. Computer Graphics and Interactive Techniques (ACM SIGGRAPH)*, New York, 2008, p. 369–378.

[5] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, 2005.

[6] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69, 2007.

[7] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[8] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photon.*, vol. 7, no. 9, pp. 739–745, 2013.

[9] R. I. Hartley and P. Sturm, "Triangulation," *Comp. Vis. Image Understand.*, vol. 68, no. 2, pp. 146–157, 1997.

[10] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nat. Commun*, vol. 3, pp. 745, Mar. 2012.

[11] O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nat. Photon.*, vol. 8, no. 10, pp. 784–790, 2014.

[12] M. Laurenzis, J. Klein, E. Bacher, and N. Metzger, Multiple-return single-photon counting of light in flight and sensing of non-line-of-sight objects at shortwave infrared wavelengths," *Opt. Lett.*, vol. 40, no. 20, pp. 4815–4818, 2015.

[13] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, Single-photon sensitive light-in-flight imaging," *Nat. Commun.*, vol. 6, Jan. 2015.

[14] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014.

[15] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 45:1–45:10, 2013.

[16] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar, "Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 167, 2013.

[17] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3-D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 3222–3229.

[18] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," *Opt. Express*, vol. 22, no. 21, pp. 26338–26350, Oct. 2014.

[19] A. Kadambi, A. Bhandari, R. Whyte, A. Dorrington, and R. Raskar, "Demultiplexing illumination via low cost sensing and nanosecond coding," in *Proc. IEEE Int. Conf. Computational Photography*, Santa Clara, CA, 2014, pp. 1–10.

[20] N. Kitsunezaki, E. Adachi, T. Masuda, and J. I. Mizusawa, "KINECT applications for the physical rehabilitation," in *Proc. IEEE Int. Symp. Medical Measurements and Applications*, Gatineau, Quebec, Canada, 2013, pp. 294–299.

[21] A. Bhandari, C. Barsi, and R. Raskar, "Blind and reference-free fluorescence lifetime estimation via consumer time-of-flight sensors," *Optica*, vol. 2, no. 11, pp. 965–973, Nov. 2015.

[22] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, 2011.

[23] B. Büttgen, E. Mechat, M. Hamed-Ali, F. Lustenberger, and P. Seitz, "Pseudonoise optical modulation for real-time 3-d imaging with minimum interference," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 10, pp. 2109–2119, 2007.

[24] A. Bhandari, A. Kadambi, and R. Raskar, "Sparse linear operator identification without sparse regularization? Applications to mixed pixel problem in TOF/range imaging," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Florence, Italy, 2014, pp. 365–369.

[25] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Opt. Lett.*, vol. 39, no. 7, 2014.

[26] Y. Hua, A. Gershman, and Q. Cheng, *High-Resolution and Robust Signal Processing*. Boca Raton, FL: CRC Press, 2003.

[27] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 31–40, 2008.

[28] L. Ljung, *System Identification*. New York: Springer, 1998.

[29] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Processing*, vol. 59, no. 4, pp. 1827–1842, 2011.

[30] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Bayesian analysis of lidar signals with multiple returns," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 29, no. 12, pp. 2170–2180, 2007.

[31] D. Shin, F. Xu, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "Computational multi-depth single-photon imaging," *Optics Express*, vol. 24, no. 3, pp. 1873–1888, 2016.

[32] A. Bhandari, A. M. Wallace, and R. Raskar, "Super-resolved time-of-flight sensing via FRI sampling theory," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 4009–4013.

[33] A. P. P. Jongenelen, D. G. Bailey, A. D. Payne, A. A. Dorrington, and D. A Carnegie, "Analysis of errors in ToF imaging with dual-frequency modulation," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1861–1868, 2011.

[34] G. Tang, N. Bhaskar, and B. Recht, "Near minmax line spectral estimation," *IEEE Trans. Inform. Theory*, vol. 61, no. 1, pp. 499–512, 2015.

SP

Ivo Ihrke, John Restrepo, and Loïs Mignard-Debise

Principles of Light Field Imaging

Briefly revisiting 25 years of research

Light field imaging offers powerful new capabilities through sophisticated digital processing techniques that are tightly merged with unconventional optical designs. This combination of imaging technology and computation necessitates a fundamentally different view of the optical properties of imaging systems and poses new challenges for the traditional signal and image processing domains. In this article, we aim to provide a comprehensive review of the considerations involved and the difficulties encountered in working with light field data.

Digital light field imaging: An overview

As we approach the 25th anniversary of digital light field imaging [1]–[3] and the technology begins to enter the industrial and consumer markets, it is a good time to reflect on the developments and trends in what has become a vibrant interdisciplinary field joining optical imaging, image processing, computer vision, and computer graphics.

The key enabling insight of light field imaging is a reinterpretation of the classic photographic imaging procedure that separates the process of imaging a scene (i.e., scene capture) from the actual realization of an image (i.e., image synthesis)—a reinterpretation that offers new flexibility in terms of postprocessing. The underlying idea is that a digital capture process enables intermediate processing far beyond simple image processing. In fact, our modern cameras are powerful computers that enable the execution of sophisticated algorithms to produce high-quality two-dimensional (2-D) images.

Light field imaging is, however, moving beyond that level by purposefully modifying classical optical designs to enable the capture of high-dimensional data sets that contain rich scene information. The 2-D images presented to the human observer are processed versions of the higher-dimensional data the sensor has acquired and only the computer sees in their raw form. This partial replacement of physics by computation enables the post-capture modification of images on a previously unimaginable scale. Most of us have seen the amazing
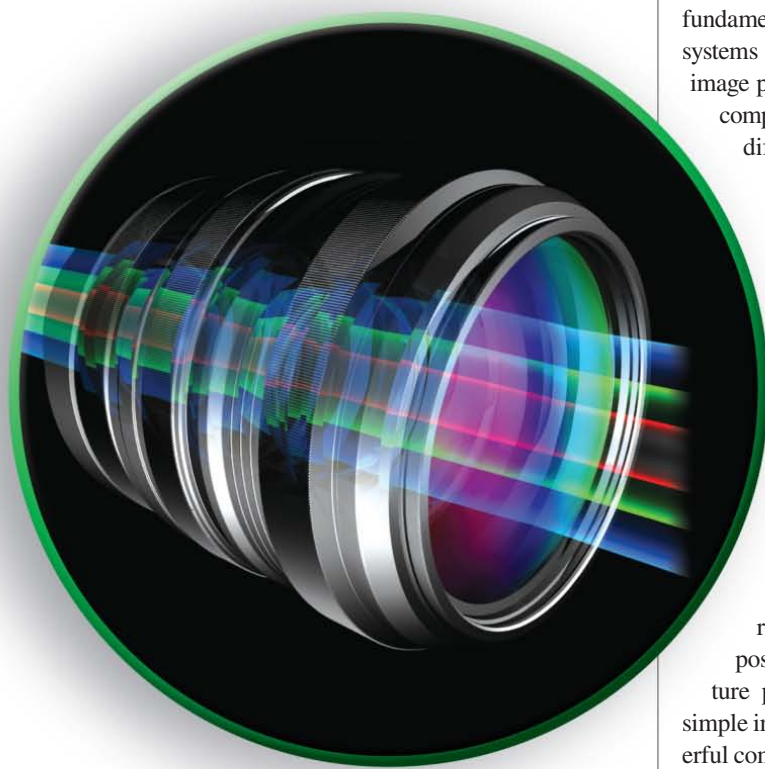
©ISTOCKPHOTO.COM/YAKOBCHUK

features that light field cameras offer: post-capture refocus, change of view point, three-dimensional (3-D) data extraction, change of focal length, focusing through occluders, increasing visibility in bad weather conditions, and improving the robustness of robot navigation, to name just a few.

In optical design terms, light field imaging presents an (as of yet unfinished) revolution. Since Gauss's day, optical designers have been thinking in terms of two conjugate planes, the task of the designer being to optimize a lens system to gather the light originating at a point on the object plane and converge it as well as possible to a point on the image plane. The larger the bundle of rays that can be converged accurately, the more light-efficient the capture process becomes and the higher the achievable optical resolution. The requirement of light-efficient capture introduces focus into the captured images, i.e., only objects within the focal plane appear sharp. Light field imaging does away with most of these concepts, purposefully imaging out-of-focus regions and inherently aiming at capturing the full 3-D content of a scene.

In terms of signal processing, we encounter a high-dimensional sampling problem with nonuniform and nonlinear sample spacing and high-dimensional spatio-directionally varying observation/sampling kernels. The light field data, however, have particular structures that can be exploited for analysis and reconstruction. This results from the fact that scene geometry and reflectance link the information contained in different samples. It also distinguishes the reconstruction problem from a classical signal processing task.

On the software side, we witness the convergence of ideas from image processing, computer vision, and computer graphics. In particular, the classical preprocessing tasks of demosaicking, vignetting compensation, undistortion, and color enhancement are all affected by sampling in four dimensions rather than in two. Additionally, image analysis by means of computer vision techniques becomes an integral part of the imaging process. Depth-extraction and superresolution techniques enhance the data and mitigate the inherent resolution tradeoff introduced by sampling two additional dimensions. A careful system calibration is necessary for good performance. Computer graphics ideas, finally, are needed to synthesize the images ultimately presented to the user.

This article aims to review of the principles of light field imaging and associated processing concepts, while simultaneously illuminating the remaining challenges. The presentation roughly follows the acquisition and processing chain from optical acquisition principles to the final rendered output image. The focus is on single-camera snapshot technologies that are currently seeing a significant commercial interest.

Background

This section, which provides background for the rest of the article, closely follows the development in [2]. An extended discussion at an introductory level can be found, e.g., in [4]. A wider perspective on computational cameras is given in [5] and [6].

Plenoptic function

The theoretical background for light field imaging is the plenoptic function [7], which is a ray-optical concept that assigns a radiance value to rays propagating within a physical space. It considers the usual 3-D space to be penetrated by light that propagates in all directions. In doing so, the light can be blocked, attenuated, or scattered.

However, instead of modeling this complexity as, e.g., computer graphics is doing, the plenoptic function is an unphysical, modelless, purely phenomenological description of the light distribution in the space. To accommodate for all the possible variations of light without referring to an underlying model, it adopts a high-dimensional description: arbitrary radiance values can be assigned at every position of space, for every possible propagation direction, for every wavelength, and for every point in time. This is usually denoted as $l_\lambda(x,y,z,\theta,\phi,\lambda,t)$, where $l_\lambda[W/m^2/sr/nm/s]$ describes spectral radiance per unit time, (x,y,z) is a spatial position, (θ,ϕ) is an incident direction, λ is the wavelength of light, and t is a temporal instance.

The plenoptic function is mostly of conceptual interest. From a physical perspective, the function cannot be an arbitrary seven-dimensional function because, e.g., radiant flux is delivered in quantized units, i.e., photons. Therefore, a time-average must be assumed. Similarly, it is not possible to measure infinitely thin pencils of rays (i.e., perfect directions) or even very detailed spatial light distributions without encountering wave effects. We may, therefore, assume that the measurable function is band-limited and that we are restricted to macroscopic settings where the structures of interest are significantly larger than the wavelength of light.

Light fields

Light fields derive from the plenoptic function by introducing additional constraints:

- They are considered to be static even though video light fields have been explored [8] and are becoming increasingly feasible. An integration over the exposure period removes the temporal dimension of the plenoptic function.
- They are typically considered as being monochromatic, even though the same reasoning is applied to the color channels independently. An integration over the spectral sensitivity of the camera pixels removes the spectral dimension of the plenoptic function.
- Most importantly, the so called "free-space" assumption introduces a correlation between spatial positions. Rays are assumed to propagate through a vacuum without objects, except for those contained in an "inside" region of the space, often called a *scene*. Without a medium and without occluding objects, the radiance is constant along the rays in the "outside" region. This removes one additional dimension from the plenoptic function [2].

A light field is, therefore, a four-dimensional (4-D) function. We may assume the presence of a boundary surface $\mathcal{S}$ separating the space into the inside part (i.e., the space region containing the scene of interest) and the outside part, where the acquisition apparatus is located. The outside is assumed to be empty space. Then, the light field is a scalar-valued function of $\mathcal{S} \times S_+^2$, where S_+^2 is the hemisphere of directions toward the outside. This definition of a light field is also applied to the term *surface*

light field [9] if the surface $\mathcal{S}$ agrees with some object geometry. In this case, the directional component of the function describes the object reflectance convolved with the incident illumination.

Commonly, the additional assumption is made that the surface $\mathcal{S}$ is convex, e.g., by taking the convex hull of the scene. In this case, the rays can be propagated to other surfaces in the outside region without loss of information. Typically, a plane p is used as the domain of (parts of) the light field function. The most popular parameterization of the spatial and directional dimensions of the light field is the two-plane parameterization, which is obtained by propagating a ray from surface $\mathcal{S}$ to the light field plane p (see Figure 1). The parameterization then consists of the intersection position (u, v) of the ray with the light field plane p and its intersection with an additional parallel plane at a unit distance $(\hat{u}, \hat{v})$. The second intersection is usually parameterized as a difference with respect to the (u, v) position and called $(s = \hat{u} - u, t = \hat{v} - v)$. This second set of coordinates measures the direction of the ray.

Phase space

The coordinates obtained in this way can be considered as an abstract space, the so-called "ray phase space" or simply phase space. A point (u, v, s, t) in this space corresponds to a ray in the physical space. It is important to remember that the phase space is always linked to a particular light field plane p. Changing the plane, in general, changes the phase space configuration, which means that a fixed ray will be associated with a different phase space point.

The phase space is interesting for several reasons. First, it allows us to think more abstractly about the light field. Second, a reduction to two dimensions (u, s) is easily illustrated and generalizes well to the full 4-D setting. Third, finite regions of the ray space, in contrast to infinitesimal points, describe ray bundles. The phase space is, therefore, a useful tool for visualizing ray bundles. Finally, an extensive literature exists on phase space optics (see, e.g., [10]) with available extensions to wave optics. The phase space is also a useful tool for comparing different camera designs [11].

The light field can now be thought of as a radiance-valued function defined in the phase space, i.e. $l(u, v, s, t)$, meaning that each ray, parameterized by (u, v, s, t), is assigned a radiance value l. The task of an acquisition system is to sample and reconstruct this function.

Light field sampling

The simplest way to sample the light field function is by placing a pinhole aperture into the light field plane p. Were the pinhole infinitesimal, ray optics a decent model of reality, and light considerations negligible, we would observe one column of the light field function at a plane a unit distance from the light field plane p. In the following, we will refer to that plane as the *sensor plane* q. Associating a directional sample spacing of Δs and shifting the pinhole by amounts of Δu enable a sampling of the function, as shown in Figure 2.

A slightly more realistic model is that the directional variation s is acquired by finite-sized pixels with a width equivalent to the directional sample spacing Δs. This introduces a directional sampling kernel that, in the phase space, can be
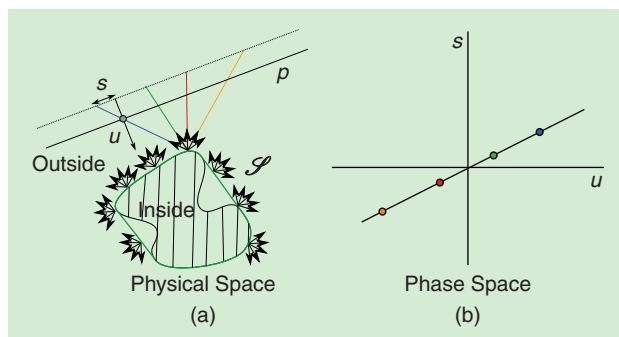
FIGURE 1. A basic description of light field. (a) The "inside" region contains the scene of interest, while the "outside" region is empty space and does not affect light propagation. The light field is a function assigning a radiance value to each of the rays exiting through the boundary surface \mathscrS . (b) A phase space illustration of the colored rays. A point in phase space determines a set of ray parameters (u, s) and, therefore, corresponds to a ray. The phase space is associated with the plane p. Because the four rays indicated in the subfigure in (a) converge to a point, the corresponding phase space points lie on a line.
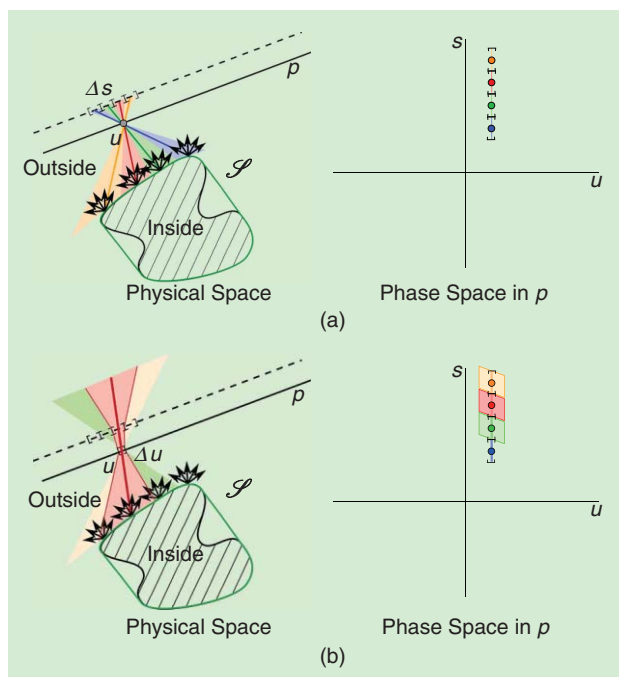
FIGURE 2. The finite sampling of a light field with real hardware. (a) Assuming a sensor placed at the dashed plane and an infinitesimal pinhole results in a discretization and averaging of only the directional light component. In phase space, this constitutes a row of vertical segments. (b) A more realistic scenario uses a finite-sized pinhole, resulting in ray bundles integrated by the sensor's pixels. In conjunction, pixels and pinholes define a two-aperture model. In the phase space, the ray bundle passed by two apertures is represented by a rhomb.

interpreted as a vertical segment, as in Figure 2(a). Of course, the pinhole has a finite dimension Δu, as well. The pinhole/pixel combination, therefore, passes a bundle of rays, as indicated in Figure 2(b). The phase space representation of the ray bundle passing this pinhole/pixel pair is a sheared rectangle, as shown on the right in Figure 2(b). It should be noted that the pinhole size and the pinhole sample spacing, as well as the
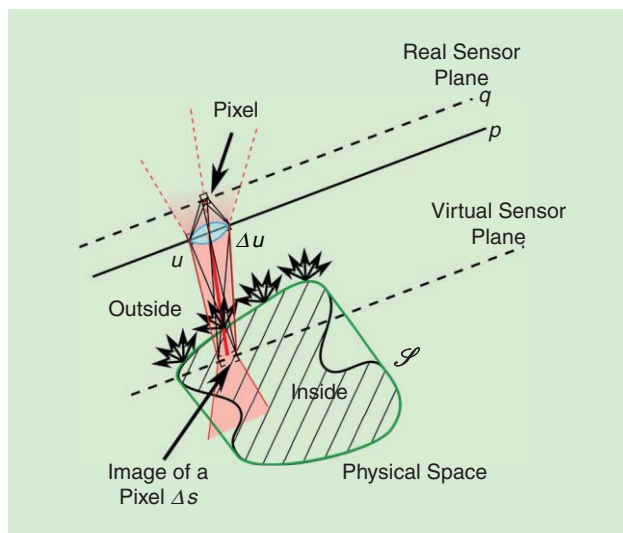
FIGURE 3. Light field imaging with a moving standard camera. Sensor pixels in the sensor plane q are mapped outside the camera and inside the world space. The camera lens and the image of the pixel constitute a two-aperture pair, i.e., a unique phase space region. The color gradient in the ray bundle indicates that the rays are considered to be virtual in the camera's image space. In reality, the rays refract and are converged onto the indicated pixel. In the world space, the ray bundle represents those rays integrated by the pixel. The sensor has more such pixels (not shown in the figure). These additional pixels effectively constitute a moving aperture in the plane of the virtual sensor position.

pixel size and the pixel spacing, may not be correlated in real applications, with corresponding implications for aliasing or oversampling (see the "Computational Processing" section).

Going back to the physical meaning of these phase space regions' respective ray bundles, we can conclude that each pinhole/pixel combination yields a single measurement (i.e., a single sample of the light field function) through integration by the pixel. The phase space region, therefore, represents the spatio-directional sampling kernel introduced by the finite size of the pixel and the pinhole, respectively, while the center ray/phase space point indicates the associated sampling position.

A key optical concept, the optical invariant, posits that an ideal optical system does not change the volume of such a phase space region (also known as *étendue*). As an example, free-space transport, as a particularly simple propagation, maintains phase space volume; it is described by a shear in the horizontal direction of the phase space. Free-space transport to a different plane is a necessary ingredient for computing refocused 2-D images from the light field.

Light field sampling with camera arrays/moving cameras
Obviously, pinhole images are of a low quality due to blurring by the finite pinhole area—or, depending on its size, diffraction effects—and to the low light throughput. Introducing a lens in the light field plane p improves the situation. This measure has the side effect of moving the apparent position of the sensor plane q in front of the light field plane p if the sensor is positioned at a farther distance than the focal length of the lens, as shown in Figure 3. The ray bundles being integrated by

a single pixel can still be described by a two-aperture model as before; however, at this point the model must be considered virtual. This implies that it may intersect scene objects. It is understood that the virtual aperture does not affect the scene object in any way. The key point is that the refracted rays in the image space of the lens can be ignored as a way of simplifying the description. Only the ray bundles in the world space that are being integrated by the pixel are considered.

With this change, the sampling of the light field remains the same as before: instead of moving a pinhole, a moving standard 2-D camera performs the sampling task. Only the parameterization of the directional component s needs to be adapted to the camera's intrinsic parameters. This is how pioneering work was performed [2], [3]. Of course, this acquisition scheme can be implemented in a hardware-parallel fashion by means of camera arrays [8], [12].

Given a sampled light field $l(u, v, s, t)$ and assuming full information to be available, the slices $I(s, t) = l(u = \text{const.}, v = \text{const.}, s, t)$ as well as $I(u, v) = l(u, v, s = \text{const.}, t = \text{const.})$ correspond to views into the scene. The function $I(s, t)$ corresponds to a perspective view, while $I(u, v)$ corresponds to an orthogonal view of the inside space. These views are often referred to as *light field subviews*.

Optics for light field cameras
While camera arrays can be miniaturized as demonstrated by Pelican Imaging Corp. [12] and differently configured camera modules may be merged as proposed by LightCo. Inc. [13], there are currently no products for end users, and building and maintaining custom camera arrays is costly and cumbersome.

In contrast, the current generation of commercial light field cameras by Lytro Inc. [14] and Raytrix GmbH [15] has been built around in-camera light field imaging, i.e., light field imaging through a main lens. In addition, attempts are being made to build light field lens converters [16] or use mask-based imaging systems [17] that can turn standard single-lens reflex cameras into light field devices. All devices for in-camera light field imaging aim at sampling a light field plane p inside the camera housing.

To understand the properties of the in-camera light field and their relation to the world space, we now extend the previous discussion of general light field imaging to the in-camera space.

In-camera light fields
In-camera light fields allow the light field to be transformed from the world space into the image space of a main lens, where it is acquired by means of miniature versions of the camera arrays, outlined earlier, that are most often implemented using micro-optics mounted on a single sensor. The commercial implementations involve microlenses mounted in different configurations in front of a standard 2-D sensor. Each microlens with its underlying group of pixels forms an in-camera (u, v, s, t) sampling scheme, as described in the previous section. We may also think of these as tiny cameras, with very few pixels, observing the in-camera light field. The image of a single microlens on the sensor is often referred to as a *micro-image*.

Unfortunately, the in-camera light field is a distorted version of the world coordinate light field due to refraction by the main lens. Here, we encounter a classic misconception: mapping the world space into the image space of the main lens, even by means of a simple thin-lens transformation, does not result in a uniformly scaled version of the world space. Instead, the in-camera light field is a projectively distorted version of the world-space light field (see Figure 4), which results from the depth-dependent magnification of optical systems.

There are different ways to describe this distortion, e.g., in terms of phase space coordinates, as suggested by Dansereau et al. [18], corresponding to a ray-remapping scheme or by appropriate projection matrices. The projection matrices commonly used in computer vision to model camera intrinsics and extrinsics are not directly usable because they model a projection onto the image plane of a 2-D camera. It is, however, important that 3-D information is preserved. The closest model is the OpenGL projection matrices used in computer graphics to transform a Euclidean world space into a space of so-called "normalized device coordinates." This space is also a 3-D space, but a perspectively distorted one.

Interpreting in-camera light field imaging in the world space

Thinking about how a miniature camera array is imaging the distorted in-camera light field is a bit difficult. It is, however, possible to apply the inverse perspective transformation to the light field plane and the virtual sensor plane—i.e., to the two aperture planes characterizing a light field sampling device—to obtain a world-space description in terms of an equivalent camera array.

The detailed position of these two planes depends on the configuration of the light field camera. There are essentially two choices:

■ an afocal configuration of the lenslets [19]
■ a focused configuration of the lenslets [20], [15].

In the first case, the sensor plane is positioned exactly at the focal distance of the microlens array. In the second, there are two possibilities for creating real or virtual imaging configurations of the microcameras: by positioning the sensor plane farther from or closer to the microlens focal length, respectively. This choice has the effect of placing the in-camera virtual sensor plane at different positions: namely at infinity for an afocal configuration or in the front or in the back of the microlens plane for a focused configuration.

In practice, the first can only be approximately achieved. First, it is difficult to mechanically set the sensor at

the right distance from the microlens array. Second, because a microlens is often a one-lens system, its focal length is strongly dependent on the wavelength of the light. The configuration may be set for green light, but the red and blue wavelengths are then focused at different distances. The finite pitch of the pixels, however, makes the system tolerant to these issues.

In microlens-based light field imaging, the microlens plane takes the role of the in-camera light field plane p. The virtual sensor plane (i.e., the sensor plane transformed by the microlens array) takes the role of the second aperture, as in Figure 3.

The inverse action of the main lens, then, is to map these two planes into the world space. In conjunction, they define the properties of the light field subviews such as focal plane, depth of field, viewing direction and angle, field of view, and—through these parameters—the sampling pattern for the world-space light field. Optically refocusing the main lens (i.e., changing its position with respect to the microlens array) affects most of these properties. The precise knowledge of the optical configuration is, therefore, necessary for advanced image processing tasks such as superresolution, and corresponding calibration schemes have been developed, as discussed in the "Calibration and Preprocessing" section.

Optical considerations for the main lens

The main optical considerations concern the (image-side) f-number of the main lens and the (object-side) f-number of the microlenses, respectively. The f-number of an imaging system is the ratio of its focal length and the diameter of its entrance pupil. It describes the solid angle of light rays that are passed by an
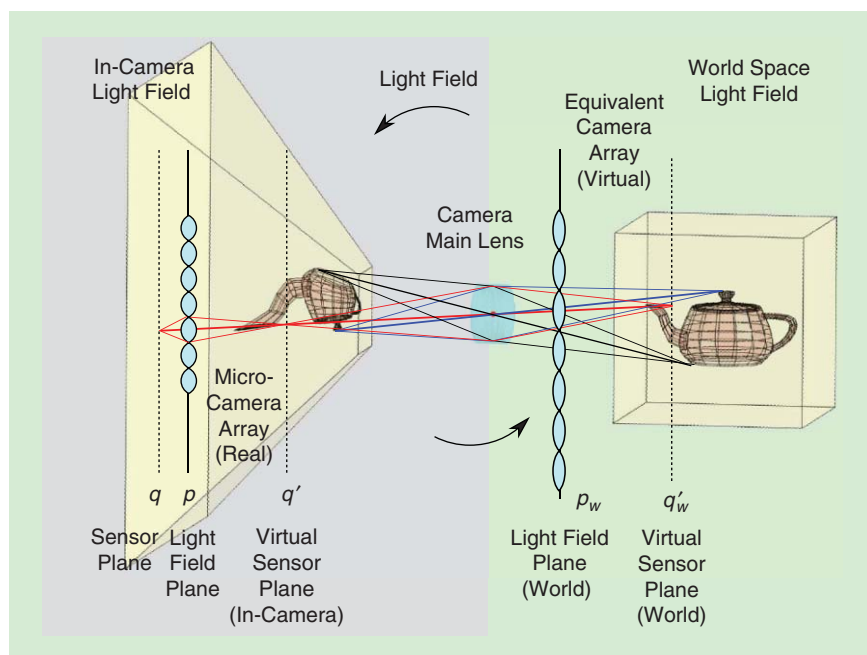
FIGURE 4. The main lens images its object space (right) into its image space (left), distorting it in the process. The world-space light field is, therefore, distorted into an in-camera light field. The distortion is a perspective projection, with its center at the center of the main lens's image-space principal plane. A micro-optics implementation of a camera array observes the distorted in-camera light field. An equivalent camera array in world coordinates can be found by mapping the light field plane p and the virtual sensor plane q to the world space.

optical system. The f-number is an inverse measure: i.e., larger f-numbers correspond to smaller solid angles. For in-camera light field systems, the f-number of the main lens must always be larger than that of the microlenses to ensure that light is not leaking into a neighboring microcamera. At the same time, for a good directional sampling, the f-number should be as small as possible. Ideally, the main lens f-number would remain constant throughout all operational conditions. This requirement imposes additional constraints, especially on zoom systems [21].

The discussion so far has involved ideal first-order optics. In reality, however, optical systems exhibit aberrations, i.e., deviations from perfect behavior. Initial investigations [22] have shown that the phase space sampling patterns are deformed by the main lens aberrations. In addition to the classic distinction between geometric and blurring aberrations, an interpretation of the phase space distortions suggests that directional shifts (i.e., a directional variant of the geometric distortions) and directional blur (i.e., a mixture of subview information) are introduced by aberrated main lenses. The effects of microlens aberrations are relatively minor and only concern the exact shape of the sampling kernel.

An example of the distortions introduced by an aberrated main lens, as opposed to an ideal thin lens, is illustrated in Figure 5. The horizontal shifts in the sampling patterns correspond to geometric distortion, typically treated by radial distortion models [18], [23]. The (slight) vertical shifts correspond to a directional deformation of the light field subviews. A known shifting pattern can be used to digitally compensate for main lens aberrations [22] or even to exploit the effect for improving light field sampling schemes [24] (see the "Computational Processing" section).

While a satisfactory treatment of first-order light field imaging can be achieved by trigonometric reasoning or updated matrix optics techniques, a complete theory of light field aberrations is missing as of this writing.

Calibration and preprocessing

Calibration and preprocessing are tightly interlinked topics for light field imaging. As outlined in the previous section, many parameters of a light field camera change when the focus of the main lens is changed. This concerns not only the geometric characteristics of the views but also their radiometric properties. The preprocessing of light field images needs to be adapted to account for these changes. In addition, different hardware architectures require adapted preprocessing procedures. We will, therefore, cover these steps based on only one example (here, a Lytro camera, which is an afocal lenslet-based light field imaging system); the underlying issues, however, affect all types of in-camera light field systems.

Color demosaicking

Using a standard Bayer color filter array to enable colored light field imaging appears to be a straightforward choice. However, as shown in Figure 6(a), for the case of an afocal light field camera each micro-image encodes the (s, t) dimensions of the light field. Different color channels, therefore, correspond to different (s, t) sampling patterns. The final image quality can be improved by taking this fact into account [25].

Vignetting

The intensity fall-off toward the sides of the micro-images, also known as *vignetting*, changes with the optical settings of
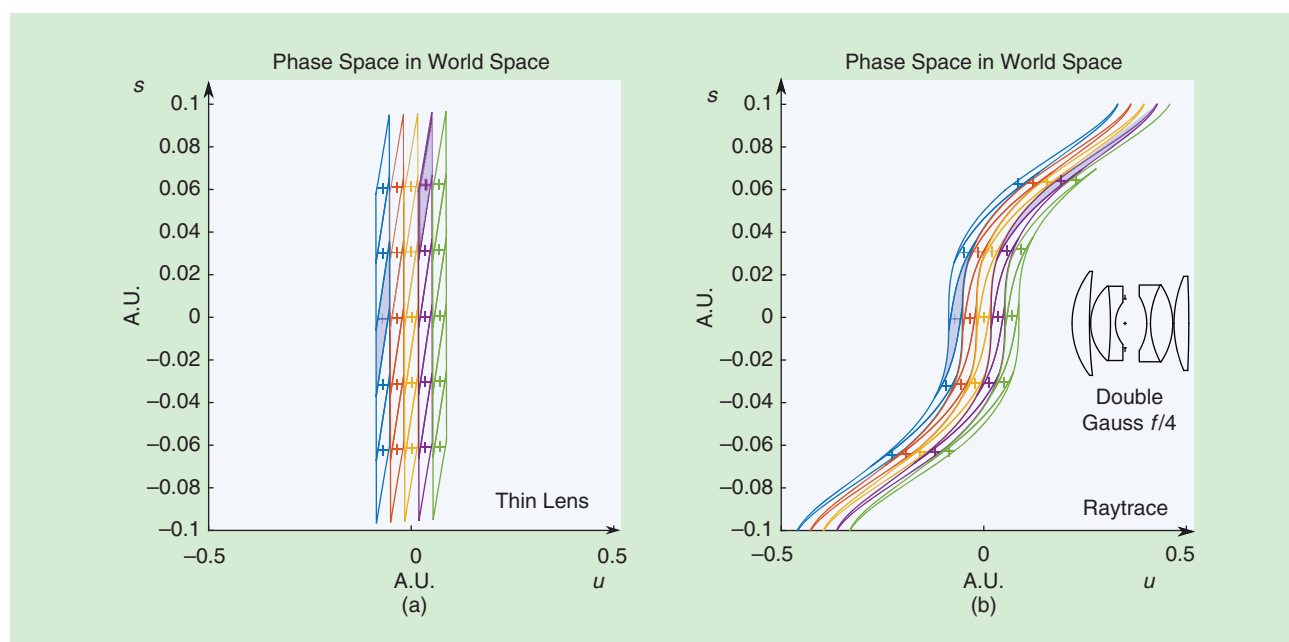
FIGURE 5. The effect of lens aberrations for an f/4 afocal light field system: (a) the phase space distribution of the sampling pattern in the world space, assuming an ideal main lens (thin lens), and (b) the phase space distribution of the sampling pattern in the world space using an f/4 double Gaussian system as a main lens. The sampling pattern is significantly distorted. The highlighted phase space regions correspond to the space between the left and the right plots. The side subview (purple) is more severely affected compared to the center subview (blue).
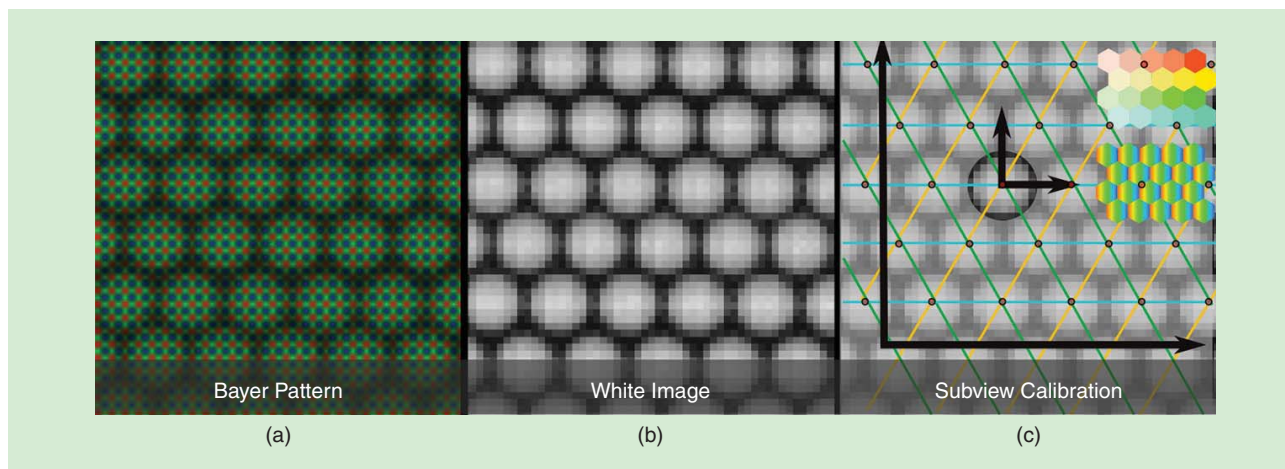
FIGURE 6. Light field preprocessing and calibration for a Lytro camera. (a) Using a Bayer pattern within the micro-images causes a shift of light field view for the color channels because different colors sample different (s, t) coordinates. (b) A white image (luminance) used for vignetting compensation, (c) A subpixel determination of the centers of the micro-images enables a calibrated (s, t)-coordinate system to be assigned to each micro-image. The (u, v) coordinates are sampled in a hexagonal fashion by the microlenses. The orientation of this global coordinate system also determines the rotation angle of the (s, t) system. The inset shows s and u calibration maps for the raw image.

the main lens. Commercial cameras, therefore, store significant amounts of calibration information in the internal camera memory. As an example, the combined vignetting of a main lens and microlenses changes across the field of view and with the focus and zoom settings of the main lens. Therefore, white images have to be taken for a sufficiently dense set of parameter settings. The closest white image to the parameters of a user shot are then used for compensation. In a lab setting, it is advisable to take one's own white images prior to data acquisition.

Calibration

To properly decode the four light field dimensions from the 2-D sensor image, it is necessary to carefully calibrate the (u, v, s, t) coordinates of every pixel that has been recorded by the sensor. With current lenslet-based architectures, to the first order this amounts to determining the center positions of the lenslets and the layout of the lenslet grid [Figure 6(c)]. More accurately, the position of the central view is given by the sensor intersection of the chief rays passing through the main lens and each one of the lenslets. In addition, microlens aberrations and angularly variable pixel responses can shift this position [26]. In general, the responses are also wavelength dependent.

The lenslet grid is typically chosen to be hexagonal so as to increase the sensor coverage. The spherical shape of the micro-images and their radius are determined by the vignetting of the main lens, which is the result of its aperture size and shape. The tight packing of the micro-images is achieved by f-number matching, as discussed in the "Optics for Light Field Cameras" section. It should also be noted that manufacturing a homogeneous lenslet array is difficult and so some variation may be expected. Further, the mounting of the lenslet array directly on the sensor may induce a variable distance between the sensor and the lenslets.

The calibration described here usually pertains to the in-camera light field coordinates. When assuming thin-lens optics for the main lens, these correspond to a linear transformation of the light field coordinates in the object space. Calibration approaches to determine this mapping are described by Dansereau et al. [18] for afocal light field cameras; the techniques, as well as the preprocessing steps described earlier, are implemented in their Matlab Light Field Toolbox. Bok et al. [27] present an alternative for performing a similar calibration by directly detecting line features of a calibration target from the raw light field images. Johannsen et al. [23] describe the calibration scheme for focused light field cameras. The handling of the effects of optical aberrations by the main lens is usually performed using classical radial distortion models from the computer vision literature. While these measures improve the accuracy, they are not completely satisfactory because the light field subviews suffer from nonradial distortions [see Figure 5(b)]. Lytro provides access to calibration information, including aberration modeling, through its software development kit. Alternatively, modelless per-ray calibrations [28] using structured light measurements have shown promising performance improvements. However, the need for a principled distortion model remains.

Once a per-pixel calibration is known, the suitably preprocessed radiance values of the light field function can be assigned to a sample position in the phase space. In principle, reconstructing the full light field function amounts to a signal processing task: given a set of irregular samples in the phase space, reconstruct the light field function on that space. In practice, additional constraints apply and are used to, e.g., achieve superresolution or to extract depth. A prerequisite for superresolution is having a known shape for the phase space sampling kernels, also called *ray-spread functions*. Calibration schemes for these still have to be developed.

Computational processing

The reconstruction of the 4-D light field function from its samples can be achieved by standard interpolation schemes [2], [29]. However, the light field function possesses additional

FIGURE 7. The light field subview and EPI corresponding to the green line in the subview. The images represent different slices of the 4-D light field function $I(u, v, s, t)$. Note the linear structures of constant color in the EPI. These structures correspond to surface points. Their slope is related to the depth of the scene point.

structure. It is not an arbitrary 4-D function, but its structure is determined by the geometry and radiometry of the scene.

As an example, if the sampled part of the light field plane p_w is small with respect to the distance to an object point within the inside region, then the solid angle of the system aperture with respect to the surface point is small. If the surface is roughly Lambertian, the reflectance does not vary significantly within this solid angle and can be assumed constant. This restriction often applies in practice, and the mixed positional–directional slices of the light field function, e.g., $l(u, v = const., s, t = const.)$, show a clear linear structure, as shown in Figure 7. These images are also known as *epipolar plane images* (*EPIs)* with reference to the epipolar lines of multiple-view computer vision. In the case of non-Lambertian surfaces, the linear structures carry reflectance information that is convolved with the illumination.

Assuming the constancy of the light field function along these linear structures to be a valid approximation and considering the 4-D case instead of our 2-D illustrations (i.e., planar structures instead of linear ones corresponding to geometric scene points), we see that the intrinsic dimensionality of the light field is only 2-D in the Lambertian case. In practice, it is necessary to have a knowledge of scene depth to exploit this fact. On the other side, the constraint serves as a basis for depth estimation. This observation is the basis for merging the steps of reconstructing the light field function (signal processing), depth reconstruction (computer vision), and superresolution (image processing). More general constraints are known. As an example, Levin et al. [30] proposed a 3-D constraint in the Fourier domain that works without depth estimation.

Intuitively, the linear structure implies that the surface point corresponding to a sloped line can be brought into focus, which in the phase space is a shear in the horizontal direction (see also Figure 1). Focus is achieved when the sloped line becomes vertical. In this case, there is only angular information from the surface point, which implies that its reflectance (convolved with the incident illumination) is being acquired. The amount of shear necessary to achieve this focusing is indicative of the depth of the scene point with respect to the light field plane p_w. The slope of the linear structures is, therefore, an indicator for depth.

Depth estimation

In light of the previous discussion, depth estimation is a first step toward superresolution. It amounts to associating a slope with every phase space sample [11]. There are several ways to estimate depth in light fields. The standard way is to extract light field subviews and to perform some form of multiview stereo estimation. Popular techniques such as variational methods [31], [16] or graph-cut techniques [32] have been explored. The literature on the topic is too large to review here, and we recommend consulting the "References" for further discussions.

The main differences between multiview stereo on images from regular multicamera arrays and for light field cameras are the sampling patterns in the phase space. Whereas the sample positions and sampling kernels of multicamera arrays are typically sparse in the phase space, for light field cameras the respective sampling patterns and kernels usually tile it. Therefore, there is a difference in the aliasing properties of these systems. Aliased acquisition implies the need to solve the matching or correspondence problem of computer vision, a notoriously hard problem. In addition, the phase space slope vectors are only estimated indirectly through (possibly inconsistent) disparity assignments in each of the subviews.

The dense sampling patterns of light field cameras allow for alternative treatments. As an example, recent work has explored the possibility of directly estimating the linear structures in the EPI images [33] based on structure tensor estimation. This technique involves directly assigning the slope vectors to each point in the phase space. However, it does not model occlusion boundaries (i.e., T-junctions in the phase space) and, therefore, does not perform well at object boundaries. Recent work is addressing this issue through estimating aperture splits [34] or exploiting symmetries in the focal stack data corresponding to the light field [35].

Superresolution

The knowledge of the slope function can be used to compute superresolved light fields [36], [33] by filling the phase space with lines that have the slope and the radiance associated to a phase space sample (see Figure 8). If the samples are jittered along the slope of the line, a geometric type of superresolution results. This effect is used in computer graphics rendering to inexpensively predict samples of high-dimensional integration for rendering depth-of-field effects [36]. As the samples are perfectly Dirac and the exact depth is a byproduct of the rendering pipeline, this fact is relatively simple to exploit, as compared to the corresponding tasks in light field imaging.
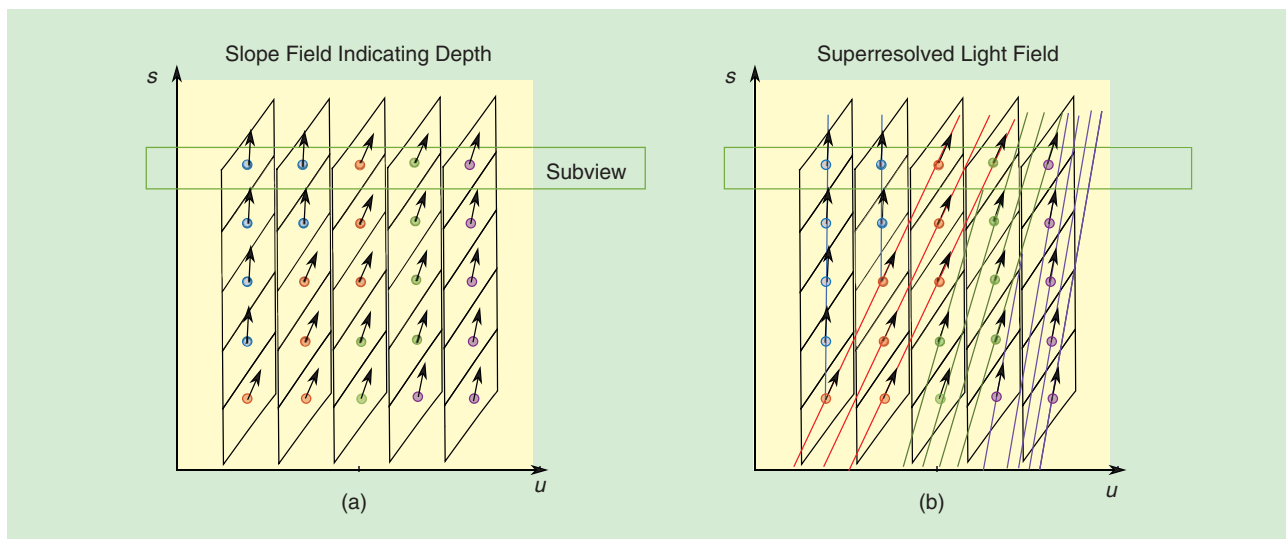
FIGURE 8. Depth estimation and superresolution. (a) Assigning a depth value to phase space samples in all subviews assigns a slope field to the light field. This can, e.g., be achieved by matching samples between subviews as in stereo or multiview stereo matching. (b) Propagating the radiance values of the samples along the slope field generates a superresolved light field and, therefore, superresolved subviews. Because the samples represent a convolution with the sampling kernel, a deconvolution step following the line propagation improves the result. The line propagation needs to consider occlusion (T-junctions).
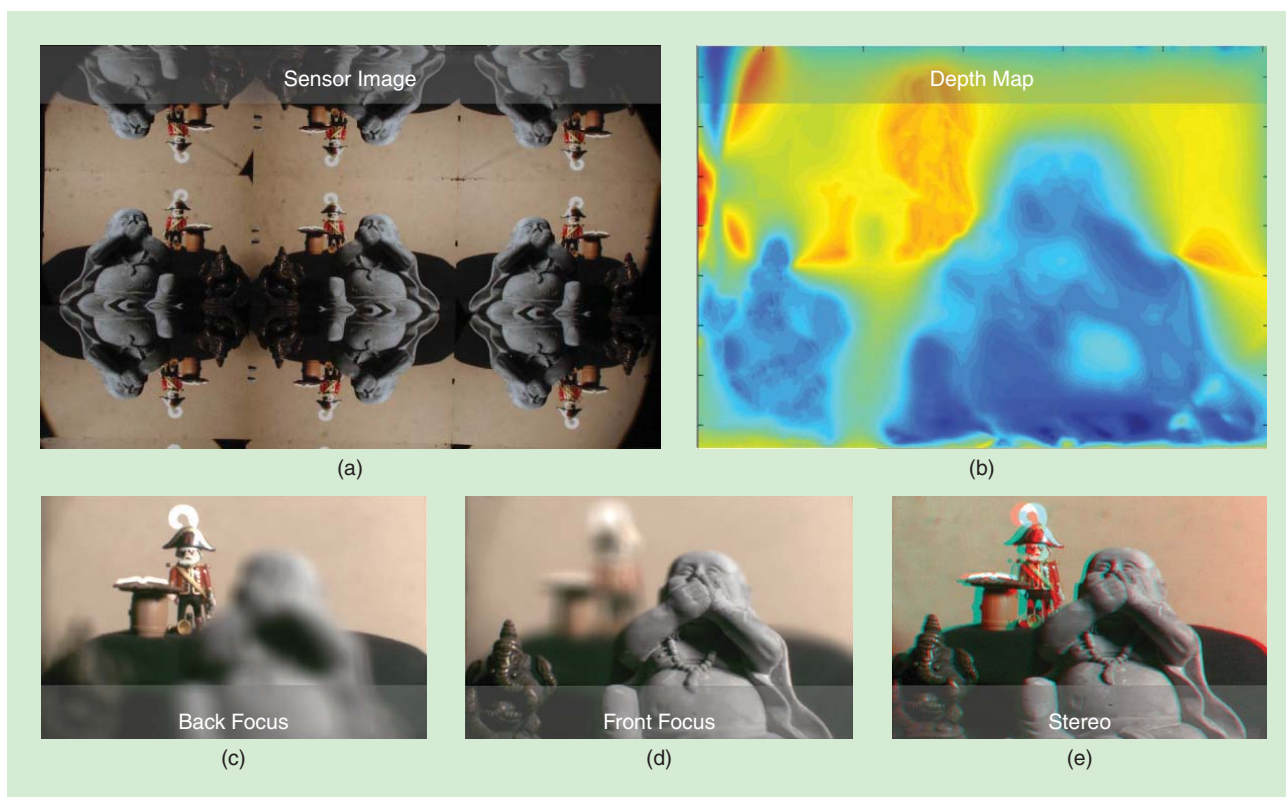
FIGURE 9. Light field image synthesis. (a) A raw image from the optical light field converter of Manakov et al. [16], (b) a depth map for the center view, computed with multiview stereo techniques, (c) and (d) a back and front focus using extrapolated light fields to synthesize an f/0.7 aperture (physical aperture f/1.4), and (e) a synthesized stereo view with user-selectable baseline.

In working with real data, the depth needs to be estimated as described earlier. Because the samples are affected by the sampling kernel (i.e., the phase space regions associated with a sample), true superresolution needs the additional step of deconvolving the resulting function [37]. For microscopic light field applications, a wave optics perspective is necessary [38], [39], and the deconvolution consequently includes wave effects.

A note on aliasing

It is commonly stated in the literature that an aliased acquisition is required for superresolution [37]. In light of the previous discussion, we may make this statement more precise by stating that 1) a Lambertian scene model is implied for geometric superresolution, 2) the samples should be jittered along the slope corresponding to a scene point's depth, and 3) smaller phase space kernels associated with the samples will be beneficial as long as there is still overlap between them when propagated along the lines to construct the superresolved subview. In conclusion, light field cameras may be more suitable for implementing superresolution schemes than multicamera arrays due to their denser sampling of the phase space.

Image synthesis

Once the light field function is reconstructed, novel 2-D views can be synthesized from the data. The simplest visualization is to extract the light field subviews, i.e., images of constant (u, v) or (s, t) coordinates, depending on the sampling pattern of the specific hardware implementation. It should be noted that both choices, in general, yield perspective views. This is because in-camera orthographic views [as synthesized by fixing the (s, t) coordinates] map to a world-space center of projection in the focal plane of the main lens. The subviews correspond to the geometry of the world-space light field plane p_w and the world-space virtual sensor plane q_w and, therefore, show a parallax between views. Interpolated subview synthesis has been shown to benefit from depth information [40]: available depth information, even if coarse, enables aliasing-free view synthesis with fewer subviews.

The goal of light field image synthesis, however, is the creation of images that appear as if they were taken by a lens system not physically in place (see Figure 9). The example most commonly shown is synthetic refocusing [29]. The technique, in its basic form, consists of performing a free-space transport of the world-space light field plane to the desired focus plane. After performing this operation, an integral over the directional axis of the light field, i.e., along the vertical dimension in our phase space diagrams, yields a 2-D view focused at the selected plane. Choosing only a subrange of the angular domain lets the user select an arbitrary aperture setting, down to the physical depth-of-field present in the light field subviews, that is determined by the sizes of the two (virtual) apertures involved in the image formation. If spatio-directional superresolution techniques (as described in the "Computational Processing" section) are employed, this limit may be surpassed.

Computing the 4-D integral allows for general settings: even curved focal planes are possible by selecting the proper phase space subregions to be integrated. However, it can be computationally expensive. If the desired synthetic focal plane is parallel to the world-space light field plane p_w and the angular integration domain is not restricted, Fourier techniques can yield significant speedups [14]. If hardware-accelerated rendering is available, techniques based on texture-mapped depth maps can be efficient alternatives [16].

Conclusions

With almost a quarter century of practical feasibility, light field imaging is alive and well, gaining popularity and progressing

into the market with several actors pushing for prime time. There are still sufficiently many scientific challenges to keep researchers occupied for some time to come. In particular, the bar of resolution loss must still be lowered in the hope of increased consumer acceptance. The megapixel race has slowed down, and pixel sizes are approaching their physical limits. This implies larger sensors and, thus, increased expense for additional resolution increases that would benefit light field technology. Improved algorithmic solutions are, therefore, of fundamental importance.

The next big step will be light field video, pushing optical flow toward scene flow and associated projected applications, such as automatic focus pulling, foreground/background segmentation, space-time filtering, etc. In terms of applications, we are seeing 4-D light field ideas penetrating in both the small and the large. In the small, we are seeing the emergence of light field microscopy [41], although we need improved aberration models and, eventually, expanded wave-optical treatments [39]. In the large, sensor networks will become increasingly important. More complex scenes—such as translucent objects [42] or, more generally, non-Lambertian scenes [43]—are made possible. Crossover to other fields, such as physics, are also appearing [44]. These are surely exciting times as we head into the second quarter century of light field technology.

Acknowledgments

Authors

Ivo Ihrke (ivo.ihrke@inria.fr) is currently a researcher with Carl Zeiss, Inc. From 2013 to 2016, he was a permanent researcher at the Inria Bordeaux Sud-Ouest Research Institute, France; previously, he was an Emmy-Noether fellow with the German Research Foundation, and he headed a research group within the Cluster of Excellence, Multimodal Computing and Interaction, at Saarland University, Germany, which was also associated with the MPI Informatik and the Max-Planck Center for Visual Computing and Communications (2010–2012). He was a Humboldt fellow at the University of British Columbia, Vancouver, Canada (2008–2009). He received his M.S. degree in scientific computing from the Royal Institute of Technology, Stockholm, Sweden, in 2002 and his Ph.D. degree (summa cum laude) in computer science from Saarland University in 2007. He coorganized a workshop on computational cameras and displays at the 2012 Conference on Computer Vision and Pattern Recognition, a Dagstuhl seminar on computational imaging in 2015, and the Zeiss Symposium Workshop on Computational Imaging in 2016.

John Restrepo (john.restrepo@inria.fr) received his physics engineering degree and his master's degree in physics from the Universidad Nacional de Colombia, Medellin campus, in 2008 and 2010, respectively. He continued his research work with an internship at the Cluster of Excellence, Multimodal Computing and Interaction, at Saarland University, Germany. Since 2013, he has been pursuing his doctorate in computer science at the Université de Bordeaux, France, working at the Inria Bordeaux Sud-Ouest Research Institute.

Loïs Mignard-Debise (lois.mignard-debise@inria.fr) is currently a Ph.D. degree student at the INRIA Bordeaux Sud-Ouest research institute, France. He received an M.S. degree in optics from the Institut d'Optique Graduate School, Palaiseau, France, in 2014.

References

[1] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Manchine Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[2] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 1996, pp. 31–42.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proc. 23rd Ann. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 1996, pp. 43–54.

[4] M. Levoy, "Light fields and computational imaging," *IEEE Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[5] C. Zhou and S. K. Nayar, "Computational cameras: Convergence of optics and processing," *IEEE Trans. Image Processing*, vol. 20, no. 12, pp. 3322–3340, 2011.

[6] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational plenoptic imaging," *Comp Graphs Forum*, vol. 30, no. 8, pp. 2397–2426, Oct. 2011.

[7] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. Cambridge, MA: MIT Press, 1991.

[8] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.

[9] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle, "Surface light fields for 3-D photography," in *Proc. 27th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 2000, pp. 287–296.

[10] A. Torre, *Linear Ray and Wave Optics in Phase Space*. Amsterdam: Elsevier, 2005.

[11] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera tradeoffs through a Bayesian analysis of light field projections," in *Proc. 10th European Conf. Computer Vision*, Berlin, 2008, pp. 88–101.

[12] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "PiCam: An ultra-thin high performance monolithic camera array," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 166, 2013.

[13] R. Laroia, "Zoom related methods and apparatus," U.S. Patent 14/327,525, 2014.

[14] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, 2005.

[15] C. Perwass and L. Wietzke, "Single lens 3-D-camera with extended depth-of-field," in *Proc. SPIE Human Vision and Electronic Imaging XVII*, 2012, vol. 8291, p. 829108.

[16] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H. P. Seidel, and I. Ihrke, "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging," *ACM Trans. Graph.*, vol. 32, no. 4, article 47, 2013.

[17] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69, 2007.

[18] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034.

[19] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford University, Stanford, CA, Stanford Tech Report CTSR 2005-02, 2005.

[20] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. IEEE Int. Conf. Computational Photography*, 2009, pp. 1–8.

[21] T. J. Knight and Y.-R. Ng, and C. Pitts, "Light field data acquisition devices and methods of using and manufacturing same," U.S. Patent 8,289,440, 2012.

[22] P. Hanrahan and R. Ng, "Digital correction of lens aberrations in light field photography," in *Proc. SPIE Int. Optical Design Conf.*, 2006, vol. 6342, p. 63421E.

[23] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß, "On the calibration of focused plenoptic cameras," in *Time-of-Flight and Depth Imaging: Sensors, Algorithms, and Applications*. New York: Springer, 2013, pp. 302–317.

[24] L. Y. Wei, C. K. Liang, G. Myhre, C. Pitts, and K. Akeley, "Improving light field camera sample design with irregularity and aberration," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 152, 2015.

[25] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "An analysis of color demosaicing in plenoptic cameras," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 901–908.

[26] C. K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 1–19, Mar. 2015.

[27] Y. Bok, H. G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," *IEEE Trans. Pattern Anal. Machine Intell.*, 2016. doi: 10.1109/TPAMI.2016.2541145

[28] F. Bergamasco, A. Albarelli, L. Cosmo, A. Torsello, E. Rodola, and D. Cremers, "Adopting an unconstrained ray model in light-field cameras for 3-D shape reconstruction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3003–3012.

[29] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proc. 27th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 2000, pp. 297–306.

[30] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, 2010, pp. 1831–1838.

[31] S. Heber, R. Ranftl, and T. Pock, "Variational shape from light field," in *Proc. Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2013, vol. 8081, pp. 66–79.

[32] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 1547–1555.

[33] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and superresolution," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, no. 3, pp. 606–619, 2014.

[34] T. C. Wang and A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *IEEE Int. Conf. Computer Vision*, 2015, pp. 3487–3495.

[35] H. Lin, C. Chen, S. B. Kang, and J. Yu, "Depth recovery from light field using focal stack symmetry," *Proc. IEEE Conf. Int. Computer Vision*, 2015, pp. 3451–3459.

[36] J. Lehtinen, T. Aila, J. Chen, S. Laine, and F. Durand, "Temporal light field reconstruction for rendering distribution effects," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 55, 2011.

[37] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Proc. IEEE Int. Conf. Computational Photography*, 2009, pp. 1–8.

[38] S. A. Shroff and K. Berkner, "Image formation analysis and high resolution image reconstruction for plenoptic imaging systems," *Appl. Opt.*, vol. 52, no. 10, pp. D22, 2013.

[39] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-D deconvolution for the light field microscope," *Opt. Exp.*, vol. 21, no. 21, pp. 25418, 2013.

[40] J. X. Chai, S. C. Chan, H. Y. Shum, and X. Tong, "Plenoptic sampling," in *Proc. 27th Ann. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 2000, pp. 307–318.

[41] M. Levoy, "Light field photography, microscopy and illumination," in *Int. Optical Design Conf. Optical Fabrication and Testing. Optical Society of America*, 2010, p. ITuB3.

[42] Y. Xu, H. Nagahara, A. Shimada, and R.I. Taniguchi, "TransCut: Transparent object segmentation from a light-field image," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 3442–3450.

[43] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous Fourier domain," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 1–13, Dec. 2014.

[44] C. Skupsch and C. Brücker, "Multiple-plane particle image velocimetry using a light-field camera," *Opt. Express*, vol. 21, no. 2, pp. 1726–1740, Jan. 2013.

SP

Kristin J. Dana

Capturing Computational Appearance

More than meets the eye
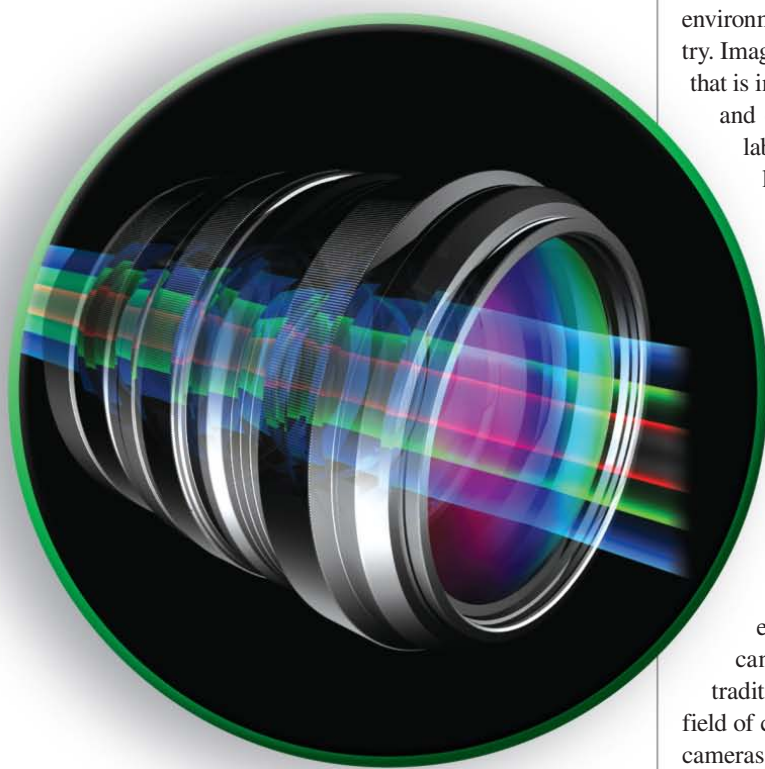
Surface reflectance and texture provides a unique signature for applications such as recognition and rendering. Intuition tells us that a camera captures appearance. However, a traditional camera captures intensity dependent on the environment lighting, camera position, and the surface geometry. Imaging for computational appearance recovers reflectance that is intrinsic to an object or scene and useful for recognition and other applications. Reflectance can be captured in a lab-based setting with a gonioreflectometer or domes of lights and cameras. Recent methods in computational imaging provide appearance-capture that is comprehensive, efficient, compact, or optimal depending on the task at hand. In this article, we review methods for capturing and modeling computational appearance. The impact of these appearance representations is significant with applications areas such as e-commerce, digital architecture, human-computer interaction, intelligent vehicles, robotics, and inspection.

Introduction

A camera captures appearance by providing a photographic image to emulate and store the visual experience. Appearance of objects, scenes, and people can be captured for human viewing and sharing. The traditional camera is analogous to the human eye, and the field of computer vision has developed methods for calibrated cameras that act as directional light sensors to record the distribution of light intensity reflected from the scene. Computational imaging takes computer vision a step further by redesigning the camera to capture scene information beyond what the human eye captures in a single view. The form and function of cameras are malleable, and creative research is taking the camera to new designs by measuring computational appearance to provide a unique signature for automated recognition and other tasks.

Capturing appearance is essentially light sampling with numerous inherent parameters and sampling choices. These

©ISTOCKPHOTO.COM/YAKOBCHUK

parameters include viewing angle, illumination angle, spatial scale, spectral range, and polarization [39], [62]. The sampled light may be emitted from a point, patch, object, or global scene. The plethora of variations for appearance specifications leads to the basic question: What is the relevant definition of appearance? The parameters are sometimes determined by the capabilities of the imaging device, but the needs of the end-use application also drives appearance capture. Just as computational imaging refers to the computation inherent in the imaging process, we introduce the term *computational appearance* to indicate how the working definition of appearance depends on the computation in the subsequent algorithms and applications. Figure 1 illustrates computational appearance as dependent on both the constraints of the imaging device and the needs of the application algorithm.

For example, appearance is often used for recognition in computer vision. Reflected intensity from an object or scene is dependent on environment lighting, camera view, and surface geometry. The human visual system handles this variation and maps the set of possible appearances to a unique object representation for recognition. The representation of an object within the human visual system is invariant to environment lighting and camera pose. For instance, a friend is recognizable at morning or at midday; recognition persists when the illumination direction has changed. Similarly, recognition is not interrupted when an observer moves and the viewpoint changes. This geometric and photometric invariance is a primary goal in computer vision and invariant representations of appearance provide a mechanism to recognize objects and scenes. A strategy to develop these representations is to observe the scene under variations in scene illumination and camera pose (camera position and orientations). These comprehensive measurements are the input to modeling algorithms which provide a computational appearance descriptor as illustrated in Figure 2. For automated recognition, the appearance descriptors form a training set can be used to build a machine-learning classifier so that a class label is provided as the output. For computer graphics rendering, computational appearance is used to render the complex appearance including effects such as gloss, sheen, texture and translucency. An important question is how best to measure appearance, i.e., how to sample the space of illumination and camera pose, to have a sufficiently detailed descriptor.

Consider the surfaces depicted in Figures 3 and 4. Figure 3 shows two surface snapshots that fail to convey appearance in a meaningful manner; i.e., an observer cannot identify the surface from either photo. Figure 4 depicts material folds that reveal how the surface reflects light over variations in the incident illumination and viewing direction. In this manner, the materials satin and velvet are easily discerned by a human observer. Notice that neither
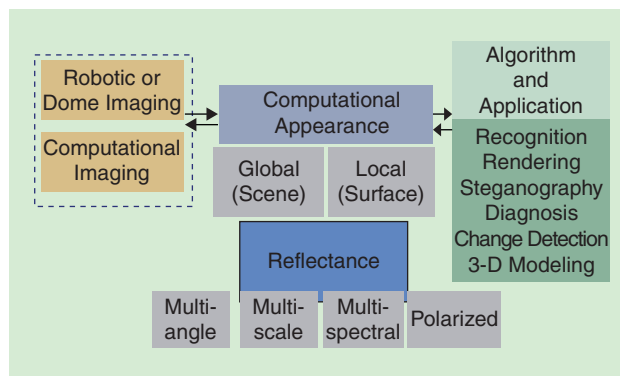
FIGURE 1. Capturing appearance with imaging. Appearance comprises reflectance captured from the global scene or from a local surface patch; it may be multiangle, multiscale, multispectral, and include polarization information. Domes of lights and cameras or robotic manipulators can be used for appearance capture. Alternatively, computational imaging provides appearance using specialized devices that are typically faster and more portable. Imaging provides appearance to the end-use algorithm or application. The bidirectional arrow indicates that the computation within the end-use algorithm may determine the specifications of computational appearance and therefore affect the imaging process.

the color nor the shape are of interest in this recognition. Instead, it is the specific way these surfaces reflect light that provide their identifying signature. Though the geometry of objects is not typically controllable, illumination and camera pose can be manipulated in a strategic manner to capture comprehensive reflectance for object and surface modeling. Figure 5 shows representative cubes of different materials; neither the color nor geometry are key in describing the object properties; the unique reflectance properties of cube discriminate or render the materials.

Beyond basic object recognition, example applications for computational appearance are numerous and include e-commerce, art archiving, robotic navigation, architecture, and biomedical applications such as dermatology. For e-commerce, conveying material appearance is an important component of the online presentation of consumer goods such as shoes, clothing, furniture, linens, and home decor. Color and geometry are typically understood as important in digital models, but materials enhance the digital presentation and convey subtle visual
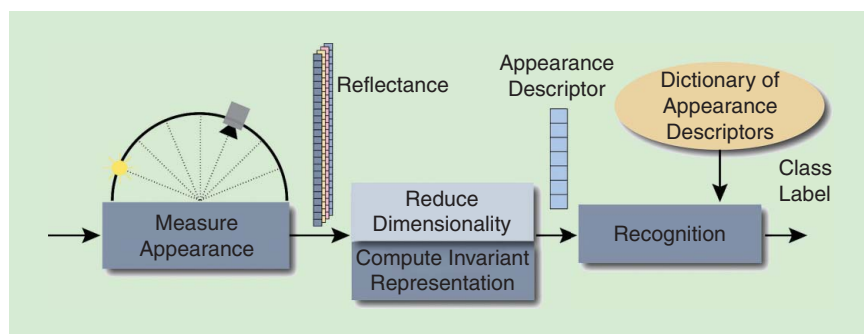
FIGURE 2. In an automated recognition pipeline, reflectance is measured and then represented by a lower-dimensional invariant appearance descriptor. A dictionary of such descriptors obtained from a training set is used to recognize the object using machine learning. The recognition result is a class label identifying the object, surface, or scene.
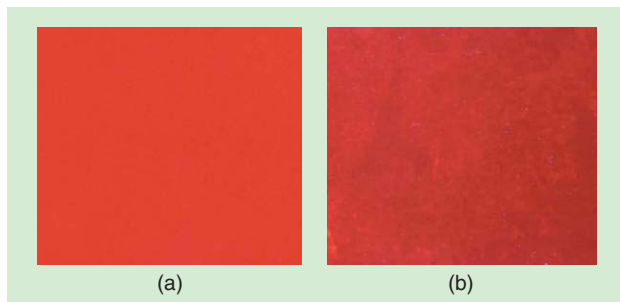
FIGURE 3. What are these? (a) and (b) show a simple snapshot of a material, but this material is not recognizable from the photo.
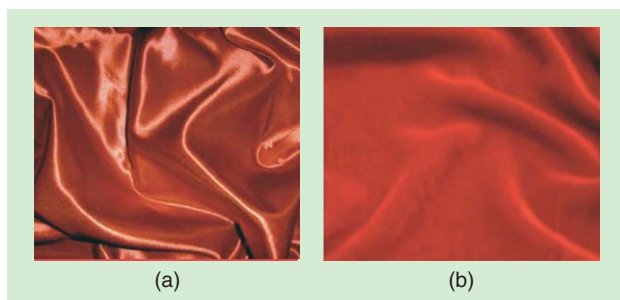
FIGURE 4. The materials [(a) satin and (b) velvet] are easily recognizable in these images because the reflectance from multiple angles is observable in the folds.

FIGURE 5. Cubes of different material composition (cork, foam, aluminum, acrylic, maple wood, granite, bakelite resin, and silicone); *Architect's Cubes* Museum of Modern Art, New York. (Photo courtesy of Kristin J. Dana.)

FIGURE 6. The BRDF (bidirectional reflectance distribution function) $f(\theta_i, \phi_i, \theta_v, \phi_v)$ is the ratio of radiance from a surface point in the direction θ_v, ϕ_v due to incident irradiance from direction θ_i, ϕ_i.

guide mobile robot navigation. In architectural applications, creating models of existing infrastructure using photography has become an important modern tool. Adding composition details creates a compendious scene description with surface labeling of material composition of built structures such as brick, limestone, stucco, wood, stone, marble, slate, granite, and copper.

Reflectance: Functions and fields

According to a geometric representation of light rays, a surface point is illuminated by an incident ray oriented by polar angle θ_i and azimuth angle ϕ_i as shown in Figure 6. The reflected light is oriented along the direction θ_v, ϕ_v. In a mirrored surface model, the incident light ray reflects at the air-surface boundary in a direction that is rotated 180° about the surface normal. For nonmirrored surfaces, a typical modelling approach is to divide reflectance into specular (mirror-like) and diffuse components. In these reflectance models, the specular component of the light intensity is maximal when the viewer is at the mirror reflection direction, but then falls off according to a parametric model as the viewer direction moves away from the peak direction. For example, in the computer graphics Phong reflection model [7], the intensity falls off as $\cos^n \alpha$, where α is the angle between the mirror reflection direction and the observer direction and n is a parameter. A diffuse component of the reflection specified by the Lambertian model is independent of observer direction and is proportional to $\cos \theta_i$, where θ_i is the angle the incident light ray makes with the surface normal. The advantage of analytical models of reflectance is the low number of parameters that can be used to describe appearance. The disadvantage is the lack of descriptive power of such models. For real-world surfaces, a portion of the light is transmitted into the surface, scatters within the surface, is

features. For art and archeological archiving, appearance can be integrated in three-dimensional (3-D) models so that fine detail such as artist brushstrokes or the translucency of a marble statue can be captured for realism. In robotics, navigation and gripping parameters depend on surface material properties. For example, robotic control variables will be different in mud than on marble, so the ground terrain material label can

absorbed for specific wavelength, and then exits the surface at a range of angles with energy in wavelengths consistent with what we perceive as the object's color. The process is complex and difficult to model precisely. The diffuse and specular components of reflection are insufficient to model the rich visual experience of real-world scenes including frosty snow, weathered copper, translucent marble, glossy paints, iridescent shells, crushed velvet, woven burlap, aged asphalt, and metallic-flake pigments. By measuring appearance directly, data-driven reflectance models can be used to build representations that are tuned to specific appearance classes.

The intensity of light reflected from a surface point when illuminated at angle θ_i, ϕ_i and viewed from angle θ_v, ϕ_v is described by the bidirectional reflectance distribution function (BRDF) expressed as $f(\theta_i, \phi_i, \theta_v, \phi_v)$ as illustrated in Figure 6. The BRDF is defined as the ratio of the radiance exiting a surface point to the irradiance incident on the surface point [44]. The units of the BRDF are inverse steradians (sr^{-1}), where the steridian is the unit of solid angle. To parse these units, consider that input light (irradiance) is the power per unit area and has units of watts per meter squared (watts/m^2). The total output light intensity from a unit area of the surface has units of watts/m^2 but it radiates in a hemisphere of possible directions, so the output light in a particular direction has units of watts/m^2 per steridian. Each direction is represented by a solid angle so that the integration over all directions represents the entire 3-D space. Therefore the ratio of output light radiance to input light irradiance as expressed by the BRDF has units sr^{-1}. To denote dependence on both viewing and illumination angles, the BRDF is expressed as $f(\theta_i, \phi_i, \theta_v, \phi_v)$. Real-world surfaces typically do not have a uniform BRDF due to both surface markings and surface texture.

The bidirectional texture function (BTF) extends the BRDF to characterize surface reflectance that varies spatially. The early concept of BTF was introduced with the Columbia–Utrecht Texture and Reflectance (CUReT) database [11], [12] and has been used for numerous texture modeling and recognition studies. The BTF expressed as $f(x, y, \theta_i, \phi_i, \theta_v, \phi_v)$ has dependence on spatial parameters x, y and angular parameters. The BRDF assumes a point-wise light transport relationship, incident light at a point in a single direction results in exitant light at the same point from multiple directions. However, because of subsurface scattering, light incident on a point exits over a surface patch. The BTF can model these effects by capturing patch-wise reflection as shown in Figure 7. Light incident at a patch at a particular direction results in reflected light from the patch over a hemispherical range of directions. BTF modeling typically assumes that incident light is uniform over the patch. In addition to subsurface scattering, the BTF representation is useful for capturing reflection from fine-scale geometry of textured surfaces such as bumps, wrinkles, and roughness. The fine-scale shadowing, occlusions, shading, and foreshortening that affect the pixel intensities of the recorded images become part of the appearance model implicitly without knowledge of the surface fine-scale geometric variation. The reflectance at each point contains the nonlinearities of the shadowing and occlusions

of fine-scale geometry. For example, surface point at x, y may be shadowed as the illumination direction changes from θ_i to $\theta_i + \delta$ for some small angle δ, causing an abrupt change in the BTF to near zero reflectance. The BTF model can also be used to texture-map a 3-D object represented by a polygonal mesh. The 3-D mesh is texture-mapped, not with a single image, but with a BTF. Traditional texture mapping maps each 3-D vertex into a two-dimensional (2-D) texture image parameterized by texture coordinates u, v. The sampled BTF is a collection of images, so that a 3-D object vertex is mapped to $f(u, v, \theta_i, \phi_i, \theta_v, \phi_v)$ where the illumination and viewing direction are defined with respect to the mesh facet. BRDF/BTF measurements are spatially local in their description, concentrating on describing the appearance of a surface point or patch. Such a description is ideal for surfaces that exhibit spatial invariance where the appearance of the patch is representative of the general appearance as in studies of textured surfaces.

While BRDF is a pointwise reflectance measurement and BTF is a patch-based reflectance measurement, the reflectance of an entire scene can also be captured globally. Light fields and reflectance fields describe the global reflectance of the entire scene or entire object. Light fields are defined as radiance as a function of position and direction [22], [32] and are four-dimensional (4-D) since they describe a spatial position with two variables and ray orientation with two angles (polar angle and azimuth angle). An eight-dimensional (8-D) reflectance field [15] describes both the incident 4-D light field as well as the 4-D exitant light field. Reflectance fields are analogous to BRDF's since both representations are bidirectional, describing the direction of incident light and exitant light. However, reflectance fields describe the input/output light over the global scene instead of a local point. Conceptually, light fields and reflectance fields construct a closed surface such as a sphere (or cube), surrounding the scene. The point on the closed surface can be parameterized by two variables that depict the spatial position. With the assumption of a convex scene, a ray emanating from each scene point can be constructed that intersects the closed surface. For each point on the closed surface,
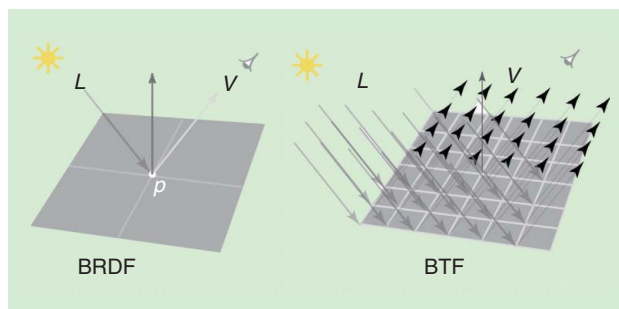
FIGURE 7. The BRDF describes light reflected from a surface point in a hemisphere of possible directions, due to light incident to a surface point at a particular angle. However, light incident at a point may be interreflected and may be partially transmitted and scattered resulting in light exiting the surface at multiple points. The BTF describes light exiting a surface patch due to light incident on the patch at a specified angle accounting for interreflections and subsurface scattering.

two variables (polar angle and azimuth angle) denote the illumination emanating from this point. Plenoptic functions are similarly defined as radiance as a function of position and exitant angle. Plenoptic functions [1] are five-dimensional (5-D) since the point can vary in 3-D space, $f(x, y, z, \theta, \phi)$ and no convexity of the scene is assumed. A common parameterization of the light rays in a light field uses two planes. Each ray emanating from the scene can be described by its 2-D intersection on two planes. Early light field rigs used a set of cameras on a plane where each camera captures many pixels (positions of the scene). Commercially available light field cameras such as Lytro use arrays of microlenses to accomplish the same effect [32], [41]. These cameras enable refocusing by recording each surface point from slightly different viewing directions in a large image format (approximately 15 MB per image). The light field camera is not designed for capturing BRDF and the angular range of local viewing directions per scene point is relatively small.

The BTF captures a spatially varying BRDF as discussed in [8]. A related concept is svBRDF which also represents a spatially varying BRDF. The BTF and svBRDF are both parameterized by a six-dimensional (6-D) function (two dimensions for spatial coordinates and four dimensions for illumination and viewing parameters). The svBRDF [31] is defined to represent an object where no texture is present, therefore the notion of a patch of interest is not relevant. The standard taxonomy usage is BTF for textured objects and svBRDF where the variations are not due to texture. The notion of local planarity is often not used in svBRDF so that curved objects without local 2-D planar embedding can be described. The BTF is used where the local variation in the patch represents the object's appearance. For example, a BTF can be used to represent tree bark, animal fur, leather, foliage, and roughness; where the appearance of a patch has the same texture as the object. The svBRDF is more often used in situations where there is no characteristic variation over the object. That is, svBRDF typically describes the variation of reflectance over an object, and BTF characterizes the reflectance variation over a patch to capture the texture over the object. Additionally, BTF representations account for local cast shadows and occlusions within the patch-based measurement allowing for nonconvex, nonsmooth fine-scale surface geometry, while svBRDF treats each point independently.

The BSSRDF [27] is another representation that is more general than the BRDF. This representation is especially useful for

surfaces that have translucency such as marble and human skin where the incident light at the surface is scattered within the volume. BSSRDF differs from the BRDF in that the latter implicitly assumes all incident light is reflected from the incident surface point. While BTF measurements account for subsurface scattering in translucent surfaces by illuminating and imaging a surface patch, BSSRDF measurements are typically done by illuminating a surface point and then measuring the light reflected from an area or patch.

Table 1 summarizes several of these reflectance representations. The 8-D reflectance field representation considers globally incident light and exitant light, the 4-D light field representation considers exitant light, the 5-D plenoptic function considers exitant light from a 3-D surface point, the 6-D BTF represents incident light at a surface patch and exitant light at this patch, and the 4-D BRDF represents incident and exitant light at a point.

Notice that incident illumination is accounted for in different ways among BRDF/BTF, light fields, and reflectance fields. For BRDF and BTF the measured values are a ratio of radiance to irradiance. The incident and exitant illumination is specified by their respective angular directions. Light fields are a measure of radiance and do not explicitly account for incident illumination. Eight-dimensional reflectance fields [15] are bidirectional in the sense that they take into account incident illumination and exitant radiance. In BRDF/BTF measurements, there is typically an implied assumption that incident illumination has uniform intensity and the measured intensity can be normalized by the uniform incident intensity to obtain the desired ratio. When measuring reflectance representations, incident light must either be known (as in controlled active lighting) or measured (as in silvered spheres for capturing environment light).

The relationship between incident light to exitant light from the global scene can also be written as a light transport equation [26], [29]. When this relation is written as a matrix function, relating a plane parameterized incident light and reflected light, we obtain the following light transport equation:

$$L_{\text{out}} = TL_{\text{in}} \tag{1}$$

where L_{in} is the vectorized 4-D input light field (or 2-D for a fixed illumination direction), L_{out} is the vectorized 4-D light field (or 2-D for a fixed camera position), and T is the transport matrix capturing the aggregate scene properties that relates input light to output light.

Instead of volumetric light measurement (3-D position and 3-D direction) for both incident and reflected illumination, the notion of reflection from a surface or scene allows a natural dimensionality reduction. For reflectance fields and light field, this reduction is done by considering light through a closed surface. For BTFs, the measurements consider light incident on and reflected from a local surface patch. Additionally, although light is an electromagnetic field that is vector-valued and depends on wavelength, further inherent dimensionality reduction is accomplished by considering intensity measurable by the camera's sensitivity functions. In this manner, the

Table 1. A comparison of the BRDF, BTF, plenoptic function, light field, and reflectance field.

	Degrees of Freedom	Bidirectional?	Local/Global
BRDF	4	Y	L
BTF	6	Y	L
Plenoptic function	5	N	G
Light field	4	N	G
Reflectance field	8	Y	G

camera acts as a light probe to measure radiance after a photometric calibration stage to relate measured image pixel values to scene intensity.

Capturing appearance

Capturing appearance densely—Robots and domes

Traditionally BRDF was measured with large devices called *gonioreflectometers* with a computer controlled light source and a photometer moved on a hemispherical gantry. Since radiometrically calibrated cameras can also measure light, BRDF methods using images from cameras calibrated to relate pixel values to radiometrically calibrated intensity replaced the traditional gonioreflectometer [11], [34], [36]. Reflectance is measured using multiple images of a surface taken with a camera moved with a

robotic arm or multiple cameras fixed at different positions (e.g., vertices of a geodesic dome). Illumination direction is controlled using a moving light source or a light dome [55]. BRDFs are also captured using objects of known geometry such as uniformly painted spheres imaged with a camera and a point source moving along a circular arc [36]. Human face and skin images taken from multiple viewing directions and under multiple illumination conditions have been used to obtain reflectance fields for successful recognition and rendering [10], [15], [63]. A comprehensive survey of BRDF/BTF capture devices is provided in [23]. Examples of reflectance measurement devices for computational appearance are depicted in Figure 8.

Both robotics-based measurements and dome-based measurements approach the measurement problem by densely sampling the hemisphere with as many samples as possible with
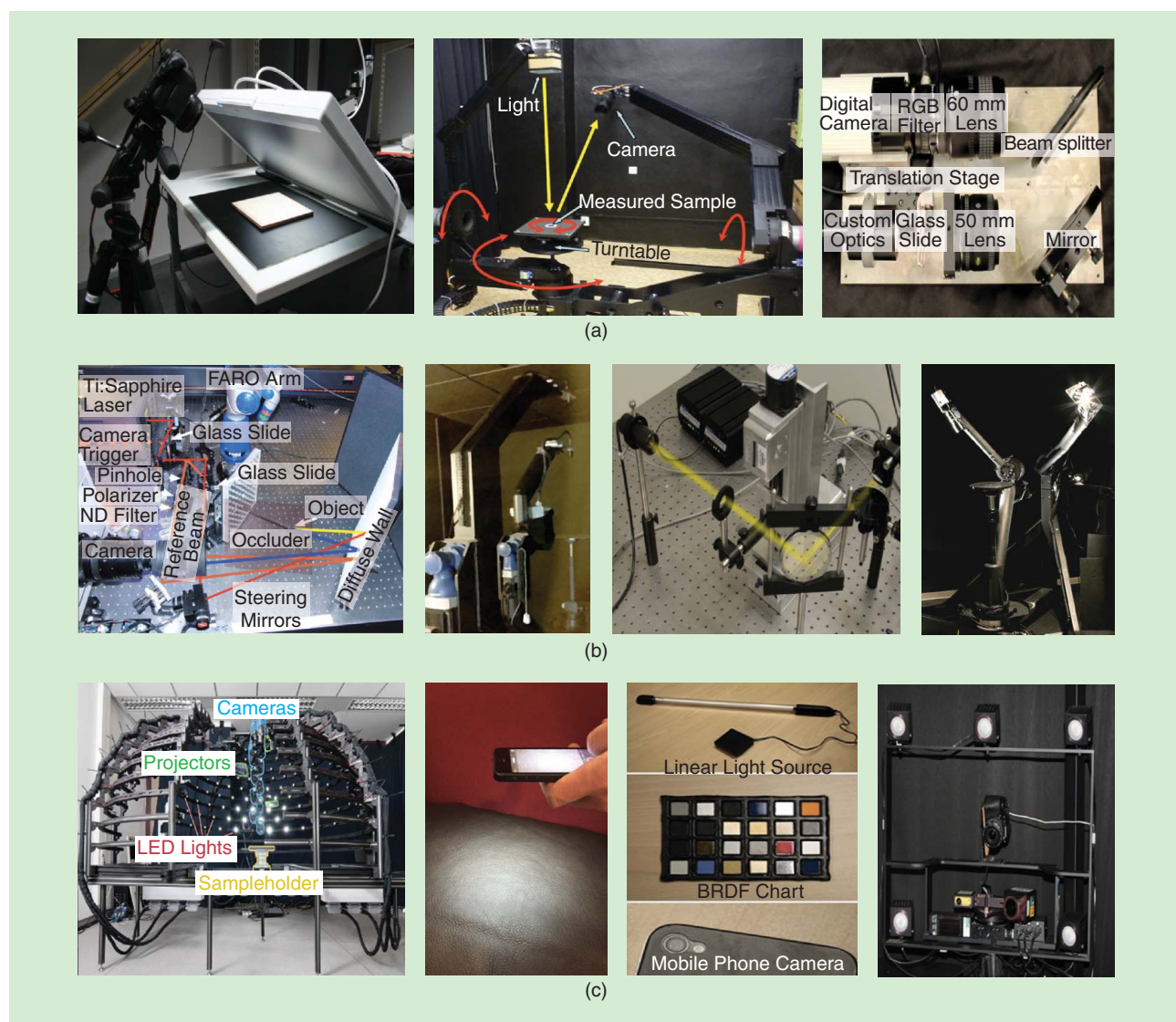
FIGURE 8. A sampling of devices to capture computational appearance. (a) From left: [2], [19], and [25]. (b) From left: [38], [53], [14], and [35]. (c) From left: [56], [3], [50], and [18]. All the devices shown measure the scene or surface as a function of viewing and illumination direction. ([2] © 2013 Association for Computing Machinery, Inc. Reprinted by permission. [25] ©2010 Association for Computing Machinery, Inc. Reprinted by permission. [38] ©2011 Association for Computing Machinery, Inc. Reprinted by permission. [53] ©2013 Association for Computing Machinery, Inc. Reprinted by permission. [3] ©2015 Association for Computing Machinery, Inc. Reprinted by permission. [50] ©2011 Association for Computing Machinery, Inc. Reprinted by permission.)

the given device. Lab-based approaches to reflectance measurement can be cumbersome, time-consuming, and nonportable. Dense angular sampling to capture the hemisphere of possible viewing directions θ_v, ϕ_v for each illumination direction θ_i, ϕ_i leads to extremely large BRDF measurement sets. However, with coarse sampling of the angular space important features of the reflectance function may be lost. Current trends in appearance capture improve on the initial dome and robotic based measurements to provide appearance capture that is fast, portable, sparse, and low dimensional.

Capturing appearance quickly—Novel cameras

Mirror-based cameras take advantage of a mirror's ability to redirect light rays in a manner that is dependent on the mirror geometry. In particular, parabolic mirrors have the property of focusing parallel light to a single point. The texture camera [14] is a device that uses a concave off-axis parabolic mirror to replace the angular movements required in a gonioreflectometer. Parallel light incident on the mirror surface is focused to a point; a planar aperture placed in the incident light field that selects a single light ray effectively chooses an angle of illumination. Motion of the planar aperture in a plane provides angular indexing of the incident light rays without the need for hemispherical motions. Since planar motion of an aperture is easier to accomplish than hemispherical motion, this design is very convenient. In a dual manner, reflected light from the surface is redirected by the parabolic mirror to a set of parallel rays. Placing a camera with a telecentric lens in the path of this set of parallel light rays provides an unusual image. It is the image of a single surface point, but from multiple viewing directions. In this manner, the camera records a multiview image of a surface point where each pixel records a different angle. While the basic operational principles can be discussed with ray optics, diffraction effects and limits on measurable angular range have also been evaluated [14]. By

FIGURE 9. Samples of the BRDF (multiple viewing directions) that show the iridescence of a peacock feather that is not observable in a single view photo. Notice the abrupt changes in color and intensity as the viewing direction changes. We see this iridescence when we tilt our heads or the feather when viewing in person [65].

scanning illumination directions and surface points, the device can measure a surface BTF and is used as a texture camera. This approach has recently been used as a reflectance sensor to collect reflectance disks as shown in Figure 9.

Point-and-shoot reflectance capture provides an opportunity to use BRDF (or a sampling of BRDF) in mobile applications. Reflectance disks are instantaneous snapshots of point reflectance obtained with small concave parabolic mirrors. The reflectance disks of multiple surface points for an iridescent sample are shown in Figure 9. Notice the reflectance disks capture the large change in surface reflectance with viewing angle (observable by color and intensity changes in the reflectance disks). Recent work [65] shows these reflectance disks are convenient to estimate angular gradients that can be effective in recognizing material classes. One-shot reflectance capture with portable devices is an important goal for acquiring data from scenes in a convenient way, to build data-driven models. Two shot reflectance capture (flash and no-flash) [3] enables BTF capture with a simple lightweight mobile device.

Another mirror-based camera is a multiview radial imaging system that obtains a dense sampling of viewing directions using a conical curved mirror placed in the light path of the camera [30]. This device also samples multiple viewpoints with a single snapshot. The system can be used to measure BRDF and recover both the geometry and the texture map of 3-D objects.

In addition to novel mirror-based cameras, lenses have been redesigned for the purpose of capturing reflectance or appearance. For example, using a pair of condenser lenses [17] also leads to a fast and compact reflectance sensor capable of measuring a BRDF with a solid angle near 90°. Another novel approach is BRDF measurement without any camera that is accomplished with a device consisting of a dome of light-emitting diode lights used as both the illuminants and photodiodes [6].

Light-field capture devices include lens array methods [32], [41], and methods that use masks instead of lenses for multiangle capture such as dappled photography [59]. This approach uses attenuation masks placed between the lens and sensor to capture weighted sums of the rays. Then the rays are reconstructructed via a decoding process. This concept of capturing weighted sums of rays from multiple angles is also referred to as *multiplexing* and has been used to capture light fields using illumination multiplexing [54] and polarization multiplexing [9].

Capturing appearance remotely

While mirror-based systems for reflectance have the advantage of using a single viewpoint to capture reflectance over multiple angles, the approach has the disadvantage of requiring an up-close observation where the distance from camera to surface is small. An issue with measuring a BRDF from a distance is that the need for hemispherical variation of the light source or viewing direction prevents large ranges of angular reflectance measurements. Several methods exist for one-shot reflectance capture, but they are limited either in the range of angles or by requiring close-up reflectance capture. Time-of-flight imaging has been used to get dense reflectance measurements from a distance [38]. This method takes advantage of ultra-fast
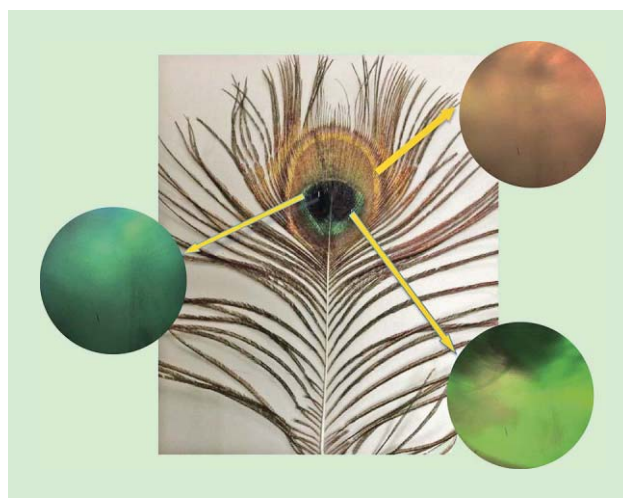
imaging in time-of-flight cameras. Different angles of exiting illumination from a surface that correspond to different path lengths are imaged at different times by the camera. The measurement of reflectance as a function of time can be mapped to different points on the surface. High-quality reconstructions of BRDFs have been demonstrated with this approach.

Capturing appearance optimally

Full appearance, even for a single surface point, is a very high-dimensional signal, especially when illumination and viewing directions are densely sampled. If every surface point is captured by densely sampling the hemisphere of possible illumination directions and viewing directions, this angular sampling gives a terabyte or more of image data at the same spatial resolution as a one megabyte traditional image. Representing high-dimensional appearance compactly is a major challenge to enable efficient capture, matching, and discrimination. Unlike standard methods of signal compression after capture, the goal of capturing appearance optimally is to avoid sampling the full appearance signal to achieve fast and efficient appearance capture.

An important question that arises in measuring reflection is: Which illumination and viewing directions matter most? Are there optimal sampling strategies that can be used to sub-sample the measurement space in a meaningful way? This question has been addressed by empirical studies on comprehensive data sets. Recent work [45] uses the MERL BRDF database [42] to determine which samples are most useful by computing the accuracy of the reconstructed the BRDFs using principal components analysis. After determining which samples are optimal, capturing only those measurements increases acquisition speed and decreases required storage requirements.

Efficient capture of the BRDF with few samples is accomplished by an isotropic approximation of the BRDF represented as a 2-D bivariate function [4]. For many approaches, the reparameterization of the BRDF as described in [52] provides a compact representation and reduces the number of basis coefficients. Lab-based BRDF-slice measurements with light-domes are used to capture a sampling of the BRDF for material classification [19], [56]. Basis illumination [21] is another approach to compress appearance at capture time. This method illuminates the scene with spherical basis functions of illumination. By expressing signals as linear combinations of basis measurements, BRDF acquisition with basis illumination requires significantly less time than dense angular sampling.

Another framework for efficient appearance capture uses the Helmoltz reciprocity principle ,which states that reflectance will be the same when the input and output light intensity is reversed. Dual photography [57] uses Helmholtz reciprocity to capture with a camera, illuminate with a light source, but then computes the image that results from switching the positions of camera and light source. This property has significant

practical importance because the scene reflectance from individual illuminants cannot typically be captured with an illumination array. That is, the reflection from a single illumination source cannot be isolated. However, an array of cameras can be used without an intereference problem and therefore dual photography can achieve the effect of multi-illuminant capture with camera arrays.

Appearance capture can also be cast as a compressive sensing problem, and sparse methods in signal processing have been used to efficiently measure 8-D reflectance fields. Symmetric photography [20] uses data sparsity of reflectance fields to make the process of measuring the entire 8-D function manageable. The method can be used to obtain virtual views of the scene by assuming the light transport matrix is symmetric; therefore transport coefficient T_{ji} need not be measured but can be estimated as the measured T_{ij}. Compressive dual photography [58] uses compressive sensing and the properties of L1-norm optimization to achieve a sparse result and exploits the compressibility of the signal in a transform domain to speed acquisition. Sparse sampling combined strategically with dense sampling is the approach of manifold bootstrapping [17]. This method samples spatially at a low angular density and samples sparse key points at a high angular density using a handheld device and 10–20 minute scanning sessions.

> Beyond basic object recognition, example applications for computational appearance are numerous and include e-commerce, art archiving, robotic navigation, architecture, and biomedical applications such as dermatology.

Reflectance hashing [65] is another method for managing high-dimensional appearance. In this method, surface reflectance is obtained using an off-axis concave parabolic mirror. These one-shot reflectance measurements are converted to binary codes in an optimization that preserves discriminative characteristics for recognizing material classes. The approach is supervised so that training reflectance data with associated material class labels must be known. The binary codes are optimized to reduce the error cost function over the training set so that the Hamming distance is close to the Euclidian distance that would be used in a nearest neighbor classification. The result is a compact binary code to represent the high-dimensional reflectance function and fast recognition is supported because the Hamming distance is very efficient to compute.

Appearance applications beyond recognition and rendering

While recognition and rendering are the traditional applications for computational appearance, there is a large set of novel applications for computational appearance in diverse application domains.

Computational appearance in dermatology

For biomedical applications, quantitative dermatology can use computational appearance captured using variations of illumination and camera pose to assess change. In standard

FIGURE 10. Basal cell carcinoma as seen under a set of illumination angles. Each panel depicts the same patient at the same time. Only the illumination angle varies, yet the structural change in appearance is significant [10].

dermatology, the visual impression of the clinical dermatologist is the main tool of initial diagnosis. A typical question in evaluating subtle change may be: Has a mole changed by 3% over the course of a year? A clinical practitioner may have difficulty in making such a precise quantitative assessment based on observation. Figure 10 illustrates an example of the dependence of skin appearance on viewing and illumination direction [9]. This image set shows basal cell carcinoma where each image depicts the same skin region captured under different illumination directions. Notice that the appearance changes significantly across the image set, revealing unique surface structures. For example, the raised translucent border is only visible in a subset of the illumination directions. These images are multiangle reflectance measurements that capture skin appearance and provides a digital representation to store, share, and compare appearance in a computationally meaningful way.

> An interesting paradigm is machine learning affecting appearance capture, driving what aspects of appearance are most relevant for the task at hand.

Separating reflectance components

Light in a scene originates not only from the primary source but also from light scattering due to translucency and from interreflections within the scene. Fast separation of direct and indirect reflectance [40] uses high-frequency binary illumination patterns projected on the scene to recover reflection of the light source separately from reflection due to scattering and interreflections. This method creates two components of appearance that can be used for scene analysis such as detecting shadowed areas and areas obscured by translucent objects. Primal-dual coding [47] also projects illumination patterns onto the scene and modulates the light capturing elements so that there is control over which light paths (between projector and camera) contribute to the image and how much they are modulated. By controlling the relative modulations of the direct and indirect light components, this method creates a tuned image that can rebalance the different scene appearance components. Structured light transport [46] creates a live video stream of indirect reflection using a high-speed imaging system comprising a digital micromirror device (DMD) projector and DMD mask. The method projects a pattern and places a modulating pixel mask at the camera. Where

primal-dual coding [47] uses liquid crystal displays (LCDs) for modulation, structured light transport uses DMDs and, therefor, the projection masks and pixel masks are binary.

Estimating geometry, reflectance, and illumination

In general, an image depends on object geometry (shape), reflectance, and illumination. If two are known, the other can be measured directly. Alternatively, when more than one of the elements are unknown, auxilliary information can be used for estimation. Just as binocular stereo can be used to determine geometry of a scene, photometric stereo uses multiple light sources to determine local surface shape in terms of local surface normals. Implicit in traditional methods of photometric stereo is a Lambertian reflectance model. Lambertian reflectance provides a simple expression for reflectance as a function of the incident illumination direction and the surface normal. Measured reflectance for known illumination directions can be used to create a set of equations to estimate the surface normal. More recent methods remove the restriction that the reflectance must be Lambertian and consider unknown reflectance. More generally, recent research considers what can be simultaneously estimated about reflectance, illumination, and shape.

When surface shape and lighting direction are known, the reflectance can be directly measured according to the reference coordinate frame defined by the local surface normal. In this case, an image of the surface provides the reflectance and the angles $\theta_i, \phi_i, \theta_v, \phi_v$ are known with respect to the given surface normals (known shape) of the object. Recent work in reflectance and illumination recovery [33] addresses the problem of estimating object reflectance and scene illumination from a single image when the object geometry is known. Algorithms in this area can also predict both the surface reflectance and the surface geometry simultaneously when the illumination is known [49]. Joint reflectance and shape can be estimated [48] when the illumination is known but uncontrolled as in measured natural illumination. Estimation of a homogenous BRDF from a single image of a known shape in unknown real-world illumination can be accomplished by utilizing natural image statistics [51]. In the case of known shape and unknown

illumination, the motion a rotating object can also enable recovery of the object BRDF. [16]. For most of these estimation methods, results are evaluated by comparing renderings to ground truth. That is, the fidelity of the representation when compared to ground truth is the performance metric.

Hiding information within appearance

When considering appearance as measured by reflectance, novel opportunities arise to hide information in the angular space of reflected light. In this context, the high dimensionality of the local appearance is an advantage. Using changes of refractive index, [13] presents a unique method to create transparent watermarks that are hidden to the human eye but can be detected by a multiview camera. By encoding a watermark with different refractive indices, the minimum reflected intensity as defined by the Brewster angle occurs at a different angle for different spatial regions of the watermark. This subtle effect cannot be visualized unless multiple angles are viewed simultaneously as done with a curved mirror-based reflectance capture device. Another approach to hiding information using reflectance angular space is Bokode [37], where a single image can capture a hidden message embedded as a spatial pattern and a lenslet. The key idea is that refocusing the image from the captured reflectance reveals the message. Photographic steganography [64] refers to hiding information in electronically displayed images that can be decoded with a camera receiver either using intensity [5], [61], [64], high-frequency modulation [28], [43], or color [60].

Conclusions

Measuring apparatus for appearance has evolved past ordinary cameras to devices for more general light capture. Specialized cameras, devices, and algorithms can more fully capture patterns of light from a scene. Dense sampling of reflectance with robotic devices and light/camera domes has evolved to computational imaging procedures for faster capture in a more compact device and a more compact representation. A large majority of the appearance capture methods make improvements that are generically useful (e.g., faster, more compact, more complete). An open research topic for future trends is tuning computational appearance for a particular task, creating a feedback from application/algorithm to capture method. For example, the image signal processors within a camera that handle demosaicing, denoising, and other image processing tasks have been optimized in an end-to-end manner [24] to optimize image quality. There are many more possibilities in the concept of combining optimization and appearance capture. Consider machine-learning methods such as deep learning and convolutional neural networks that can use appearance measurements as input. An interesting paradigm is machine learning affecting appearance capture, driving what aspects of appearance are most relevant for the task at hand. Since the computational camera is a tunable device, the question of how to tune it leads to new cost functions that balance algorithm requirements, hardware constraints, physical and optical constraints, and application goals. Future trends for appearance capture are not replicating what the eye would see, but rather tuning the sampling of light to meet what an algorithm requires.

Author

Kristin J. Dana (kdana@ece.rutgers.edu) received her B.S. degree in 1990 from the Cooper Union, New York; her M.S. degree from the Massachusetts Institute of Technology in 1992; and her Ph.D. degree from Columbia University, New York, in 1999. She is a full professor in the Department of Electrical and Computer Engineering at Rutgers University. Her research interests in computer vision include computational photography, machine learning, illumination modeling, texture and reflectance, bioimaging, motion estimation, optical devices, optimization in vision, and applications of robotics. She is the recipient of the 2001 National Science Foundation Career Award and a team recipient of the Charles Pankow Innovation Award in 2014 from the American Society of Civil Engineers.

References

[1] E. H. Adelson and J. R. Bergen, *Computational Models of Visual Processing*. Cambridge, MA: MIT Press, 1991, pp. 3–20.

[2] M. Aittala, T. Weyrich, and J. Lehtinen, "Practical SVBRDF capture in the frequency domain," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 110:1–110:12, July 2013. [Online]. Available: http://doi.acm.org/10.1145/2461912.2461978

[3] M. Aittala, T. Weyrich, and J. Lehtinen, "Two-shot SVBRDF capture for stationary materials," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 110:1–110:13, July 2015. [Online]. Available: http://doi.acm.org/10.1145/2766967

[4] M. A. Ali, I. Sato, T. Okabe, and Y. Sato, "Toward efficient acquisition of BRDFs with fewer samples," in *Computer Vision–ACCV 2012*, New York: Springer, 2013, pp. 54–67.

[5] A. Ashok, S. Jain, M. Gruteser, N. Mandayam, W. Yuan, and K. Dana, "Capacity of screen–camera communications under perspective distortions," *Pervasive Mob. Comput.*, vol. 16, pp. 239–250, Jan. 2015.

[6] M. Ben-Ezra, J. Wang, B. Wilburn, X. Li, and L. Ma, "An LED-only BRDF measurement device," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[7] J. F. Blinn and B. T. Phong, "Models of light reflection for computer synthesized pictures," in *Proc. 4th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '77*, 1977, pp. 192–198.

[8] O. G. Cula and K. J. Dana, "Compact representation of bidirectional texture functions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 1041–1067.

[9] O. G. Cula, K. J. Dana, and D. K. Pai, and D. Wang, "Polarization multiplexing for bidirectional imaging," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2005, pp. 1116–1123.

[10] O. G. Cula, K. J. Dana, F. P. Murphy, and B. K. Rao, "Bidirectional imaging and modeling of skin texture," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2148–2159, Dec. 2004.

[11] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," Columbia University, New York, Tech. Rep. CUCS-048-96, Dec. 1996.

[12] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.

[13] K. J. Dana, G. Livescu, and R. Makonahalli, "Transparent watermarking using bidirectional imaging," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*, 2009, pp. 31–38.

[14] K. J. Dana and J. Wang, "Device for convenient measurement of spatially varying bidirectional reflectance," *J. Optical Soc. Am. A*, vol. 21, pp. 1–12, Jan. 2004.

[15] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. "Acquiring the reflectance field of a human face," in *Proc. 27th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '00*, New York, 2000, pp. 145–156.

[16] Y. Dong, C. Guojun, P. Pieter, Z. Jiawan, and T. Xin, "Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 193, 2014.

[17] Y. Dong, J. Wang, X. Tong, J. Snyder, Y. Lan, M. Ben-Ezra, and B. Guo, "Manifold bootstrapping for SVBRDF capture," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 98, July 2010.

[18] M. Farouk, I. El-Rifai, S. El-Tayar, H. El-Shishiny, M. Hosny, M. El-Rayes, J. Gomes, F. Giordano, H. Rushmeier, F. Bernardini, and K. Magerlein, "Scanning and processing 3D objects for web display," in *Proc. 3DIM, Fourth Int. Conf. 3D Digital Imaging and Modeling*, 2003, pp. 310–317.

[19] J. Filip, R. Vavra, M. Haindl, P. Id, M. Krupika, and V. Havran, "BRDF slices: Accurate adaptive anisotropic appearance acquisition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 1468–1473.

[20] G. Garg, E.-V. Talvala, M. Levoy, and H. P. Lensch. "Symmetric photography: Exploiting data-sparseness in reflectance fields," in *Proc. 17th Eurographics Conf. Rendering Techniques Eurographics Association*, 2006, pp. 251–262.

[21] A. Ghosh, W. Heidrich, S. Achutha, and M. OToole, "A basis illumination approach to BRDF measurement," *Int. J. Computer Vision*, vol. 90, no. 2, pp. 183–197, 2010.

[22] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. "The lumigraph," in *Proc. 23rd Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '96*, New York, 1996, pp. 43–54.

[23] M. Haindl and J. Filip, *Visual Texture: Accurate Material Appearance Measurement, Representation, and Modeling.* New York: Springer, 2013, pp. 23–62.

[24] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. P. Pulli, "Flexisp: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 231, 2014.

[25] M. Holroyd, J. Lawrence, and T. Zickler, "A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance," *ACM Trans. Graph.*, vol. 29, no. 4, p. 99, 2010. [Online]. Available: http://doi.acmorg/10.1145/1778765.1778836

[26] D. S. Immel, M. F. Cohen, and D. P. Greenberg, "A radiosity method for non-diffuse environments," *ACM SIGGRAPH Comp. Graph.*, vol. 20, no. 4, pp. 133–142, 1986.

[27] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan, "A practical model for subsurface light transport," in *Proc. 28th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '01*, New York, 2001, pp. 511–518.

[28] K. Jo, M. Gupta, and S. K. Nayar, "Disco: Displays that communicate," Dept. Comput. Sci., Columbia Univ., Tech. Rep. CUCS-028-14, pp. 1–20, Dec. 2014.

[29] J. T. Kajiya, "The rendering equation," *ACM SIGGRAPH Comp. Graph.*, vol. 20, no. 4, pp. 143–150, 1986.

[30] S. Kuthirummal and S. K. Nayar, "Multiview radial catadioptric imaging for scene capture," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 916–923, July 2006.

[31] H. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, "Image-based reconstruction of spatially varying materials," in *Proc. 12th Eurographics Conf. Rendering, Eurographics Association*, 2001, pp. 103–114.

[32] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, New York, 1996, pp. 31–42.

[33] S. Lombardi and K. Nishino, "Reflectance and illumination recovery in the wild," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 1, pp. 129–141, 2016.

[34] S. R. Marschner, S. H. Westin, E. Lafortune, and K. E. Torrance, "Image-based bidirectional reflectance distribution function measurement," *Appl. Optics*, vol. 39, no. 16, pp. 2592–2600, June 2000.

[35] S. R. Marschner, S. H. Westin, A. Arbree, and J. T. Moon, "Measuring and modeling the appearance of finished wood," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 727–734, 2005.

[36] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 759–769, 2003.

[37] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar, "Bokode: Imperceptible visual tags for camera based interaction from a distance," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 98:1–98:8, July 2009.

[38] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala, "Single view reflectance capture using multiplexed scattering and time-of-flight imaging," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 171:1–171:10, Dec. 2011.

[39] S. K. Nayar, "Computational cameras: Redefining the image," *IEEE Comp.*, vol. 39, no. 8, pp. 30–38, 2006.

[40] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 935–944, 2006.

[41] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Dept.

Computer Science, Stanford Univ., Stanford, CA, Tech. Rep. CSTR 2005-02, Apr. 2005.

[42] A. Ngan, F. Durand, and W. Matusik, "Experimental analysis of BRDF models," *Rendering Tech.*, vol. 2005, no. 16, pp. 2, 2005.

[43] V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wengrowski, and N. Mandayam, "High-rate flicker-free screen-camera communication with spatially adaptive embedding," in *Proc. IEEE Int. Conf. Computer Communications*, 2016.

[44] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. *Geometric Considerations and Nomenclature for Reflectance.* National Bureau of Standards (US), vol. 160, 1977.

[45] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi, "On optimal, minimal BRDF sampling for reflectance acquisition," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 186:1–186:11, Oct. 2015.

[46] M. O'Toole, J. Mather, and K. Kutulakos, "3D shape and indirect appearance by structured light transport," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 3246–3253.

[47] M. O'Toole, R. Raskar, and K. N. Kutulakos, "Primal-dual coding to probe light transport," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 39–31, 2012.

[48] G. Oxholm and K. Nishino, "Shape and reflectance estimation in the wild," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 2, pp. 376–389, Feb. 2016.

[49] G. Oxholm and K. Nishino, "Shape and reflectance from natural illumination," in *Computer Vision–ECCV 2012*, Heidelberg, Germany: Springer, 2012, pp. 528–541.

[50] P. Ren, J. Wang, J. Snyder, X. Tong, and B. Guo, "Pocket reflectometry," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 45, 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1964940

[51] F. Romeiro and T. Zickler, "Blind reflectometry," in *Proc. Computer Vision–ECCV*, 2010, pp. 45–58.

[52] S. M. Rusinkiewicz, "A new change of variables for efficient brdf representation," in *Proc. Rendering Techniques*, 1998, pp. 11–22.

[53] I. Sadeghi, O. Bisker, J. De Deken, and H. W. Jensen, "A practical microcylinder appearance model for cloth rendering," *ACM Trans. Graph.*, vol. 32, no. 2, pp. 14:1–14:12, Apr. 2013.

[54] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," in *Proc. 9th IEEE Int. Conf. Computer Vision*, 2003, pp. 808–815.

[55] C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein, "Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the University of Bonn," *Sensors*, vol. 14, no. 5, pp. 7753, 2014.

[56] C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein, "Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the University of Bonn," *Sensors*, vol. 14, no. 5, pp. 7753–7819, 2014.

[57] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. Lensch, "Dual photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 745–755, 2005.

[58] P. Sen and S. Darabi, "Compressive dual photography," *Comp. Graph. Forum*, vol. 28, no. 2, pp. 609–618, 2009.

[59] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69, 2007.

[60] E. Wengrowski, K. Dana, M. Gruteser, and N. Mandayam. (2016). *Reading between the pixels: Photographic steganography for camera display messaging.* [Online]. Available: arXiv:1604.01720

[61] E. Wengrowski, W. Yuan, K. Dana, A. Ashok, M. Gruteser, and N. Mandayam, "Optimal radiometric calibration for camera-display communication," in *Proc. IEEE Winter Conf. Applications of Computer Vision*, 2016.

[62] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational plenoptic imaging," *Comp. Graph. Forum*, vol. 30, no. 8, pp. 2397–2426, 2011.

[63] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross, "Analysis of human faces using a measurement-based skin reflectance model," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1013–1024, July 2006.

[64] W. Yuan, K. Dana, A. Ashok, M. Varga, M. Gruteser, and N. Mandayam. "Photographic steganography for visual mimo: A computer vision approach," in *Proc. IEEE Workshop on the Applications of Computer Vision*, 2012, pp. 345–352.

[65] H. Zhang, K. Dana, and K. Nishino, "Reflectance hashing for material recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3071–3080.

SP

Aswin C. Sankaranarayanan, Pavan Turaga,
Matthew A. Herman, and Kevin F. Kelly

Enhanced Compressive Imaging Using Model-Based Acquisition

Smarter sampling by incorporating domain knowledge
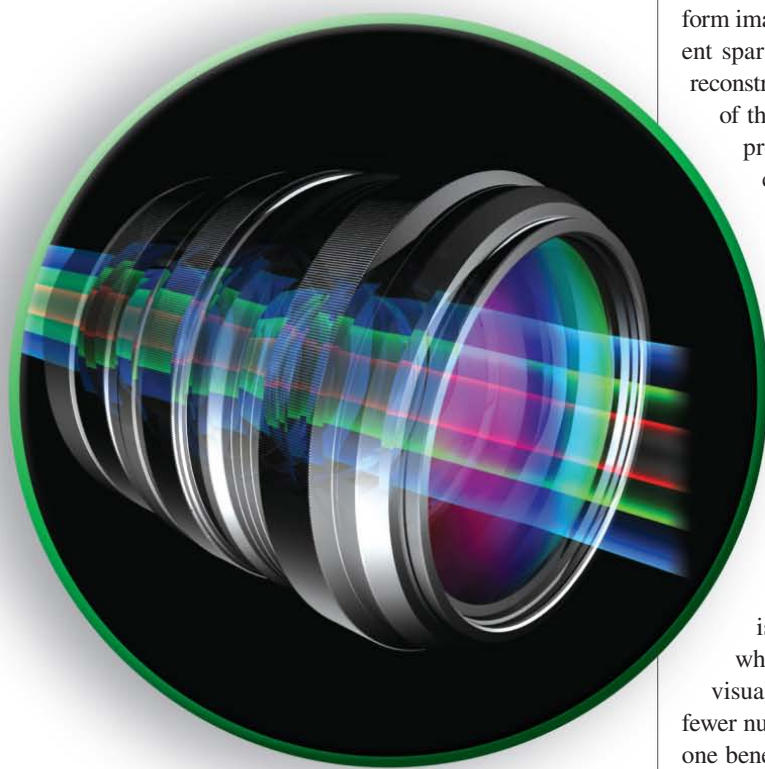
©ISTOCKPHOTO.COM/YAKOBCHUK

ompressive imaging (CI) is a subset of computational photography where a scene is captured via a series of optical, transform-based modulations before being recorded at the detector. However, unlike previous transform imagers, compressive sensors take advantage of the inherent sparsity in the image and use specialized algorithms to reconstruct a high-resolution image with far lower than 100% of the total measurements. Initial CI systems exploited the properties of random matrices used in other areas of compressive sensing (CS); however, in the case of imaging, there are immense benefits to be derived by designing measurement matrices that optimize specific objectives and enable novel capabilities. In this article, we survey recent results on measurement matrix designs that provide the ability of real-time previews, signature-selective imaging, and reconstruction-free inference.

Compressive imaging

The last decade has seen rapid advancements in computational imaging, especially in the context of high-dimensional acquisition in resource-constrained regimes. At the forefront of these advances is the idea of CI [2], the optical embodiment of CS, which enables one to sense images, videos, and other visual signals compressively, i.e., sensing a signal from far fewer number of measurements than its dimensionality. Thus, one benefit of such a system is that the measurement process simultaneously compresses as it acquires the image. More importantly, such a design allows far cheaper imaging outside the visible spectrum, such as the infrared wavelengths. While silicon-based imagers cost less than US$100 for tens of megapixels, infrared imaging systems that are not silicon based are thousands of dollars, even for far less than a megapixel resolution, and tens of thousands of dollars for a megapixel or more. One can use the same imaging system and mathematics to also exploit sparsity in time and turn a visible video camera system into a high-speed one [28]. Finally, CI systems

can perform image-sensing tasks with far less overhead in processing, transmission, and data storage than the current camera-based counterparts. While the examples in this article focus on images and videos, the use of such CI systems for depth and hyperspectral acquisition can further leverage such benefits and are also being explored by many researchers.

Specifically, CI (and, more generally, CS) aims to sense a signal $\mathbf{x} \in \mathbb{R}^N$ from an underdetermined linear system, i.e., measurements of the form

$$\mathbf{y} = \Phi\mathbf{x} + e, \tag{1}$$

where $\Phi \in \mathbb{R}^{M \times N}$ with $M < N$, and e is the measurement noise. For an arbitrary signal in $\mathbb{R}^N$, this is impossible since the map $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^M$ is many to one and noninvertible. CS handles this by restricting the signal $\mathbf{x}$ to belong to a distinguished class, e.g., signals that are sparse. The main results of CS state that when the measurement matrix Φ has a special structure and $\mathbf{x}$ is K-sparse (or having K or fewer nonzero entries), then we can robustly recover $\mathbf{x}$, provided that M is sufficiently large. These basic results have been extended beyond real-space sparse imaging to include signals that are sparse in a transform domain, have sparse gradients [26], low-rank matrices [10], and low-dimensional manifolds [5].

A classic example of CI in practice is the single pixel camera (SPC), which consists of an optical modulator and a single photodetector that obtains coded linear or compressive measurements of the scene. A schematic and further description of the SPC is highlighted in "Signal Pixel Camera Basics." The compressive measurements $y_t \in \mathbb{R}$ taken by an SPC at the sample instants $t = 1, \ldots, T$ can be modeled as $y_t = \langle \phi_t, \mathbf{x}_t \rangle + e_t$, where T is the total number of acquired samples, $\phi_t \in \mathbb{R}^{N \times 1}$ is the measurement vector, $e_t \in \mathbb{R}$ represents measurement noise, and $\mathbf{x}_t \in \mathbb{R}^{N \times 1}$ is the scene (or frame) at sample instant t. We assume that the two-dimensional (2-D) scene consists of $n \times n$ spatial pixels that, when vectorized, results in the vector $\mathbf{x}_t$ of dimension $N = n^2$. We also use the notation $\mathbf{y}_{1:W}$ to represent the vector consisting of a window of $W \le T$ successive compressive measurements (samples), i.e.,

$$\mathbf{y}_{1:W} = \begin{bmatrix} y_1 \\ \vdots \\ y_W \end{bmatrix} = \begin{bmatrix} \langle \phi_1, \mathbf{x}_1 \rangle + e_1 \\ \vdots \\ \langle \phi_W, \mathbf{x}_W \rangle + e_W \end{bmatrix}. \tag{2}$$

If we further assume that the scene is static (as is the case when we are sensing an image), then $\mathbf{x}_1 = \mathbf{x}_2 = \cdots = \mathbf{x}_W = \mathbf{x}$ and we obtain the imaging model in (1).

The theoretical results of CI rely heavily on the properties of random matrices, i.e., matrices whose entries are sampled from certain distributions. A central result states that, when a matrix $\Phi \in \mathbb{R}^{M \times N}$ satisfies the so-called restricted isometry property (RIP) on all K-sparse signals, then it is possible to stably recover all K-sparse signals from linear measurements as in (1). Specifically, the measurement operator Φ is said to satisfy the RIP with constant $\delta > 0$ if, for every K-sparse signals $\mathbf{x}_1, \mathbf{x}_2$, the following relations hold:

$$(1 - \delta)\|\mathbf{x}_1 - \mathbf{x}_2\|^2 \le \|\Phi\mathbf{x}_1 - \Phi\mathbf{x}_2\|^2 \le (1 + \delta)\|\mathbf{x}_1 - \mathbf{x}_2\|^2. \tag{3}$$

The quantity δ encapsulates the deviation from perfect isometry and is called the *isometry constant*. Random matrices provide a simple and elegant method to construct measurement operators that satisfy the RIP. When the entries of Φ are sampled independent and identically distributed (i.i.d.) from sub-Gaussian distributions, then Φ satisfies the RIP with overwhelming probability provided $M = O(K \log(N/K))$ [4]. Similarly, measurement operators Φ obtained via randomly subsampling the rows of certain orthonormal matrices satisfy the RIP with overwhelming probability [11]. Both designs for enabling measurement matrices with the RIP are universal, i.e., they are independent of the application.

Many of the early CI systems used random constructions for measurement matrices. However, despite their conceptual simplicity, random projections suffer from certain shortcomings that make them impractical. Their theoretical guarantees are probabilistic, i.e., there is a nonzero chance that the obtained embedding does not satisfy a (near) isometry, and asymptotic, i.e., the guarantees hold only when the problem dimensions are sufficiently high. Further, by virtue of universality, random matrices are independent of both the data under consideration as well as the eventual inference task that we seek to perform. As a consequence, the use of random projections precludes us from leveraging special geometric structure that might be present in the data or the inference task. From a practical standpoint, large random matrices are also extremely cumbersome, requiring storage and processing requirements that become prohibitive when sensing high-dimensional signals.

In this article, we survey recent trends in the construction of deterministic matrices for CI and highlight key areas where the use of specifically designed measurement matrices provide significant improvements over random constructions. We discuss the following three applications.

- **Structured CI using signature-preserving matrices:** Careful design of measurement operators can enable us to sense certain structures effectively. We present a new interpretation/construction of Hadamard codes using signature-blocks that arranges patterns into groups that share a certain local signature or sequency. The codes can aid in analyzing the scene without having to computationally reconstruct an image, e.g., assessing the signal-to-noise ratio (SNR) of different blocks, and determining which signatures (i.e., features) are most prominent. This new approach to constructing acquisition patterns benefits in both faster recovery and enhanced image quality as well as in object recognition and tracking tasks. Further, the block structure also permits low-resolution previews of different signature-filtered versions of the observed scene.
- **Motion predictive video CS using dual-scale sensing (DSS) matrices:** We show that measurement matrices can be endowed to sense the scene at multiple spatial scales, simultaneously. This enables real-time recovery of the video, albeit at a lower spatial resolution, and can

Single Pixel Camera Basics

While CI has also been employed with large focal plane imaging arrays, the most extreme realization of CI is the SPC. This combines an optical modulator displaying a time sequence of patterns where the image forms on an intermediate image plane. A series of patterns displayed on this modulator convolves with the image and directs a portion of this light towards a single optical sensor while discarding the rest. This provides a set of image measurements that are recorded for later reconstruction into the original image as shown in Figure S1(a). While many choices of optical modulators exist, the digital micromirror array devices from Texas Instruments provide a large number of broad-spectrum modulators that toggle between patterns up to a rate of 30 kHz.

Such an optical system can encode the image in many possible ways, as illustrated in Figure S1(b). One of the most intuitive ways is when one simply scans a single ON mirror one at a time in a raster-encoding manner. The first drawback in this case is that each measurement collects very little light. Second, anything lower than 100% of the measurements results in only a partial image of the scene. Another approach is to use a transform-based acquisition and thus results in roughly half the total scene light being acquired with each measurement. Given the binary modulation nature of the digital micromirror device (DMD), the Hadamard transform appears ideal. However, with less than 100% of the measurements, this case still reconstructs only part of the information contained in the image. Finally, if one employs random measurements as suggested by [11], then only a small fraction of the total measurements are needed to reconstruct nearly the full image. Many of the initial CI systems employed permuted-Hadamard measurements as a means of generating pseudo-random acquisition sequences. However, as this article discusses, prior information and/or assumptions regarding the scene to be imaged can result in large improvements in acquisition and processing for various tasks.
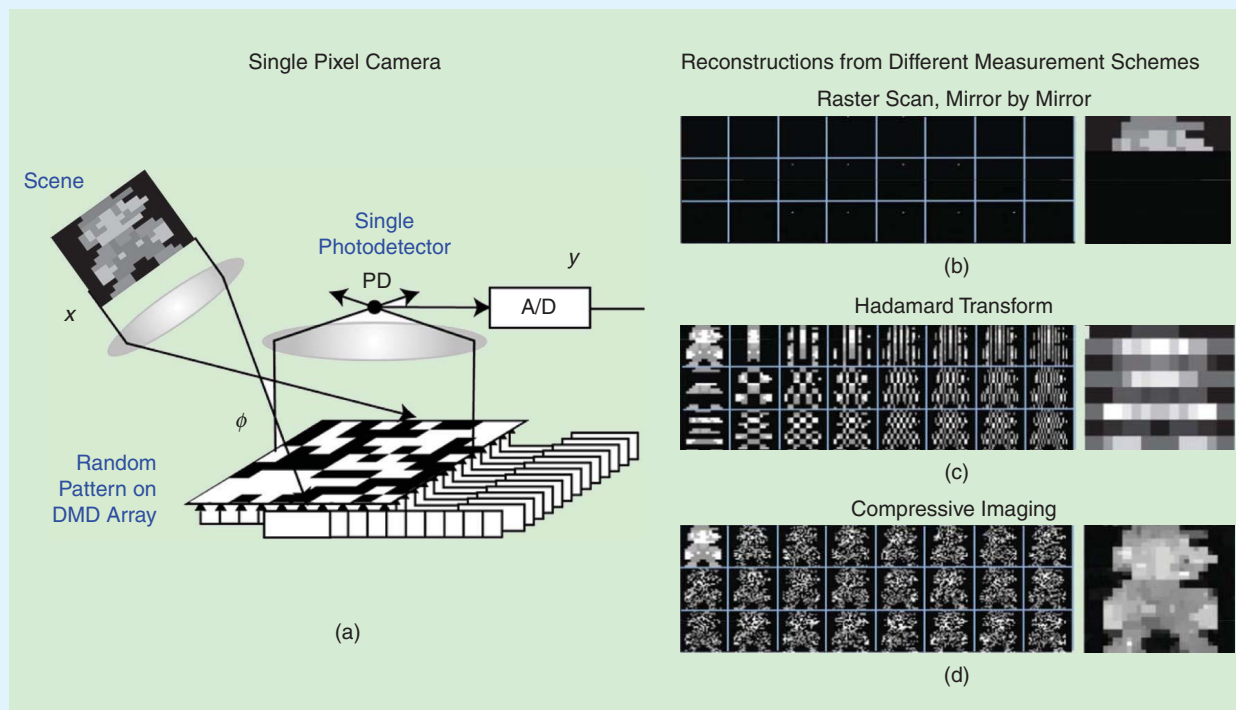
FIGURE S1. (a) A schematic representation of the SPC where the image is focused on an optical modulator and encoded by a series of patterns before being relayed and focused onto a single detector element. (b)–(d) Three examples of inner products of possible encoding patterns and their convolution with the scene being imaged: (b) a raster scan, (c) Hadamard basis, and (d) random basis. Each example also shows the reconstruction obtained from using only 25% of the total measurements.

provide a digital viewfinder for compressive cameras. This low-resolution preview can be exploited to build signal-specific models. An example of this is to estimate scene motion by computing optical flow between frames of the preview. Motion flow provides extremely precise relationships between the frames of a video that can subsequently be exploited to recover the video at high spatial and temporal resolutions.

- **Reconstruction-free inference:** The eventual goal in many applications is often not just reconstruction of an image but an inference problem pertaining to detection, tracking, recognition, and/or classification. While inference can be performed postreconstruction on the output of a reconstruction procedure, there are important benefits to be gained by performing them directly on the compressive domain. First, many inference tasks are inherently simpler than reconstruction; hence, there is hope that we can perform them with fewer measurements. Second, CS reconstruction is intrinsically tied to the signal models used for the unknown signal and these signal models prioritize features that deal with visual perception, which often is not the most relevant for the subsequent processing tasks. Third, reconstruction algorithms associated with CS have high computational complexity; hence, avoiding a reconstruction step in the overall processing pipeline can be beneficial. To highlight these beneficial aspects of compressive inference and the critical role that measurement operator design plays in it, we will review techniques that let us solve high-level computer vision problems (e.g., object, face, and activity recognition) by foregoing reconstruction in favor of inference.

> **The Kronecker product is used throughout mathematical sciences in countless applications such as signal/image processing, control theory, quantum computing, etc.**

In many ways, the examples that we discuss fall under the broad category of model-based CS [3], where signal models beyond simple sparsity are used to obtain recovery guarantees with fewer measurements. A key distinction is that the results of model-based CS rely on random matrix constructions, while we seek alternative methods that are domain- and task-specific.

Structured compressive imaging

In this section, we present a method of generating measurement matrices that are endowed with unique local signatures. These waveforms can be used to make measurements of a scene of interest with applications in imaging and detection/classification. We introduce a generalized Kronecker product that generates a matrix with blocks of rows where, within each block, the rows all share the same local signature (i.e., a specific spatial pattern). The individual rows can be used as patterns, e.g., on a spatial light modulator (SLM) in an optical system that observes a scene.

The Kronecker product

The Kronecker product is used throughout mathematical sciences in countless applications such as signal/image processing, control theory, quantum computing, etc. Part of its utility comes from the ability to tensor together low-dimensional ideas into larger systems [33]. The Kronecker product has long been used whenever operators are separable. A classical example is when an image is represented as a matrix and the transformed image can be separated into two functions, one that operates on the rows and another that operates on the columns. Extending this to higher dimensions is straightforward. Recently, Kronecker product-based CS has been explored in the context of multidimensional signals [13]. Arguably though, the Kronecker product's most crucial role is in enabling fast implementations of important mathematical functions, such as the discrete Fourier transform, the Haar wavelet, and the Hadamard transform. We are particularly interested in the Kronecker product-based constructions of Hadamard matrices, which have had a long history in imaging and optics [16].

A Hadamard matrix is an orthogonal matrix with entries restricted to just $\{\pm 1\}$. Power-of-two Hadamard matrices are often thought of as a square-wave version of the discrete cosine transform (DCT) and are attractive since they have an associated fast transform. However, Hadamard matrices of sizes other than powers of two exist as well, and they can also have fast transforms. The rows of Hadamard matrices are usually described in terms of their sequency, which is similar to the notion of the frequency of a sinusoid. Sequency is simply the number of ± 1 transitions contained in a Hadamard waveform.

The local signature-based measurement matrix design presented here is a simple generalization of the standard Kronecker product. The typical Kronecker product of matrices A and B is defined [33] as $A \otimes B := [a_{ij} B]$, where a_{ij} is the (i, j)th element of matrix A. From the definition, we see that commuting the factors A and B in general yields $A \otimes B \neq B \otimes A$. Note that the elements of the left-hand factor provide the weights for the copies of the right-hand factor. Qualitatively, we can think of the left-hand factor A as the modulator and the right-hand factor B as that which is modulated. In this sense, the rows of B provide the local patterns that ultimately generate the global patterns in the rows of $A \otimes B$. These shorter, local patterns can be thought of as signatures.

The signature row-block Kronecker product

Suppose matrices A and B, respectively, have K and L rows, i.e., $A = [a_i]_{i=0}^{K-1}$, $B = [b_i]_{i=0}^{L-1}$. The signature row-block (SRB) Kronecker product is defined as

$$A \otimes_{SRB} \{B\} := \begin{bmatrix} \mathfrak{B}_0 \\ \vdots \\ \mathfrak{B}_{L-1} \end{bmatrix}, \quad \mathfrak{B}_j := A \otimes b_j = \begin{bmatrix} a_0 \otimes b_j \\ \vdots \\ a_{K-1} \otimes b_j \end{bmatrix}. \quad (4)$$

Here, $\mathfrak{B}_j$ is the jth SRB, which consists of the K rows of A that modulate just signature b_j. The signature rows of matrix B analyze/synthesize the local patches of pixels in the image, and the rows of matrix A simply multiplex these into the larger, global measurement patterns. Hence, it is matrix B that is directly tied to the model-based acquisition strategy. For example, in CS applications, matrix B could be a dictionary previously trained by a principal component analysis to have maximal incoherence with respect to the sparsity basis of an
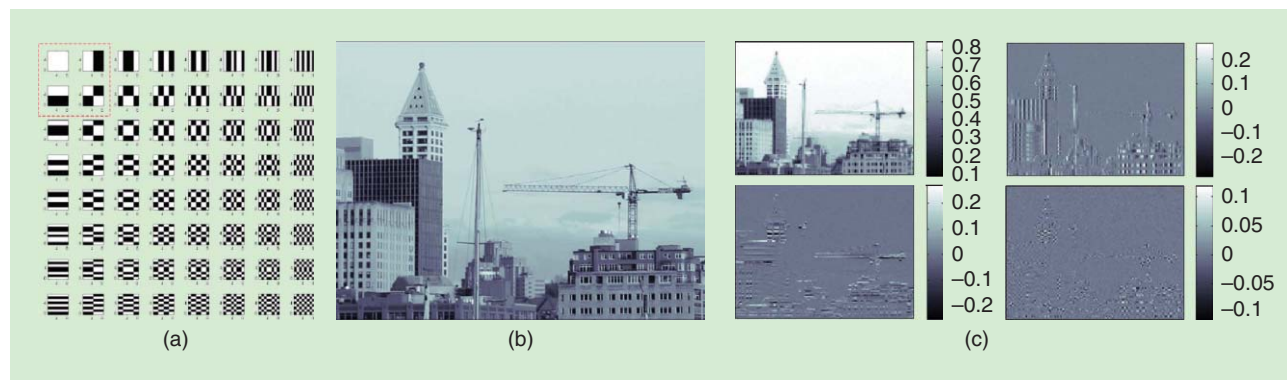
FIGURE 1. (a) 8 × 8 signature tiles (2-D sequence patterns) corresponding to the 64 SRBs that partition the Hadamard domain. (b) A 768 × 1024 ground truth scene. (c) Four low-resolution 96 × 128 signature previews from the four lowest-sequency blocks [marked in red in (a)], which all have 100% complete sampling. (Figure courtesy of www.dpreview.com.)

observed scene, or it could be a standard unitary matrix such as the DCT.

The SRB Kronecker product (4) can easily be obtained from either of the typical Kronecker products $A \otimes B$ or $B \otimes A$. That is, there exist permutation matrices P and Q such that $A \otimes_{\text{SRB}} \{B\} = P(A \otimes B) = (B \otimes A)Q$. This is significant because, if observations of a scene with the rows of an SRB Kronecker product matrix are collected, the typical Kronecker product and its inverse can be used (and their fast implementations, if they exist) for global processing of the whole image. Yet, at the same time, the SRB structure lets us group measurements according to the local signatures, which has value since it is the local information that contains details such as edges, textures, or anomalies within a signal/scene of interest. This enables properties such as the ability to view the data either within the context of particular SRBs or within the context of the larger transform space, analyze or solve an imaging/inference problem as L separate smaller problems or as one large problem, assess the SNR of a particular SRB's coefficients and/ or determine optimal bit allocation, quickly generate downsampled previews of the scene filtered through each signature, and subsample within certain SRBs, e.g., in CI applications. The last two items are examined in the next section.

Using signature row-block Kronecker products for CI

It is easy to extend the SRB Kronecker product (4) to 2-D imaging applications. In this case, the SRB structure naturally endows the measurement space with a convenient 2-D partitioning now based on local signature tiles, instead of the one-dimensional signature rows discussed previously. For computational imaging, the rows of a Hadamard matrix can be reshaped into the 2-D spatial modulating waveforms used, e.g., on the digital micromirror device (DMD) used in the SPC. In this application, each element of a given row is mapped to one mirror of the DMD, and the ±1 values determines whether it is in an ON or OFF state. Hadamard matrices have been used extensively as sensing matrices in CI since they have been shown to be incoherent to sparse signals. Further, the fast implementation of many Hadamard transforms

means that the reconstruction algorithms can quickly converge to a solution. Many CI applications also apply a scrambling operation to the Hadamard matrix, e.g., randomly permuting the columns. This breaks up the structure and results in pseudorandom binary patterns that can be beneficial in certain situations. However, this is different than the approach taken here.

At the same time, Hadamard patterns can be used in a more traditional transform coding/decoding manner. Power-of-two Hadamard transforms have good energy compaction properties, similar to the DCT. We can utilize this fact in conjunction with the partitioned block structure provided by the SRB Kronecker product. Now, with A as a Hadamard matrix, each individual SRB $\mathfrak{B}_j$ in (4) is an orthogonal basis for a subspace encoded or filtered by the signature b_j. Further, with B also as a Hadamard matrix, the set of SRBs $\{\mathfrak{B}_j\}$ are orthogonal to each other. With B specifically as a power-of-four Hadamard matrix H_{4^n}, for some n, the signatures will span all possible sequences when observing $2^n \times 2^n$ patches of pixels, which is similar to the range of spatial frequencies in the 2-D DCT. For example, if matrix B is a Walsh-Hadamard matrix H_{64} in (4), the measurement space is divided into 64 SRBs associated with the 2-D signature 8×8 tiles shown in Figure 1. Each of these signature tiles correspond to one row of H_{64} that has been reshaped to 2-D.

The SRB structure of the deterministic sensing matrix $A \otimes_{\text{SRB}} \{B\}$ lends itself to selective and model-based sampling strategies. We are free to choose which SRBs we want to sample from, and we can choose to sample them partially or completely. This leads to a partial-complete sensing approach [20] that is essentially a block-structured version of variable density sampling. Note that SRBs that are completely sampled at 100% can be easily and quickly demodulated by removing the multiplexing effect of matrix A (more details can be found in [19]). This provides a low-resolution preview of the scene filtered through the signature tile associated with a particular SRB. To see this, consider the scene with $N = 768 \times 1024 = 12 \cdot 2^{16}$ pixels shown Figure 1. If we want to construct an SRB Kronecker matrix with $B = H_{64}$

in (4), then we must choose $A = H_{12 \cdot 2^{10}}$ so that $A \otimes_{SRB} \{B\}$ is $N \times N$. Suppose that we completely sample the four lowest sequency blocks associated with the signature tiles outlined in the red-dashed line in Figure 1. The low-resolution previews of the scene filtered through these signature tiles can be seen in Figure 1. These signature previews reveal that the buildings in the observed city skyline have strong vertical and horizontal components. In general, any a priori information will dictate which SRBs to focus on.

Returning to the goal of CI, we selected modes from each of the 64 SRBs of the 2-D Hadamard domain such that the total number of measurements M was 15% of N, as shown in Figure 2(a). Besides the four completely sampled SRBs, the other 60 SRBs were partially sampled at random with a canonical variable density described in [21]. Most natural scenes have their energy focused in the lower sequency modes, so we sampled more densely in these SRBs. Note that the measurements had additive white Gaussian noise with an SNR of 30 dB. We used these measurements in conjunction with a reconstruction algorithm that minimized the total variation (TV) that resulted in the recovered image seen in Figure 2(b). Compared to the ground truth in Figure 1, good detail can be seen even though only 15% of the possible Hadamard modes were used to observe

the scene. Furthermore, the algorithm converged in just 13 iterations. This approach has some similarity to hybrid sampling methods that gather low-frequency measurements, followed by higher-frequency detail measurements. A somewhat related strategy is to assemble a union of bases consisting of chirp or Reed–Muller sequences as the sensing matrix [27].

Next, we examine how our structured SRB sensing matrix compares with a typical CI sensing matrix. We used the same 2-D Hadamard domain, but the sensing modes were chosen uniformly at random, again such that $M/N = 15\%$ as seen in Figure 2(d). Note that sensing in this manner usually results in extremely poor reconstructions. We ameliorated this by scrambling each of the Hadamard patterns before sensing so that they appeared as binary noise. (In this sense, the comparison with the SRB method is not exact but is the most fair.) As before, white Gaussian noise was added to the measurements such that the SNR was 30 dB. However, now the TV-minimization algorithm required 88 iterations to converge to the solution in Figure 2(e). Compared to Figure 2(b), we see that the SRB technique produces a slightly sharper image and does so in more than six-fold fewer iterations, which is a significant improvement. In addition, the SRB technique, by providing low-resolution signature previews, demonstrates how intelligent sampling of the transform domain can offer simultaneous supplementary information.

Although the Hadamard coefficients with the highest energy tend to be concentrated in the lowest-sequency blocks, this is not always the case. Further, a canonical variable density strategy is not always ideal. It is possible to use the SRB structure to find out which are the best blocks to sense as well as what the ideal subsampling density is for each block [21]. By simply subsampling just a few Hadamard modes (e.g., much less than 1%) from each SRB, we can assemble a sufficient statistic that can guide us to the blocks with the most energy, which should have the best SNR. Hence, we can adapt to an observed scene by doing a fast initial query of the transform measurement domain and thereby get the most bang for the buck.

Overall, we see that the SRB Kronecker product provides a structure that enables flexible sensing strategies. By properly designing the matrix and choosing which SRBs to use, important local information can be gleaned from the global measurements of an observed scene. The previous example shows how it can be used for imaging in a CS manner as well as in providing low-resolution signature previews. However, the SRB
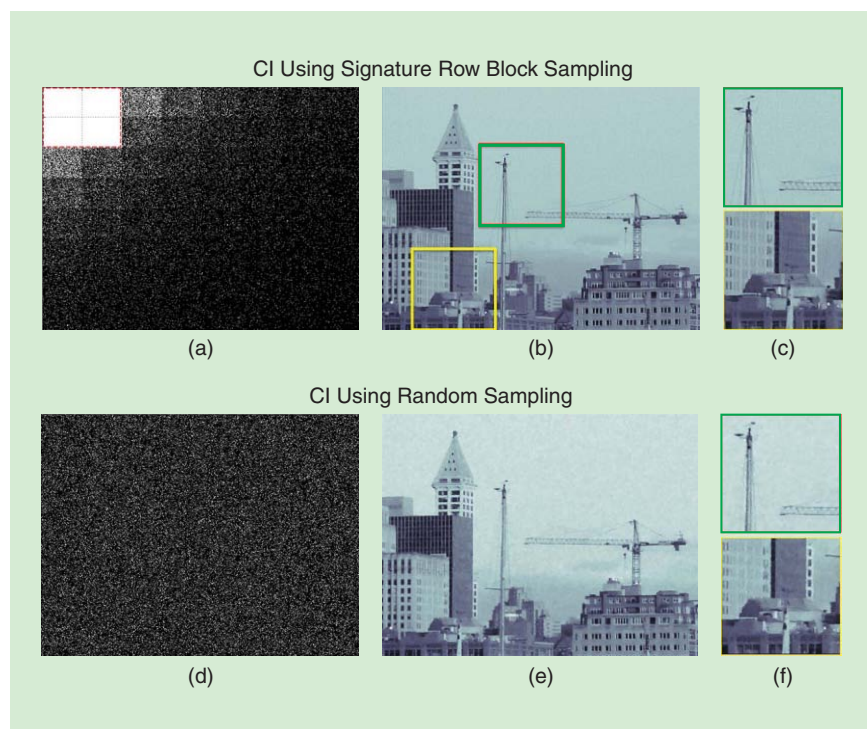
FIGURE 2. (a) The 2-D Hadamard domain partitioned by 64 SRBs (the 8 × 8 sequency patterns unique to each block are shown in Figure 1). The white dots indicate which Hadamard modes were used to observe the ground truth in Figure 1. The number of white dots equals M, the number of samples, such that $M/N = 15\%$. (b) The resulting reconstructed image using TV minimization with insets shown in (c). Note how, in spite of the high compression, certain signatures like horizontal and vertical stripes are well preserved. (d) The 2-D Hadamard domain uniformly sampled at random as is done in typical CS. (e) The resulting reconstructed image using the same TV minimization with insets shown in (f). The quality is slightly degraded and took more than six times as long to converge. (Figure courtesy of www.dpreview.com.)

structure can also be used for detection and classification applications. Hadamard matrices were chosen in our examples because they possess several attractive properties, i.e., they are well suited for the binary nature of the DMD, they have fast transforms, they have good energy compaction for most natural images, and they are incoherent with respect to popular sparsity bases used in CI. Hadamard matrices used in conjunction with the SRB Kronecker product structure can provide powerful model-based measurement matrices.

Spatial-temporal resolution tradeoffs and motion predictive video compressive sensing

In this section, a case study for video CS is presented using the SPC with a specially designed measurement matrix that provides the ability to tradeoff spatial and temporal resolutions. Recall that the SPC uses a single photodetector to obtain compressive measurements of a scene. However, this use of a single photodetector is, in general, not sufficient for acquiring complex scenes at high resolution. Hence, it is common to assume that the scenes to be acquired are static and acquire multiple measurements over time. This approach, however, fails for time-variant scenes since each measurement acquires information of a (slightly) different scene. Figure 3 illustrates the effect of the violation of the static scene assumption. Put simply, grouping too few measurements for reconstruction results in poor spatial resolution, and grouping too many measurements results in severe temporal aliasing artifacts.

Characterizing spatial-temporal tradeoffs

We first study the effect of downsampling on the measurement process and make two assumptions. First, the time-varying scene can be approximated as static, i.e., $\mathbf{x}_t = \mathbf{b} + \Delta\mathbf{x}_t$ where $\mathbf{b}$ is the static component, and $\Delta x_t = x_t - b$ is the error at

sample instant t caused by the static-scene assumption. Second, we decompose the static image $\mathbf{b}$ into a low-resolution component, $\mathbf{Ub}_L$, where $\mathbf{b}_L = \mathbf{Db} \in \mathbb{R}^{N_L}$, and the high-frequency residual $\mathbf{b} - \mathbf{Ub}_L$, where $\mathbf{D}$ and $\mathbf{U}$ are a pair of downsampling and upsamplign operators. Now the measurement equations in (2) can be written as

$$\mathbf{y}_{1:w} = \Phi(\mathbf{Ub}_L + \mathbf{b} - \mathbf{Ub}_L) + \mathbf{z}_{1:w} + \mathbf{e}_{1:w}$$
$$= \Phi\mathbf{Ub}_L + \Phi(\mathbf{I} - \mathbf{UD})\mathbf{b} + \mathbf{z}_{1:w} + \mathbf{e}_{1:w}. \quad (5)$$

The term $\mathbf{z}_{1:w}$ accounts for the motion blur terms in $\Delta\mathbf{x}_t$. Inspection of (5) reveals three sources of error in the CS measurements of the low-resolution static scene $\Phi\mathbf{Ub}_L$: 1) the spatial-approximation error $\Phi(\mathbf{I} - \mathbf{UD})\mathbf{b}$ caused by downsampling, 2) the temporal-approximation error $\mathbf{z}_{1:w}$ caused by assuming the scene remains static for W samples, and 3) the measurement error $\mathbf{e}_{1:w}$. Note that when $W \geq N_L$, the matrix ΦU has at least as many rows as columns, and, hence, we can get an estimate of $\mathbf{b}_L = (\Phi U)^\dagger \mathbf{y}_{1:w}$.

$$\widehat{\mathbf{b}}_L = (\Phi\mathbf{U})^\dagger \mathbf{y}_{1:w} = \mathbf{b}_L + (\Phi\mathbf{U})^\dagger(\Phi(\mathbf{I} - \mathbf{UD})\mathbf{b} + \mathbf{e}_{1:w} + \mathbf{z}_{1:w}), \quad (6)$$

where $(\cdot)^\dagger$ denotes the pseudoinverse. The window length W controls a tradeoff between the spatial-approximation error $\Phi(\mathbf{I} - \mathbf{UD})\mathbf{b}$ and the error $\mathbf{z}_{1:w}$ induced by assuming a static scene $\mathbf{b}$, and the least squares estimator matrix $(\Phi\mathbf{U})^\dagger$ (potentially) amplifies all three error sources.

Dual-scale sensing matrices

The choice of the sensing matrix Φ and the upsampling operator $\mathbf{U}$ are critical to arrive at a high-quality estimate of the low-resolution image $\mathbf{b}_L$. Indeed, if the effective matrix $\Phi\mathbf{U}$ is ill-conditioned, then application of the pseudoinverse
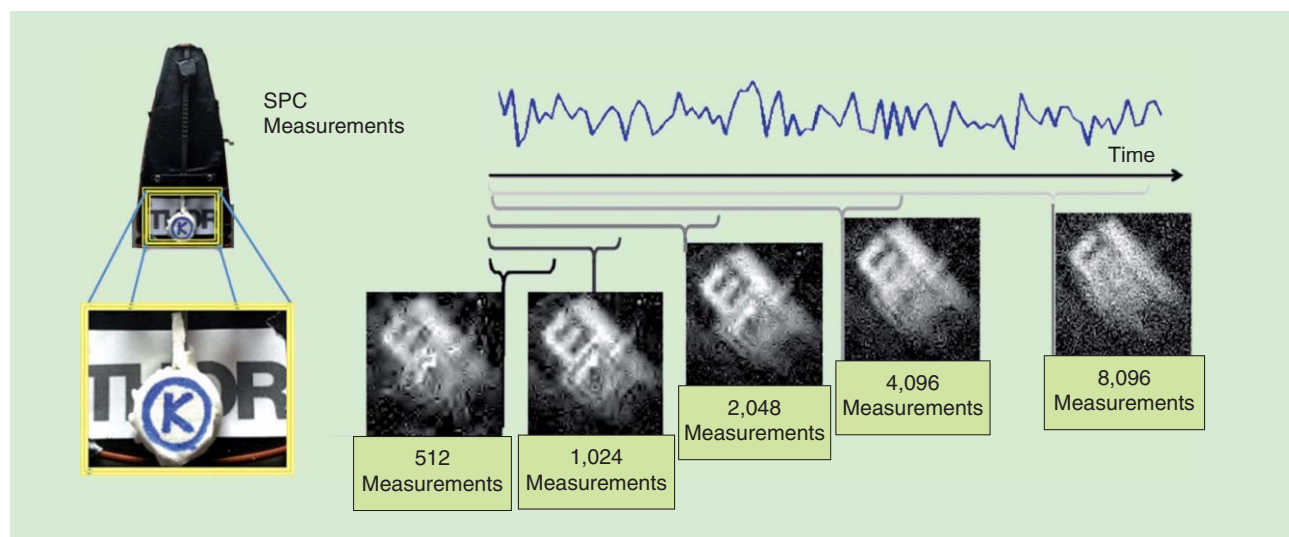
FIGURE 3. SPC and the static scene assumption. An SPC acquires a single measurement per time instant. If the scene were static, one can aggregate multiple measurements over time to recover the image of the scene via sparse signal recovery; for dynamic scenes, however, this approach fails. Shown above are reconstructs of a scene comprising of a pendulum with the letter R swinging from right to left. We show reconstructed images using different numbers of aggregated (or grouped) measurements. Aggregating only a small number of measurements results in poor image quality. Aggregating a large number of measurements violates the static scene assumption and results in dramatic temporal aliasing artifacts. (Figure adapted from [31].)

$(\Phi U)^{\dagger}$ amplifies all three sources of errors in (6), eventually resulting in a poor estimate. For virtually all sensing matrices Φ commonly used in CS, such as i.i.d. (sub)Gaussian matrices, right multiplying them with an upsampling operator U often results in an ill-conditioned matrix or even a rank-deficient matrix. This is a consequence of the smallest singular value of square random matrices approaching zero when the size of the matrix is large [34].

A novel class of sensing matrices, referred to as DSS matrices, achieve good CS recovery performance and have minimum noise enhancement when computing a low-resolution preview $\widehat{\mathbf{b}}_L$ according to (6). These matrices will satisfy the RIP to enable CS and remain well-conditioned when right multiplied by a given up-sampling operator U. One approach to the design of DSS matrices Φ (see Figure 4) relies on upsampling low-resolution Hadamard matrices. A simple way is to start with a $W \times W$ Hadamard matrix $\mathbf{H}$ and to write the CS matrix as $\Phi = \mathbf{HD} + \mathbf{F}$, where $\mathbf{F} \in \mathbb{R}^{W \times N}$ is a ± 1 matrix such that $\mathbf{FU} = 0$, i.e., each block of $\mathbf{F}$ should sum to zero. A powerful example of such a construction is the so-called sum-to-one transform [15], where a Hadamard matrix is carefully designed to satisfy the dual scale property across any pairs of scales.

> **Hadamard matrices have fast inverse transforms that can significantly speed up the recovery of the low-resolution preview frames.**

Preview mode

The use of Hadamard matrices for the low-resolution part in the proposed DSS matrices has an additional benefit. Hadamard matrices have fast inverse transforms that can significantly speed up the recovery of the low-resolution preview frames. Such a fast DSS matrix has the key capability of generating a high-quality preview of the scene (see Figure 4) with very low computational complexity. This is beneficial for video CS as it allows one to easily and quickly extract an estimate of the scene motion. The motion estimate can then be used to recover the video at its full resolution. In addition, the use of fast DSS matrices can be beneficial in various other ways, including as a digital viewfinder.

Motion flow-based video CS

State-of-the-art video compression methods rely on estimating the motion in the scene, compress a few reference frames, and use the motion vectors that relate the remaining parts of a scene to these reference frames. While this approach is possible in the context of video compression, i.e., where the algorithm has prior access to the entire video, it is significantly more difficult in the context of CS.

A general strategy to enable the use of motion flow-based video CS is to use a two-step approach [28]. In the first step, an initial estimate of the video is generated by recovering each frame individually using sparse wavelet or gradient priors. The initial estimate is used to derive motion flow between consecutive frames, which enables a powerful description in terms of relating similar intensities at pixels across frames. In the second step, the video is re-estimated with the aid of enforcing the extracted motion-flow constraints in addition to the measurement constraints. The success of this two-step strategy critically depends on the ability to obtain reliable motion estimates, which, in turn, depends on obtaining robust initial estimates. Further, since we are upsampling the motion estimates, this approach only works for scenes where the moving objects occupy a few pixels in the low-resolution preview. We can obtain such robust initial estimates by using the DSS matrices. Figure 5 illustrates the outline of a video CS algorithm [31] that uses the DSS matrices to enable robust previews, compute motion flow using the previews, and exploit the motion flow to sense videos at high compressions.

Process of Generating Rows of DSS Matrices — Row of the Hadamard Matrix → Upsampling (NN) → ⊕ High-Frequency Sparse Pattern → Row of the DSS Matrix

Previews of Three Scenes — Windmill, Pendulum, Hand
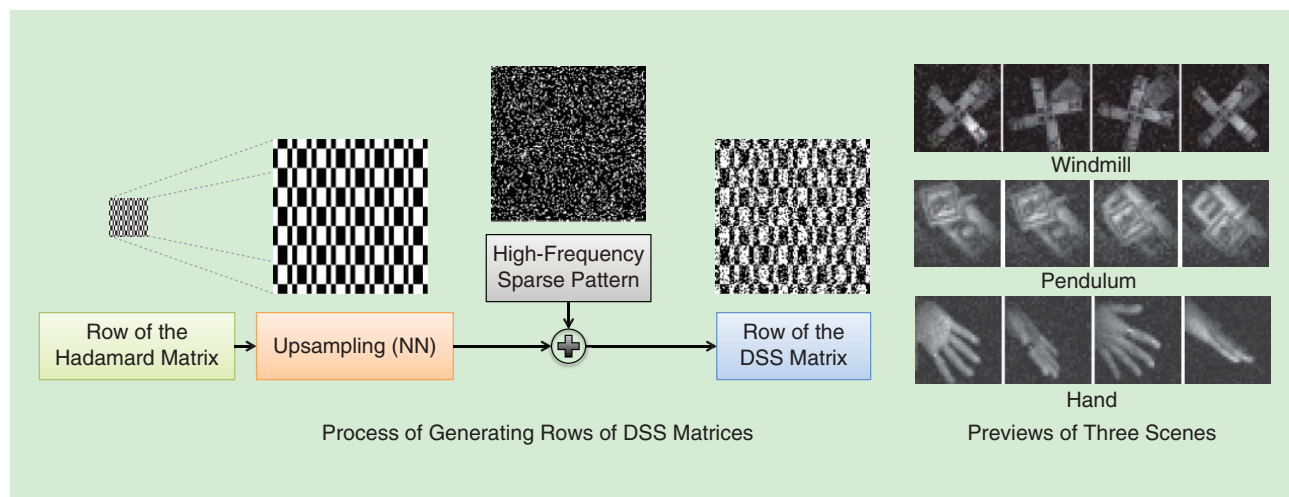
FIGURE 4. Generating DSS patterns and previews. DSS matrices are generated such that downsampling each row produces a row of a low-resolution Hadamard matrix. Also shown are previews generated for three different scenes with real data. (Figure adapted from [31].)
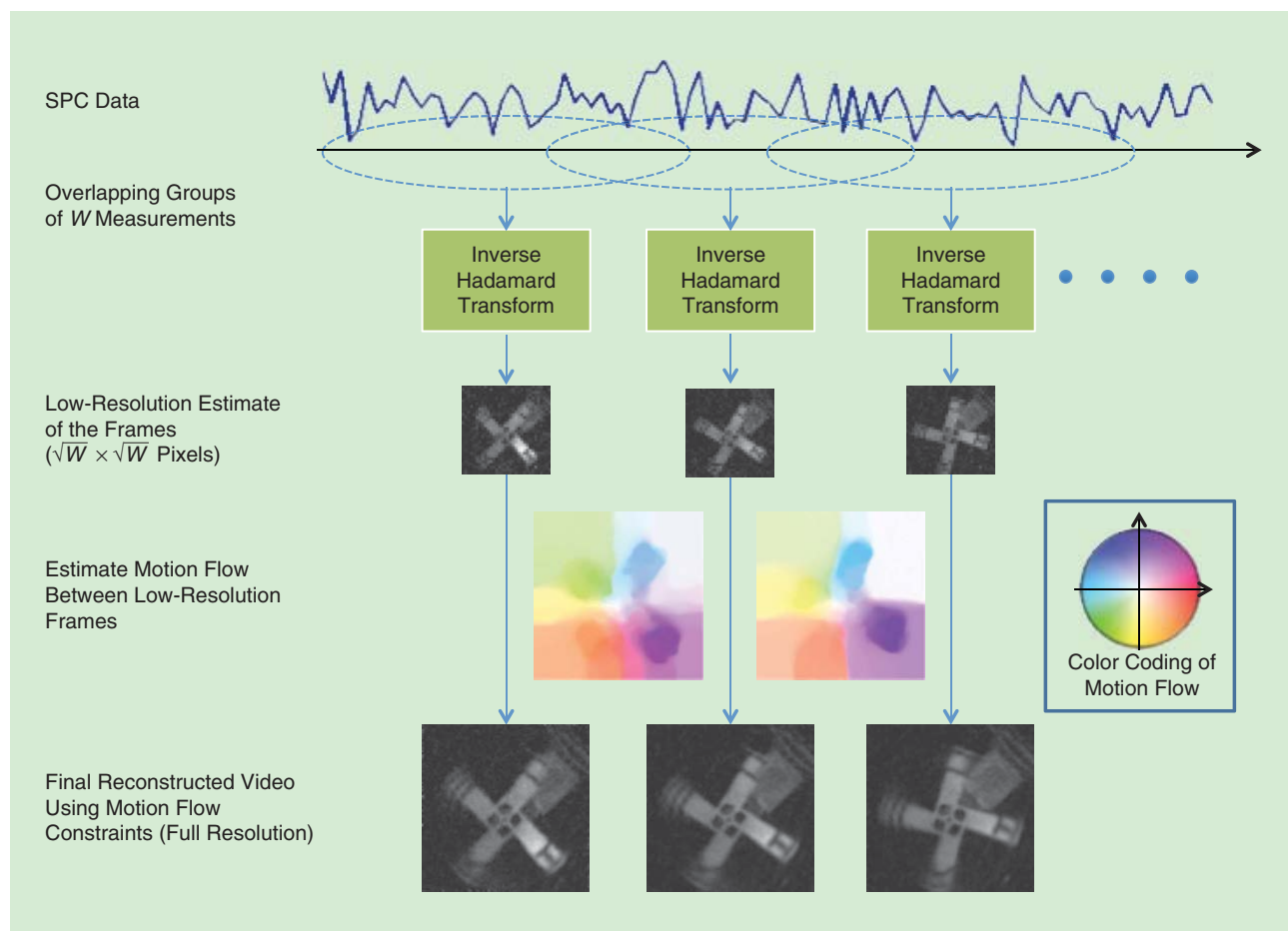
FIGURE 5. An outline of the CS multiscale video (CS-MUVI) recovery framework. Given a total number of T measurements, we group them into overlapping windows of size W, resulting in a total of F frames. For each frame, we first compute a low-resolution initial estimate using a window of W neighboring measurements. We then compute the optical flow between upsampled preview frames (the optical flow is color coded as in [24]). Finally, we recover F high-resolution video frames by enforcing a sparse gradient prior along with the measurement constraints as well as the brightness constancy constraints generated from the optical-flow estimates. (Figure adapted from [31].)

Reconstruction-free inference

In this section, we discuss high-level inference problems that can be solved directly in the compressive domain without the need for reconstruction of the imagery. The primary example we consider is that of action recognition, which is one of the long standing research areas in computer vision with widespread applications in video surveillance and human-computer interaction. In many applications of action recognition, one is faced with resource constraints such as limited power supply, limited storage, and limited computational capabilities, all of which make CI a great solution. To this end, we focus specifically on action recognition from videos and review the results that demonstrate that it is indeed possible to perform action recognition at extremely higher compression ratios without reconstructing the video in the first place.

The central idea that allows for detection/recognition problems to be solved directly in the compressed domain is the Johnson–Lindenstrauss (JL) lemma and its application to compressive convolutional filtering (also referred to as *smashed filtering*). Specifically, the JL lemma states that the correlation between any two signals is nearly preserved even when the data is compressed to a much lower-dimensional space. This property allows one to evaluate the response of a signal to a given convolutional kernel directly in the compressed domain. This is extremely useful in practice since correlational features have traditionally been used extensively in computer-vision problems such as automatic target recognition, face recognition [35], palm-print identification [18], and even activity recognition [29]. Davenport et al. [25] introduced the concept of smashed filters that provides a way to extract such correlational features in the compressed domain.

We first show that approximate correlational features can be extracted directly from CS measurements even in very challenging activity videos. Using this in conjunction with the widely used correlational filters approach to recognition tasks in computer vision, one can develop a spatiotemporal smashed filtering approach to action recognition that results in robust performance at extremely high compression ratios.

The overall algorithmic pipeline for directly extracting correlational features from compressed video is illustrated in

Figure 6. We start with pretrained space-time filters for each specific action. We chose to train maximum average correlation height (MACH) filters for each action class [29]. The JL lemma then provides a way to evaluate the filter response for a given video directly in the compressed domain, thereby avoiding reconstructing the frames of the test video. To reduce computational complexity, the three-dimensional (3-D) response volume is calculated in the frequency domain via 3-D fast Fourier transform. For a given test video, we obtain N_A correlation volumes. For each correlation volume, we adapt three level volumetric max pooling to obtain a 73-dimensional feature vector [30]. In addition, we also compute peak-to-sidelobe ratio for each of these 73 max-pooled values. This framework can be used in any reconstruction-free application from compressive cameras that can be implemented using 3-D correlation filtering. The action localization in each frame is determined by a bounding box centred at location $(l^{\max})$ in that frame, where $l^{\max}$ is determined by the peak response (i.e., the response corresponding to the classified action) in that frame and the size of the filter corresponding to the classified action. To determine the size of the bounding box for a particular frame, the response values inside a large rectangle of the size of the filter and centered at $l^{\max}$ in that frame are normalized so that they sum up to unity. Treating this normalized rectangle as a 2-D probability density function, we determine the bounding box to be the largest rectangle centered at $l^{\max}$, whose sum is less than a value λ. For our experiments, we use λ equal to 0.7.

Experimental results

We present sample results obtained on Weizmann [7] and the University of Central Florida (UCF) sports [29] data sets. More extensive results can be found in [22]. For all of our experiments, we use a measurement matrix ϕ, whose entries are drawn from i.i.d. standard Gaussian distribution, to compress the frames of the test videos. We note that it is possible to use more esoteric measurement matrices to improve either reconstruction and/or recognition performance. For example, variants of wavelet bases are better suited for reconstruction and task-driven measurement operators are better suited for inference. In this section, we use the random Gaussian matrix to level the playing field for reconstruction and inference.

Results on Weizmann data set

The Weizmann data set contains ten different actions, each performed by nine subjects, thus making a total of 90 videos. For evaluation, we used the leave-one-out approach, where the filters were trained using actions performed by eight actors and tested on the remaining one. The results shown in Figure 7 indicate that our method clearly outperforms the reconstruct-then-recognize using the improved dense trajectories (IDT) method, a state-of-the-art recognition algorithm. At compression ratios of 100 and above, recognition rates are very stable for the compressive recognition framework, while reconstruct-then-recognize fails completely. The recognition rates are stable even at high compression ratios and are comparable to the recognition accuracy for the Oracle MACH (OM) method [1]. The average time taken by both methods to process a video of $144 \times 180 \times 50$ size are shown in parentheses in Figure 7. Recon+IDT takes about 20–35 minutes to process one video, with the frame-wise reconstruction of the video being the dominating component. In contrast, compressive inference takes only a few seconds. The sample spatial localization results are shown in Figure 7(a) in a few frames for various actions of the data set.

Results on UCF sports data set

The UCF sports action data set [29] contains a total of 150 videos across nine different actions. It is a challenging data
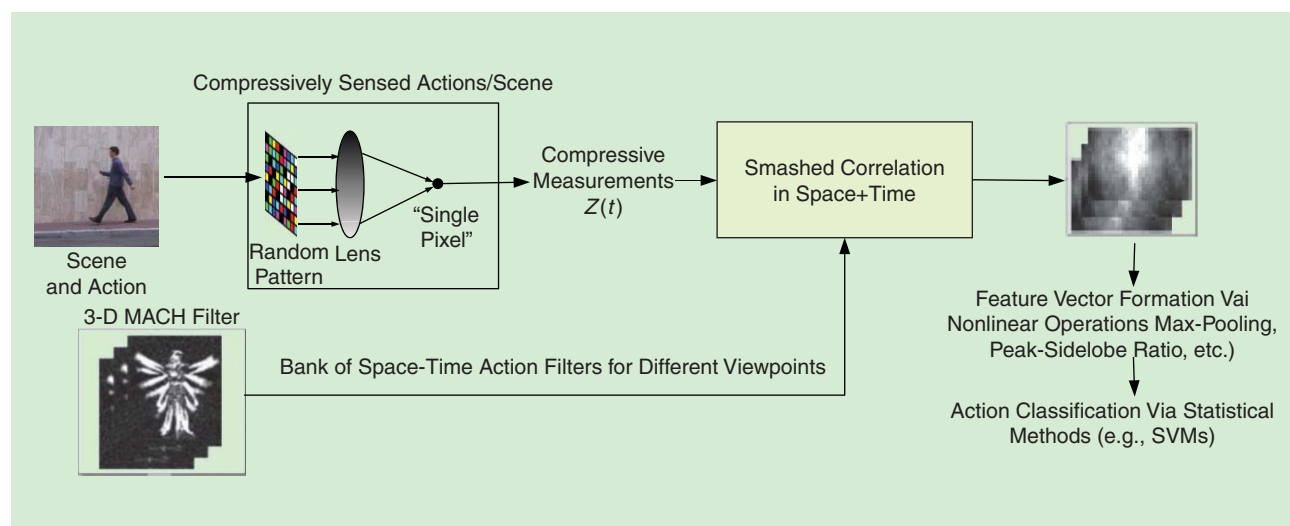
FIGURE 6. Compressive inference via smashed filters. (a) Every frame of the scene is compressively sensed by optically correlating random patterns with the frame to obtain CS measurements. (b) An overview of our approach to action recognition from a compressively sensed test video. First, MACH [29] filters for different actions are synthesized offline from training examples and then compressed to obtain smashed filters. Next, the CS measurements of the test video are correlated with these smashed filters to obtain correlation volumes that are analyzed to determine the action in the test video. (Figure adapted from [22].)

Compression Factor	Compressive Recognition		Reconstruct-Then-Recognize (IDT)	
	Recognition Rate	Run Time	Recognition Rate	Run Time
1	81.11%	3.22 seconds	100%	3.1
100	81.11%	3.22 seconds	5.56%	1,520 seconds
200	81.11%	3.07 seconds	10%	1,700 seconds
300	76.66%	3.1 seconds	10%	1,800 seconds
500	78.99%	3.08 seconds	7.77%	2,000 seconds
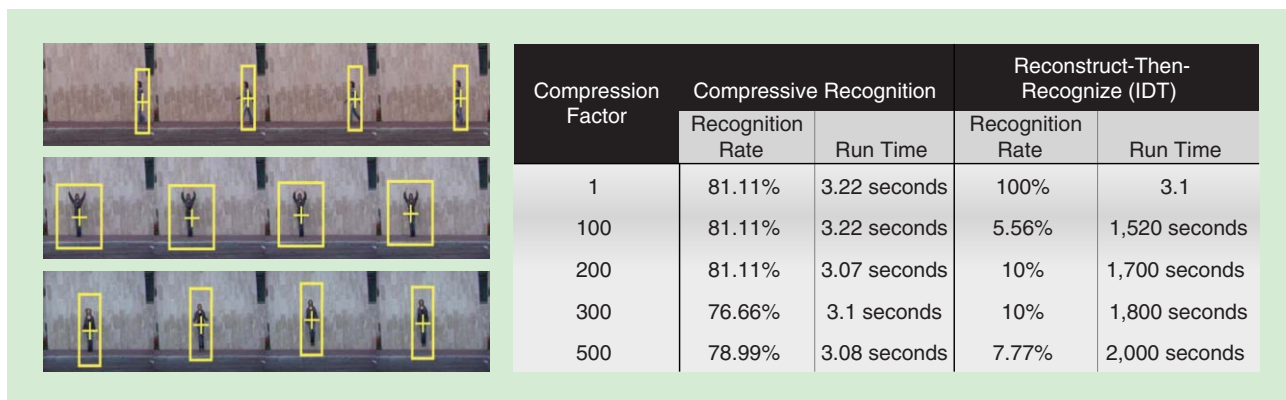
FIGURE 7. (a) The spatial localization of actions without reconstruction at a compression ratio of 100 for different actions in the Weizmann data set. (b) The table compares recognition rates and runtimes for direct recognition in the compressive domain versus reconstruct-then-recognize. Direct recognition outperforms the reconstruct-then-recognize with IDT method and achieves a recognition rate that is comparable to the recognition rate of 81.11% in the case of OM [1], [29] and is much faster. (Figure adapted from [22].)

set with scale and viewpoint variations. For testing, we use leave-one-out cross validation. At the compression ratios of 100 and 300, the recognition rates are 70.67% and 68%, respectively. The rates obtained are comparable to those obtained in the OM setup [29] (i.e., 69.2%). Considering the difficulty of the data set, these results are very encouraging. Sample results of the spatial localization and confusion matrix for recognition are shown in Figure 8.

Learning tuned measurement operators for task-specific inference
While the results presented previously indicate that effective recognition of actions is feasible from compressive imagers without the need for reconstruction, the results suggest that one can potentially achieve better results with a task-specific design of measurement matrices. As already indicated at the beginning of the experiments section, task-specific design can be achieved in a number of ways using

Action	Golf Swing	Kicking	Riding Horse	Run Side	Skateboarding	Swing	Walk	Driving	Lifting
Golf swing	**77.78**	16.67	0	0	0	0	5.56	0	0
Kicking	0	**75**	0	5	5	10	5	0	0
Riding Horse	16.67	16.67	**41.67**	8.33	8.33	0	8.33	0	0
Run Side	0	0	0	**61.54**	7.69	15.38	7.69	7.69	0
Skateboarding	0	8.33	8.33	25	**50**	0	5	0	0
Swing	0	3.03	12.12	0.08	3.03	**78.79**	3.03	0	0
Walk	0	9.09	4.55	4.55	9.09	9.09	**63.63**	0	0
Driving	0	0	0	0	7.14	0	0	**92.86**	0
Lifting	0	0	0	0	0	0	0	16.67	**83.33**

FIGURE 8. (a) The reconstruction-free spatial localization of the subject for OM (shown as a yellow box) and spatiotemporal smashed filter (STSF) (shown as a green box) at a compression ratio of 100 for some correctly classified instances of various actions in the UCF sports data set. (b) The confusion matrix for UCF sports database at a compression factor of 100. The recognition rate for this scenario is 70.67%, which is comparable to OM [29] (69.2%). (Figure adapted from [22].)

Can We Learn Good Measurement Matrices?

The criterion of near isometry is geared toward enabling reconstruction that is not necessarily conducive to inference. As an example, consider classification of two classes using a nearest-neighbor (NN) classifier. The near-isometry property underlying RIP ensures that distances are approximately preserved, and, therefore, NNs are approximately preserved. Yet, the preservation of distance is not necessary for NN classification. Indeed it is sufficient if the measurement operator does not perturb the class membership of the NN of a point. Intuitively, this is a significantly simpler constraint to satisfy, and we can hope to achieve it with far fewer measurements.

Domain-specific design of measurement operators

Can we learn measurement operators that can be tuned to the specifics of an application or data domain and, fur-

ther, incorporate task-specific constraints? Specifically, given a collection of data $\{x_1, x_2, \ldots, x_Q \in \mathbb{R}^N\}$, we pose the problem of *learning* a measurement matrix $\Phi \in \mathbb{R}^{M \times N}$ that satisfies the RIP on this data set, i.e., we seek a matrix $\Phi \in \mathbb{R}^{M \times N}$ that satisfies

$$\forall i, j, \quad (1 - \delta) \leq \frac{(x_i - x_j)^T \Phi^T \Phi (x_i - x_j)}{\| x_i - x_j \|^2} \leq (1 + \delta). \quad (S1)$$

Unfortunately, solving for a matrix Φ with the fewest rows that satisfy (S1) is a nonconvex problem. In particular, while this is a hard optimization problem over Φ, by using a lifting operation, we can pose this as an optimization problem over the gram matrix $P = \Phi^T \Phi$. Note that the RIP constraints that are nonlinear in Φ can be written as
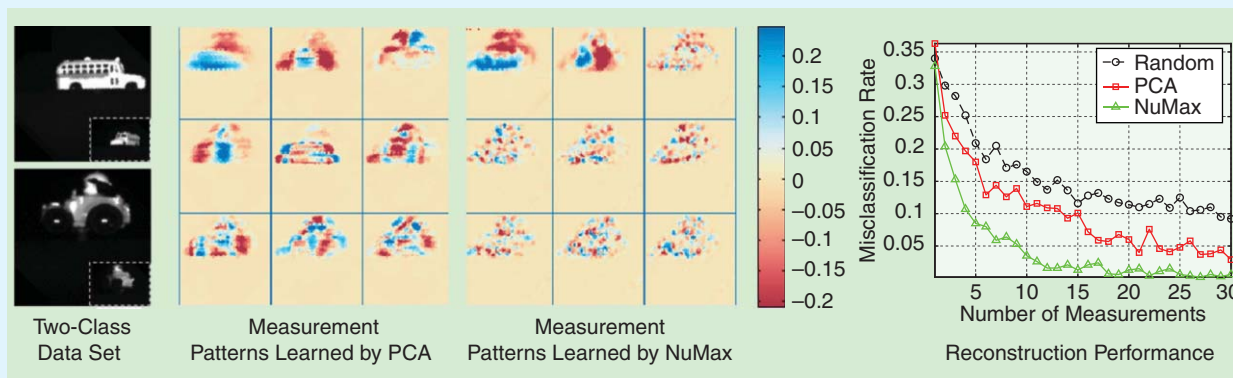
FIGURE S2. Learning measurement operators using NuMax. We compare the performance of NuMax, principal component analysis (PCA) and random projection on a two-class problem. NuMax outperforms the other methods due to is reliance on both the domain (i.e., the data set) and its preservation of neighborhoods via the RIP. (Images used courtesy of [23].)

several criterion, from preservation of the fidelity of extracted features or promotion of class separability. Exploring these ideas could form the basis for impactful future work in this area.

Open problems

The earliest results in CS were broad in their scope and universal in their applicability to all sparse signals. Yet, this implicit simplicity created a significant mismatch to real-world signals that are often enriched with structures that are more complex than sparsity. Efficient sensing and inference with such signals requires a fundamental rethinking of all aspects of CS, including the prime role played by measurement matrices. While this article highlights this important aspect using three case studies, there are many important open problems that need to be addressed to truly harness the potential of nonrandom matrix constructions.

The need for deterministic matrix constructions that rival the performance of random matrices

There has been some limited work on matrix constructions that satisfy the RIP [12] (see "Can We Learn Good Measurement Matrices?"). However, these require $M \sim K^2$ measurements that are significantly worse than random constructions, and they further involve polynomials functions that are not amenable to imaging where there are physical constraints that allow only matrices where entries need to be nonnegative and satisfy energy-preservation constraints. Further, SLMs that are typically used in CI often constrain the measurement matrix to be binary valued. A theory of measurement design that is deterministic while respecting the physical laws of imaging would spur many novel applications.

Going beyond the RIP as the metric of choice

As noted in many earlier works including [12], the RIP is only a sufficient condition for signal recovery and it is well known as a

expressions that are linear in P. In [17], it is shown that a measurement matrix satisfying the RIP can be obtained by solving a semidefinite program given as

$$(\text{NuMax}) \min_{P \in \mathbb{R}^{N \times N}} \| P \|_*, \quad \text{s.t}$$
$$P \geq 0, (1 - \delta) \leq \frac{(x_i - x_j)^T P (x_i - x_j)}{\| x_i - x_j \|^2} \leq (1 + \delta).$$

There are two key properties underlying this optimization problem. First, the objective of the optimization in NuMax is $\| P \|_*$, and the nuclear norm of the matrix P is the convex relaxation of the rank objective and is known to be low-rank promoting [14]. Second, the constraints specify that the square root of P be near isometric on the data set. As a consequence, the optimization problem solves for a low-rank symmetric matrix $P = \Phi^T \Phi$ so that Φ satisfies the RIP on the data set. The key advantage here is that we are learning a measurement matrix that is tailored to the particular data set of interest. Hence, the solution provided by NuMax typically provides significant improvements over random projections in its ability to enable inference at far fewer measurements (see Figure S2), especially as the noise increases.

Task-specific design of measurement operators

Consider the example of NN classification (NNClass). Suppose that we have a data set of labeled points from multiple classes. We can define two sets of difference vectors: 1) intraclass-difference vectors $v_{ij} = x_i - x_j$ when the points x_i and x_j belong to the same class and 2) interclass-difference vectors $w_{mn} = x_m - x_n$ when the points x_m and x_n belong to difference classes. Intuitively, intra-class differences should not expand under measurement operator, and interclass differences should not shrink. This intuitive idea can be formulated into the following optimization problem:

$$(\text{NuMax–Class}) \min_{P \in \mathbb{R}^{N \times N}} \| P \|_*, \quad \text{s.t}$$
$$P \geq 0, \frac{v_{ij}^T P v_{ij}}{v_{ij}^T v_{ij}} \leq C_1 < 1, \frac{w_{mn}^T P w_{mn}}{w_{mn}^T w_{mn}} \geq C_2 > 1.$$

The main difference between NuMax and NuMax-NNClass is that the latter no longer seeks near-isometric embeddings. Instead, we seek a measurement matrix that shrinks the distances between points from the same class, which is enabled by the constraint $v_{ij}^T P v_{ij} / v_{ij}^T v_{ij} \leq C_1$, and increases the distances between points from different classes, which is enabled by the constraint $w_{mn}^T P w_{mn} / w_{mn}^T w_{mn} \geq C_2$. There are two key advantages to this modified optimization. First, note that the constraints in NuMax-NNClass are a subset of those in NuMax, which is easily seen if we set $C_1 = 1 - \delta$ and $C_2 = 1 + \delta$. Hence, the feasibility set of NuMax-Class is larger, and we can hope for a lower nuclear-norm solution (and, hence, a lower-rank solution). Second, given that we are shrinking intraclass distances preferentially and expanding interclass distances, we can promote the NNClass rate directly because there is a greater likelihood that the NN belongs to the same class in the embedded space, resulting in improved classification compared to the pseudorandom measurements.

weak condition [6]. Hence, a criterion that better predicts the phase transition associated with compressive recovery problems would be an invaluable tool for evaluating the efficacy of deterministic matrix constructions. It is worth mentioning that several frameworks have been proposed with the aim of providing a better characterization of performance using metrics other than the RIP, typically using ideas in coherence [8], [9] and spectral norm [32]. The tradeoff here is that the guarantees are probabilistic on the signal space as opposed to universal.

Local features versus global measurements

The inference problems in computer vision often use part-based modeling where local features are constructed to represent an image and its constituents. This approach is extremely robust to changes in background and occlusion, which is critical in machine vision. Inference from compressive measurements, on the other hand, rely on global linear measurements from which

the extraction of local features is extremely hard, if not impossible. Hence, a framework for local feature extraction from compressive measurements would enable us to apply the vast literature in machine vision to compressive inference.

Authors

Aswin C. Sankaranarayanan (saswin@andrew.cmu.edu) is an assistant professor in the Electrical and Computer Engineering (ECE) Department at Carnegie Mellon University. He received his Ph.D. degree from the University of Maryland, College Park, where he was awarded the distinguished dissertation fellowship by the ECE Department in 2009. He was a postdoctoral researcher at the Digital Signal Processing Group at Rice University. His research encompasses problems in compressive sensing and computational imaging. He has received best paper awards at the IEEE Computer Vision and Pattern Recognition Workshops on Computational

Cameras and Displays (2015) and Analysis and Modeling of Faces and Gestures (2010).

Pavan Turaga (pturaga@asu.edu) is an assistant professor in the School of Arts, Media, Engineering, and Electrical Engineering at Arizona State University. He received his B.Tech. degree from the Indian Institute of Technology Guwahati, India, in 2004 and his M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 2008 and 2009, respectively. His research interests are in computer vision and computational imaging with a focus on non-Euclidean and high-dimensional statistical techniques for these applications. He was awarded the Distinguished Dissertation Fellowship in 2009 and received the National Science Foundation CAREER Award in 2015. He is a Senior Member of the IEEE.

Matthew A. Herman (matthew.herman@inviewcorp.com) is a senior algorithm engineer at InView Technology Corporation. He received his Ph.D. degree in applied math in 2009 from the University of California, Davis, with a focus on applications of compressive sensing, such as radar and model mismatches of the sensing/system matrix. His work on compressive sensing radar won the 2013 Best Paper Award from the IEEE Signal Processing Society. From 2009 to 2011, he was a postdoctoral research fellow at the University of California, Los Angeles. Since starting at InView in 2011, he has worked on a variety of novel compressive sensing acquisition and reconstruction algorithms.

Kevin F. Kelly (kkelly@rice.edu) is currently an associate professor in the Electrical and Computer Engineering Department and the chair of the Applied Physics Program at Rice University. He received his B.S. degree in engineering physics from the Colorado School of Mines in 1993 and his Ph.D. degree in applied physics from Rice University in 1999. He was a postdoctoral fellow at the Institute for Materials Research in Sendai, Japan, and in the Chemistry Department at Penn State University. His compressive imaging research was selected as one of the top ten emerging technologies by *Technology Review Magazine*. He cofounded the InView Technology Corporation, a company focused on commercializing compressive imaging.

References

[1] S. Ali and S. Lucey, "Are correlation filters useful for human action recognition?," in *Proc. Intl. Conf. Pattern Recognition*, 2010, Istanbul, Turkey, pp. 2608-2611.

[2] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–121, 2007.

[3] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[4] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[5] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[6] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?," *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, 2011.

[7] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proc. IEEE Intl. Conf. Computer Vision*, 2005, pp. 1395–1402.

[8] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 358–374, 2010.

[9] E. J. Candes and Y. Plan, "A probabilistic and ripless theory of compressed sensing," *IEEE Trans. Info. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[10] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comp. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[11] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[12] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, no. 4, pp. 918–925, 2007.

[13] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Processing*, vol. 21, no. 2, pp. 494–504, 2012.

[14] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. IEEE American Control Conf.*, 2001, pp. 4734–4739.

[15] T. Goldstein, L. Xu, K. F. Kelly, and R. Baraniuk, "The stone transform: Multiresolution image enhancement and compressive video," *IEEE Trans. Image Processing*, vol. 24, no. 12, pp. 5581–5593, Dec. 2015.

[16] M. Harwit and N. J. Sloane, *Hadamard Transform Optics*. New York: Academic Press, 1979.

[17] C. Hegde, A. C. Sankaranarayanan, W. Yin, and R. Baraniuk, "NuMax: A convex approach for learning near-isometric linear embeddings," *IEEE Trans. Signal Processing*, vol. 63, no. 22, pp. 6109–6121, 2015.

[18] P. H. Hennings-Yeoman, B. V. K. V. Kumar, and M. Savvides, "Palmprint classification using multiple advanced correlation filters and palm-specific segmentation," *IEEE Trans. Inform. Forens. Security*, vol. 2, no. 3, pp. 613–622, 2007.

[19] M. A. Herman, "Signature row-block Kronecker products and the perfect shuffle," to appear.

[20] M. A. Herman, "Compressive sensing with partial-complete, multiscale hadamard waveforms," in *Proc. Optical Society of America, Imaging, and Applied Optics*, Arlington, VA, June 2013, p. CM4C-3.

[21] M. A. Herman, T. Weston, L. McMackin, Y. Li, J. Chen, and K. F. Kelly, "Recent results in single-pixel compressive imaging using selective measurement strategies," in *Proc. SPIE: Sensing Technology + Applications*, Compressive Sensing IV, 94840A, Baltimore, MD, May 2015, p. 94840.

[22] K. Kulkarni and P. Turaga, "Reconstruction-free action inference from compressive imagers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 772–784, 2016.

[23] Y. Li, C. Hegde, A. C. Sankaranarayanan, R. Baraniuk, and K. F. Kelly, "Compressive image acquisition and classification via secant projections," *J. Opt.*, vol. 17, no. 6, p. 065701, 2015.

[24] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. thesis, Mass. Inst. Tech., 2009.

[25] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS and T/SPIE Symp. Electronic Imaging: Computational Imaging*, 2007, p. 64980H.

[26] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 1035–1058, 2013.

[27] K. Ni, S. Datta, P. Mahanti, S. Roudenko, and D. Cochran, "Efficient deterministic compressed sensing for images with chirps and Reed-Muller codes," *SIAM J. Imaging Sci.*, vol. 4, no. 3, pp. 931–953, 2011.

[28] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 329–336.

[29] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatio-temporal maximum average correlation height filter for action recognition," in *Proc. IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[30] S. Sadanand and J. J. Corso, "Action bank: A high-level representation of activity in video," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012, pp. 1234–1241.

[31] A. C. Sankaranarayanan, L. Xu, C. Studer, Y. Li, K. F. Kelly, and R. G. Baraniuk, "Video compressive sensing for spatial multiplexing cameras using motion-flow models," *SIAM J. Imaging Sci.*, vol. 8, no. 3, pp. 1489–1518, July 2015.

[32] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harm. Anal.*, vol. 25, no. 1, pp. 1–24, 2008.

[33] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl. Math.*, vol. 123, no. 1–2, pp. 85–100, Nov. 2000.

[34] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," arXiv:1011.3027 (2010).

[35] X. Zhu, S. Liao, Z. Lei, R. Liu, and S. Z. Li, "Feature correlation filter for face recognition," *Adv. Biometrics*, vol. 4642, pp. 77–86, Aug. 2007.

SP

Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan,
Qionghai Dai, Lawrence Carin, and David J. Brady

Computational Snapshot Multispectral Cameras

Toward dynamic capture of the spectral world
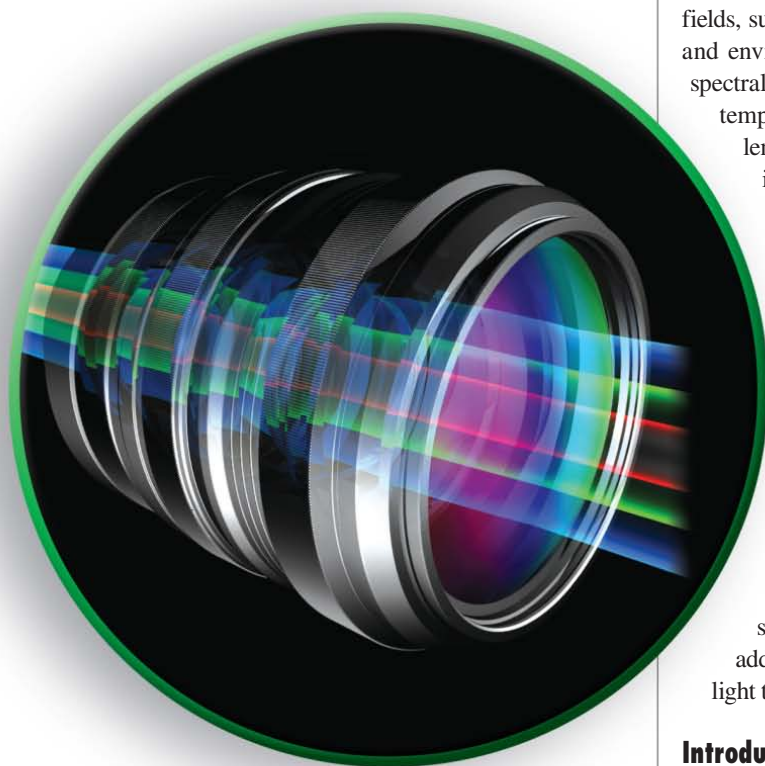
Multispectral cameras collect image data with a greater number of spectral channels than traditional trichromatic sensors, thus providing spectral information at a higher level of detail. Such data are useful in various fields, such as remote sensing, materials science, biophotonics, and environmental monitoring. The massive scale of multispectral data—at high resolutions in the spectral, spatial, and temporal dimensions—has long presented a major challenge in spectrometer design. With recent developments in sampling theory, this problem has become more manageable through use of undersampling and constrained reconstruction techniques. This article presents an overview of these state-of-the-art multispectral acquisition systems, with a particular focus on snapshot multispectral capture, from a signal processing perspective. We propose that undersampling-based multispectral cameras can be understood and compared by examining the efficiency of their sampling schemes, which we formulate as the spectral sensing coherence information between their sensing matrices and spectrum-specific bases learned from a large-scale multispectral image database. We analyze existing snapshot multispectral cameras in this manner, and additionally discuss their optical performance in terms of light throughput and system complexity.

Introduction

The spectrum of a point in a scene is represented by the distribution of its electromagnetic radiation over a range of wavelengths. In conventional digital imaging devices, spectra are measured using three-channel red, green, blue (RGB) sensors, which are designed to coincide with the tristimulus color measurements in the human visual system. However, a triple representation fails to capture the intricate details of natural scene spectra, which arise from the diversity and complexity of illumination and reflectance spectra in the real world. Since various material and object properties can be inferred

©ISTOCKPHOTO.COM/YAKOBCHUK

from detailed spectra, acquisition systems for precise spectral measurements can be effective tools for scientific research and engineering applications. For instance, spectral data can greatly facilitate cancer detection and diagnosis, since certain types of cancer cells have spectral characteristics that differ from those of normal cells [1]. Spectral data can also yield a rich set of features for image analysis. To take advantage of this, spectral capture technology has become widely used in military security, environmental monitoring, biological sciences, medical diagnostics, scientific observation, and many other fields [1]–[7].

Studies in spectrum acquisition have been conducted for decades. Early spectrometers acquire only a single beam of light at a time, which significantly limits their utility for measuring full scenes. Later work focused on efficient, high resolution capture of both the spectral and spatial dimensions. Recently, breakthroughs in temporal resolution have been achieved, which enable simultaneous acquisition of dynamic scenes in the spatial, temporal and spectral dimensions [8]–[10].

Traditional sampling methods [11]–[17], which are based on the Nyquist–Shannon sampling theorem, measure the signal at a certain constant sampling rate on each of the three dimensions. Each sample contains the signal information at a single sampling location, time and wavelength. Sampling multispectral images in all three spatiospectral dimensions requires measurement at a massive scale, and thus making full-sampling schemes, such as those based on scanning or interferometry, impractical in this scenario. That is because scanning a scene on either the spatial dimension or the spectral dimension entails a major sacrifice in the temporal sampling rate. As a result, a full-sampling approach can only be applied in practice on static or slow-moving scenes.

Capitalizing on recent advances in compressive sensing theory, several techniques have been developed based on undersampling and constrained reconstruction, such as computed tomography imaging spectrometry (CTIS) [18] and coded aperture snapshot imaging (CASSI). Within the CASSI paradigm, there are single dispersive CASSI [19], dual dispersive CASSI [20], [21], its dual-coded three-dimensional (3-D) version called the dual-coded snapshot imager (DCSI) [22], the colored 3-D version called the colored coded aperture spectral camera imager (CCASSI) [23]–[25], [47], prism-mask video imaging spectrometry (PMVIS) [26], [27], and single pixel camera spectrometry (SPCS) [28]. The aforementioned systems are all snapshot multispectral cameras, which means that the spectral data are measured in a single exposure (shot) on the camera sensor. There are also other systems that capture multispectral data at video rates, but with more than one measurement per frame, by taking advantage of a rapidly varying optical element such as a spatial light modulator (SLM) or digital micromirror device (DMD), or by adding another camera

into the optical path [30]–[32], [48]. These methods all capture fewer measurements than full-sampling schemes and reconstruct spectra from incomplete data with the aid of regularized reconstruction theory (e.g., utilizing knowledge of signal sparsity in some basis).

A diagram of several coded-aperture-based undersampling snapshot schemes is shown in Figure 1. For better visualization, the target 3-D spectral data cube (x, y, λ) is shown using a two-dimensional (2-D) matrix representing both the spatial (x) domain and the spectral (λ) domain. Such a high-dimensional spectral data cube is not possible to capture in a single exposure using prevalent camera sensors. This has motivated the aforementioned undersampling systems that first capture a low-dimensional projection of the original high-dimensional spectral data. The projection process can be represented as a sensing matrix that projects the spectral and spatial information into a low-dimensional measurement, which is then computationally decoded. To multiplex the spectral and spatial information in a solvable manner, as shown in Figure 1, the coded aperture-based undersampling schemes usually manipulate the original data matrix in two ways: shearing and spatial modulation. These two transforms effectively reorganize the entries of the data matrix and are operable in practice (shearing by a prism or diffraction grating, and spatial modulation by an occlusion mask, spatial light modulator, or digital micromirror device).

> *Since various material and object properties can be inferred from detailed spectra, acquisition systems for precise spectral measurements can be effective tools for scientific research and engineering applications.*

Depending on their optical configurations and exploiting statistical properties of the spectrum data, the aforementioned methods employ different sampling strategies, which result in different sensing performance. In fact, the sampling scheme of a multispectral acquisition system has a significant effect on the reconstruction quality of spectra. On the other hand, in spectrometer design, sampling is also determined by the spectrometer optics and practical issues (e.g., calibration). With the optical design flexibility that is possible through the combination of optical elements (e.g., gratings and prisms) and computational elements (e.g., spatial light modulators or digital micromirror devices), we posit that the effectiveness and efficiency of the sampling scheme should become the principal factor in the design of spectrometers.

Our intent in this article is to present a comprehensive discussion and analysis of existing coded aperture-based multispectral snapshot systems, and link them to different sampling schemes from the signal processing perspective. For each of these coded aperture-based undersampling schemes, efficiency is examined based on the spectral sensing coherence information between its sensing matrix and sparse spectral bases constructed from a multispectral image data set. In addition, the optical properties of the spectrometers, i.e., light throughput, noise tolerance, feasibility, and complexity, are discussed as well. We hope that these analyses and discussions not only provide readers with fresh insight on multispectral imaging,
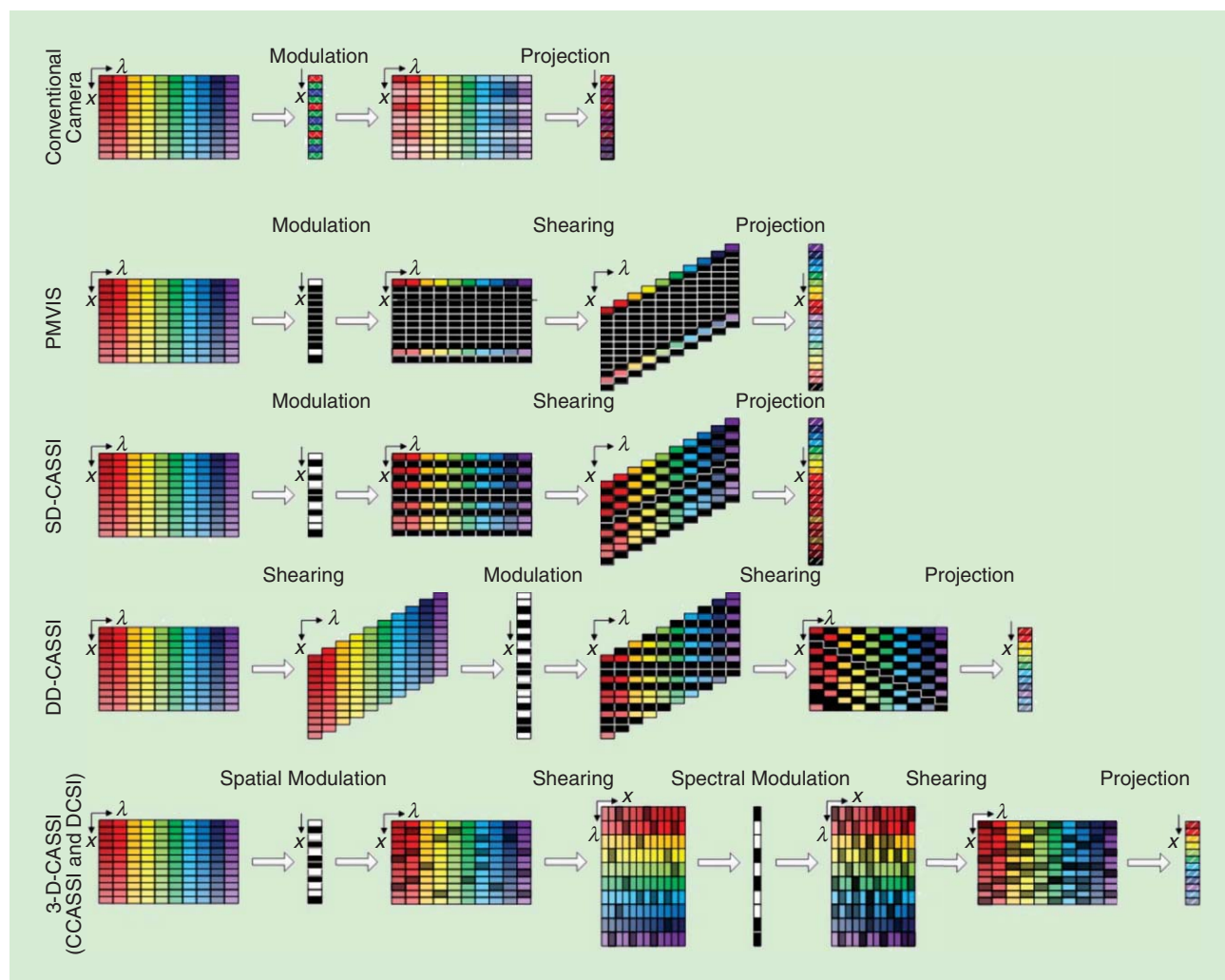
FIGURE 1. Sensing matrices of existing sampling schemes for multispectral acquisition. The spectral and spatial data matrix is high dimensional, and current camera sensors can capture only a low-dimensional projection of the spectral data. The projection process can be regarded as a sensing matrix for the high-dimensional spectral data cube, and state-of-the-art computational multispectral imaging methods can be summarized as different sensing matrices.

but also serve as guidance for designing new multispectral cameras and conducting further study of existing methods.

Full-sampling systems

While mostly focusing on undersampling techniques for multispectral capture, we also paint a fuller picture of multispectral imaging by first reviewing systems designed for full-sampling schemes. Conventional multispectral image acquisition systems are generally based on the Nyquist–Shannon sampling theorem, and thus they sample the signal at twice its maximal frequency. Therefore, due to the considerable amount of data, a sacrifice in either spatial or temporal resolution is needed for these cameras. Such a sacrifice may make full-sampling schemes less practical, thus motivating systems based on compressive measurements.

In spite of the low latency of capture, full-sampling methods for multispectral image acquisition have become widely used in practice. We introduce the basic principles and analyze the performance of full-sampling multispectral acquisition

systems in this section, including three conventional multispectral cameras: filter-based spectrometers, scanning spectrometers, and interferometry-based methods.

Filter-based spectrometers record a sequence of images using a different color filter with each imaging exposure, which effectively samples a set of full spatial resolution images over the spectral range at the expense of temporal resolution. These spectrometers can be easily implemented using a rotating wheel of gel filters, or electronically tunable filters that are typically based on birefringent liquid crystal plates. The measurement scheme of filter-based spectrometers can be viewed as spectral sampling over the temporal domain, with the spatial resolution fully preserved. In such a system, it is important for the color filters to be fabricated with an antireflective coating, to minimize reductions in light throughput.

Instead of varying the filters temporally, scanning spectrometers sweep a spectral sensing device over the scene, sacrificing temporal resolution to gain spatial resolution. Typically, scanning is performed in a whiskbroom or a pushbroom manner.

The whiskbroom design captures the spectrum of a single spatial location at each time instant, and thus requires substantial time to obtain an entire 3-D data cube. Rather than a pinhole aperture, the pushbroom design employs a slit aperture aligned with one of the two spatial dimensions (either x or y), and the spectrometer is translated along the other direction, providing much lower latency than the whiskbroom design. With a scanning-based sensor, the exposure time can be lengthened to increase signal intensity. However, scanning spectrometers involve more mechanical and calibration complexity in practice.

Interferometry techniques (also known as *Fourier transform spectral imaging*), which are based on the principle of interference, project several subimages onto the image sensor, each corresponding to a different color channel. Though a Fourier transform is required to reconstruct multispectral images from raw measurements, interferometry spectrometers are considered to be full-sampling systems because the number of measurements is equal to the number of pixels in the final reconstructed image. These methods sacrifice spatial resolution but avoid spatial discrepancies by directly measuring the spectra of scene points. For such systems, their complexity (with multiple imaging lenses) and precision requirements (on the order of nanometers) make them difficult to build and calibrate.

Undersampling systems

While many methods have been used to construct spectral imagers, this article specifically compares coded aperture-based undersampling designs. The multispectral image information of a dynamic scene spans three domains—spatial, spectral and temporal—presenting an immense amount of data. Just a single second of uncompressed multispectral video

> In spite of the low latency of capture, full-sampling methods for multispectral image acquisition have become widely used in practice.

with a typical 60 spectral bands and only one mega-pixel of spatial resolution is close to two gigabytes. Measuring this amount of data even with short exposure times is infeasible with full-sampling schemes. In capturing multispectral information at video rates, significant undersampling is thus required. Several coded-aperture-based systems have been proposed for multispectral snapshot imaging or video capture. In limiting ourselves to such systems, it is helpful to explain why they are of particular interest to compressive spectral imaging. All spectral imagers take measurements of the form

$$g_i = \int f(x,\lambda) h_i(x,\lambda)\,dx d\lambda, \qquad (1)$$

where $f(x,\lambda)$ is the unknown spectral image and $h_i(x,\lambda)$ is the instrument function for the ith measurement. Such measurements may be point-wise, as in pushbroom systems for which $h_i = \delta(x - x_i, \lambda - \lambda_i)$ ($\delta(\cdot)$ is the Dirac delta function), or multiplexed, as in coded aperture or tomographic systems. Point-wise measurements, however, lack forward model coherence properties consistent with compressive measurement. For compressive measurement one would like to measure weighted groups of unrelated pixels. Representing the spectral data cube as a 2-D space-wavelength structure, CTIS-style systems integrate along lines through the data cube as illustrated in Figure 2.

In the ideal case, one might instead integrate groups of pixels randomly selected from the data cube. Fully random strategies have been implemented for 2-D imaging using single pixel cameras [50]. For tomographic imagers, such as spectral cameras, however, no simple physical mechanism exists for integrating random and independent voxel groups.

The most common form of spectral imager is, of course, the RGB camera, which uses color filter arrays to periodically isolate different color planes. The ideal spectral imager might be similar to an RGB camera but with more diverse and complex spectral filters. Several groups have indeed proposed or implemented spectral imagers using filter arrays [51], [52]. Complex spectral filters are constructed from interference devices. Pixelated interference filters with complex spectral structure are, however, both expensive and difficult to fabricate. In using coded apertures, we find physical advantages in the use of spatial modulation to measure spectral information analogous to the use of spatial delay lines to measure time. Femtosecond pulses are commonly measured using piezoelectric positioning systems with nanometer scale-resolution [53]. Native femtosecond time measurement devices do not exist. Similarly, it is much easier to use a coded aperture with micron scale features to encode a pixelated spectral filter with 10–100 features than to create a similarly complex interference filter. To understand the basic resolution of a coded aperture system, we return to the 2-D data cube discussed previously. As illustrated in Figure 3, we consider a coded aperture with code feature size $\triangle$. A spectral imaging system observes the unknown scene modulated by this code with the spectral planes dispersed by a
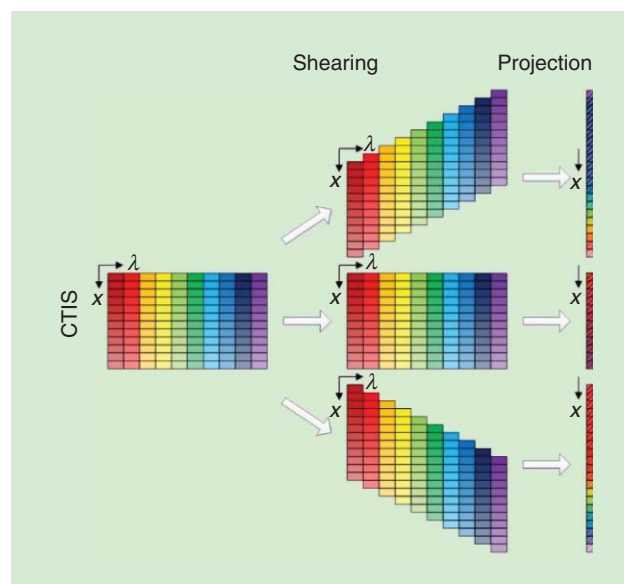
FIGURE 2. Spectral data measurement in the CTIS system.

grating or prism at the rate $(dx/d\lambda)$. A simple single disperser coded aperture system integrates along the wavelength dimension on detection, basically taking tomographic projections along this dimension. Due to the coded aperture, however, features along this dimension are modulated to improve the coherence of the forward model relative to simple tomographic projections. The rate of modulation is easily determined by considering the number of independent code features observed at each detection point. A given code feature is shifted spatially by $\Gamma(dx/d\lambda)$, where Γ is the separation between the shortest and longest wavelength observed. Therefore, integrating along a single wavelength channel, the number of independent wavelength coding elements observed is $N = (\Gamma dx/\Delta d\lambda)$. The spectral resolution is $\Gamma/N = \Delta d\lambda/dx$. For a grating of period L imaged with a lens of focal length F, $(d\lambda/dx) = (L/F)$. With $L = 3$ microns and $F = 3$ cm, for example, a code feature of size 10 microns yields a spectral resolution of 10 nm, corresponding to 30–40 spectral features over the visible range. Better spectral resolution can be obtained with faster gratings or longer focal lengths, but a multiplexing level of 30–40 is already fairly aggressive for snapshot imaging. Multiform integration methods will likely be necessary for more heavily multiplexed systems.

We see, therefore, that coded apertures present a simple and straightforward mechanism for complex spectral filter implementation. In addition, depending on the implementation, they have reasonably local kernels that allow spatially separable data cube estimation.

Even within the family of coded aperture spectral imagers, numerous design choices may be considered for code implementation, dispersive elements, and sensing. Since we cannot comprehensively consider all design choices, here we focus on comparing the coherence of the forward model for several model systems based upon compressive coded aperture designs proposed and demonstrated over the past decade. We specifically do not consider implications of static codes implemented on slides versus dynamic codes implemented using spatial light modulators. While spatial light modulators suffer scatter and numerical aperture limitations not found with static codes, we hope that the reader will find our comparisons without detailed physical implementations sufficiently compelling to postpone full consideration of practical issues.

The coded aperture-based undersampling systems employ different sampling strategies according to their optical configurations and exploit statistical properties of multispectral data, which leads to different sensing performance in terms of spectral reconstruction quality. Figure 4 displays diagrams of four undersampling multispectral cameras. It is worth noting that to facilitate comparison, the diagrams of the systems are drawn not according to the physical configurations proposed

> *The coded aperture-based undersampling systems employ different sampling strategies according to their optical configurations and exploit statistical properties of multispectral data, which leads to different sensing performance in terms of spectral reconstruction quality.*

in the original papers [19]–[27], but rather so that their light paths are equivalent. For DD-CASSI, the original implementation in [20] has two dispersers to realize the dispersion and pixel-wise focusing (i.e., all the spectra of a single point passed through the mask focus on a single pixel), but its diagram in Figure 4 has only one disperser (grating) to achieve the same focusing by just tuning the location of the spatial modulator (mask) and the image sensor. By representing systems with different kinds of modulation (i.e., point-wise coding and sheared coding) and imaging (pixel-wise focusing and dispersed imaging) using similar optical paths in Figure 4, the intrinsic differences between the four systems are revealed. As shown in Figure 4, the PMVIS, SD-CASSI, and DD-CASSI systems only use a single mask to modulate the input light. The main difference between them is the placement of the mask. Both PMVIS and SD-CASSI place the mask on the imaging plane, leading to point-wise coding (i.e., all the spectra of a single point are either passed through or blocked by the mask), while the DD-CASSI places the mask in front of the image plane, which leads to a spectrally sheared coding (i.e., the 3-D code is generated by stacking the same 2-D code with different offsets). In contrast, the sensor of DD-CASSI is put on the image plane to achieve pixel-wise focusing, while PMVIS and SD-CASSI place the sensor behind the imaging plane, which leads to dispersed imaging (i.e., spectra of a single point dispersed to a set of pixels). As for 3-D-CASSI, two masks are utilized to achieve both the spatial and spectral modulation simultaneously, and the sensor is put on the focus plane to ensure pixel-wise focusing.

Prism-mask video imaging spectrometry
PMVIS [26], [27] straightforwardly acquires the spectra of scene points with the aid of a prism and utilizes a mask with uniformly distributed holes that prevent overlaps of the
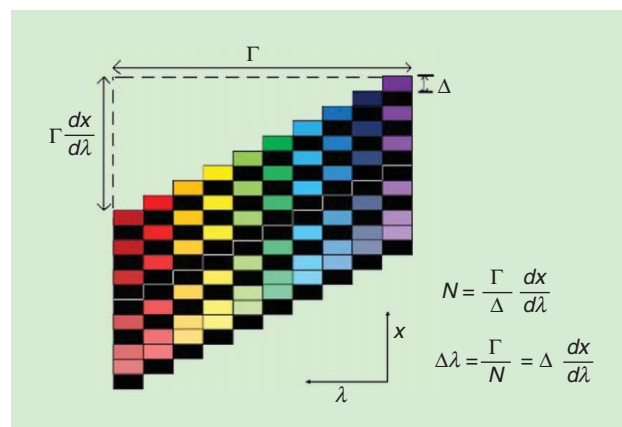
FIGURE 3. A diagram of the relationship between the spectral resolution and the code feature size.

$$N = \frac{\Gamma}{\Delta}\,\frac{dx}{d\lambda}$$

$$\Delta\lambda = \frac{\Gamma}{N} = \Delta\,\frac{dx}{d\lambda}$$
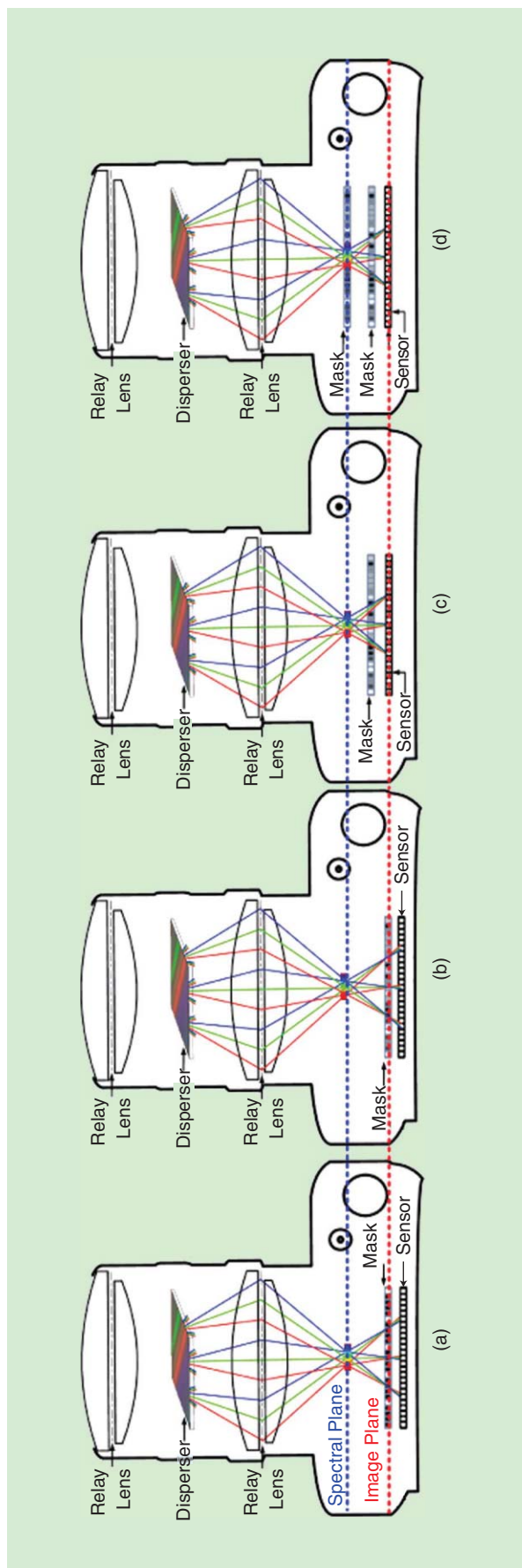
FIGURE 4. Diagrams of four undersampling multispectral cameras: (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI.

dispersed spectra on the sensor, as shown in Figure 4(a). The spectral values of the sample points are measured directly without any spectral modulation, and there exists a known correspondence between spectral bands and sensor plane locations. Each of the measurements represents a certain spectral intensity value of its corresponding scene point. As shown in Figure 5(a), this system sacrifices spatial resolution to achieve high spectral resolution. Cao et al. [27], [28] extended this idea with a hybrid PMVIS scheme in which a high spatial resolution RGB image is simultaneously acquired with each multispectral snapshot. Through a spatial interpolation within the spectral frame that is guided by the high resolution RGB image, a final result is computed with high resolution in both the spatial and spectral dimensions.

Single dispersive coded aperture snapshot imager

According to compressive sensing theory [54], if a signal has a low-dimensional representation (e.g., it can be represented as a sparse combination of orthonormal bases, like wavelets), then it can be reconstructed from a small set of measurements. With an appropriate sampling scheme, the samples needed to reconstruct a signal can be fewer than those specified by the Nyquist–Shannon limit. Based on this concept, various undersampling systems have been developed to reconstruct entire spectra from fewer measurements.

Wagadarikar et al. proposed the CASSI system using a single disperser [19], which we will refer to as SD-CASSI. The spectral data cube is modulated by a coded mask and dispersion, as shown in Figure 4(b). Light rays of different wavelengths are modulated by an aperture code and then are offset differently by a dispersive element, which results in a coded and sheared 3-D cube as illustrated in Figure 4 before projection onto the CCD sensor. The imager captures a 2-D projection of the coded and sheared cube as shown in Figure 4. After the undersampling and spectrally multiplexed capture, the complete data cube is reconstructed based on the prior that spatial-spectral information is sparse in the wavelet domain.

The CASSI system implements measurement matrices of a specific structure, i.e., a replicated and slanted 2-D code along the spectral dimension, illustrated as the SD-CASSI projection in Figure 5(b). The mathematical formulation of the 3-D code can be expressed as

$$C(x,y,\lambda) = C_{2D}(x - \Delta s \cdot \lambda, y) = \text{reshape}(\mathbf{TC}), \qquad (2)$$

where (x,y) and λ are the spatial and spectral indices, C_{2D} is a randomly generated 2-D spatial coding pattern, Δs is the offset of each channel caused by dispersion, C is the column vector form of the unsheared modulation code, T is the shearing operation matrix, and $reshape(\cdot)$ is the reshape function to transform the column vector to the original 3-D data cube.

The image is modulated before the dispersive element, and then the disperser shears the modulated image. Thus the measurements can be modeled as

$$\mathbf{M} = \mathbf{\Phi S} = \text{vector}\left(\sum_\lambda \text{reshape}(T\text{diag}(C)S)\right), \qquad (3)$$

where $\mathbf{M}$ is the column vector form of the measured values, $\mathbf{\Phi}$ is the sensing matrix, $\mathbf{S}$ is the column vector form of the spectral data cube, diag() is the diagonal operator to transform a vector to a same order square matrix with the elements of the input on the diagonal, reshape() is the reshaping operator to transform the column vector to the 3-D data cube, and vector() is the vectorizing operator for transforming the 2-D matrix to a column vector. Mathematically, the operation $\mathrm{vector}\Sigma_\lambda\mathrm{reshape}$ can be regarded as a down-sampling operator, and each element of the output $\mathbf{M}$ is the summation of a set of certain elements of the input vector $\mathbf{T}\mathrm{diag}(\mathbf{C})\mathbf{S}$. In other words, the combined operator $\mathrm{vector}\Sigma_\lambda\mathrm{reshape}$ can be represented by a single short matrix. Then, the sensing matrix of SD-CASSI can be represented by

$$\mathbf{\Phi} = \mathbf{\Sigma T}\mathrm{diag}(\mathbf{C}), \qquad (4)$$

where $\mathbf{\Sigma}$ denotes the short matrix form of the combined summation operator $\mathrm{vector}\Sigma_\lambda\mathrm{reshape}$. According to (4), the imaging procedure represented by the summation matrix $\mathbf{\Sigma}$ just follows the shearing operation $\mathbf{T}$, which implies dispersed imaging, i.e., the spectra of a single point is dispersed to a set of pixels. The coding matrix $\mathrm{diag}(\mathbf{C})$ manipulates the original spectral data cube directly, which corresponds to point-wise coding (i.e., all the spectra of a single point are either passed through or blocked by the mask).

Dual dispersive coded aperture snapshot imager

Since the basic CASSI system forms a sheared 3-D spectral data cube, the observed snapshot is blurred by dispersion. To overcome this effect, Gehm et al. [20] proposed a dual-disperser architecture (DD-CASSI) in which two dispersers are symmetrically placed on the two sides of the coded aperture to produce an unsheared spectral cube with replicated slanted code. Lin et al. [21] proposed a single disperser (grating) system called spatial-spectral encoded compressive spectral imager (SSCSI) to realize the same function as shown

> **DD-CASSI has exactly the same coding space as the SD-CASSI system, but the modulated 3-D spectral data cube is not sheared (i.e., it is sheared back after modulation).**

in Figure 4(b) but with less calibration difficulty than the dual-disperser CASSI. Although these two systems adopt a similar code, they employ different sampling and reconstruction methods. A recursive offset code is applied by Gehm et al. [20] to achieve fast block-wise processing, while a random and nonuniform code as well as a dictionary-based reconstruction algorithm are employed by Lin et al. [21] to obtain high-quality performance.

DD-CASSI has exactly the same coding space as the SD-CASSI system, but the modulated 3-D spectral data cube is not sheared (i.e., it is sheared back after modulation), as shown in Figure 5(c). Thus, the measurement matrix becomes

$$\mathbf{\Phi} = \mathbf{\Sigma T}^T\mathrm{diag}(\mathbf{C})\mathbf{T} = \mathbf{\Sigma}\mathrm{diag}(\mathbf{TC}). \qquad (5)$$

According to (5), the transpose $\mathbf{T}^T$ is the inverse shearing matrix, which is used to unshear the sheared cube. As for the second term on the right of (5), the diagonal coding matrix $\mathrm{diag}(\mathbf{TC})$ modulates the original spectral data cube, which implies spectrally sheared coding (i.e., the 3-D code is generated by stacking the same 2-D code with different offsets). In this case, the shearing matrix $\mathbf{T}$ only shears the 3-D code, and the spectral data is not affected. Meanwhile, the summation matrix $\mathbf{\Sigma}$ integrates the coded data cube along the spectral dimension, which represents the pixel-wise focusing, i.e., all the passed spectra of a single point are focused on a single pixel.

Spatial-spectral coded compressive spectral imager

The feasible codes for both the basic CASSI and the nonspatially modulated imager are limited by their physical modulation capabilities. Theoretically, 3-D-CASSI, which encodes the spatial-spectral data cube randomly, can achieve more feasible codes and higher performance, as shown in Figure 4(b). However, the physical implementation of 3-D-CASSI is not trivial.

To approximate the 3-D modulation in spatial-spectral data cube, two coded aperture-based systems, i.e., the color-coded
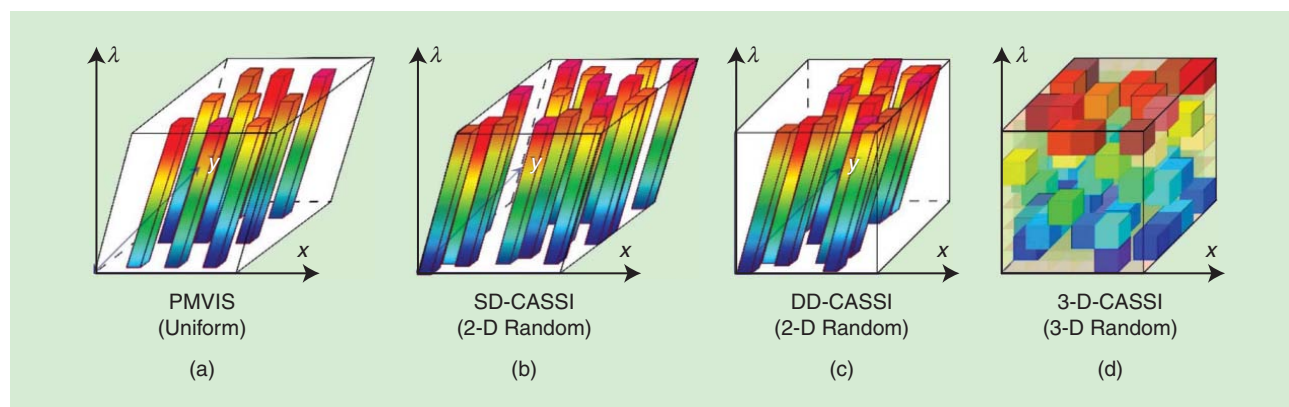
FIGURE 5. An illustration of voxel sensing schemes of four types of undersampling multispectral cameras. The sampling schemes of the (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI systems.
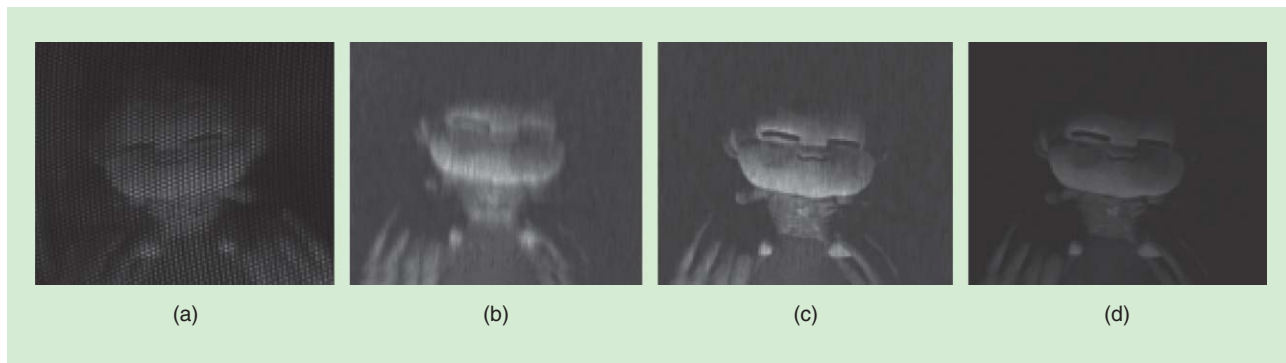
FIGURE 6. Measurement snapshots of the (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI systems.

aperture spectral camera imager (also known as CCASSI) and DCSI are proposed. Correa et al. [23], [47], Arguello et al. [24], and Rueda et al. [25] utilize the RGB colored sensor as a spatial-spectral modulator to achieve the specific dual coded (in spatial and spectral dimensions) compressive spectral imager (CCASSI). By combing the separable codes of all the spectral channels, the CCASSI can achieve more complex modulation than SD-CASSI and DD-CASSI. Similarly, Lin et al. [22] proposed a system that consists of two controllable modulators (e.g., digital mirror devices) on both the spectral and spatial plane, and introduce the dynamic modulation, i.e., changing the codes of the spectral and spatial planes during the exposure time, to enable more flexible modulation. Mathematically, the composited 3-D spectral code $C_{CCASSI}(x, y, \lambda)$ of CCASSI and $C_{DCSI}(x, y, \lambda)$ of DCSI can be represented by the sum of a set of separable codes:

$$C_{CCASSI}(x, y, \lambda) = \sum_{c \in \{\text{channel set}\}} C_c^{\text{spatial}}(x, y) C_c^{\text{spectral}}(\lambda),$$

$$C_{DCSI}(x, y, \lambda) = \sum_t C_t^{\text{spatial}}(x, y) C_t^{\text{spectral}}(\lambda), \quad (6)$$

where (x, y) and λ are the spatial and spectral indices, c and t index the spectral channels and time slices respectively, C_t^{spatial} and C_t^{spectral} are the spatial and spectral codes of the CASSI system for channel c, and C_t^{spatial} and C_t^{spectral} are the spatial and spectral codes of the DCSI system at time t. Since both the spectral data cube and the coding pattern are not sheared in this system, the measurement matrix is

$$\mathbf{\Phi} = \Sigma diag(\mathbf{C}). \quad (7)$$

The ideal 3-D-CASSI can in principle produce any 3-D code, as shown in Figure 5(d). Both CCASSI and DCSI are the approximate implementations of the ideal 3-D-CASSI. According to (7), there is no shearing matrix in the sensing matrix, which implies pixel-wise focus and nonrestricted coding [as in (6)] on both the spatial and spectral dimensions. Thus, 3-D-CASSI

> **For the four types of undersampling systems, we examine their sampling efficiency based on the spectral sensing coherence information of their sensing matrices, and then evaluate their reconstruction accuracy on a diverse multispectral database containing images of various scenes.**

provides a larger feasible code space than the SD-CASSI and DD-CASSI systems.

All of the coded aperture-based systems capture images with the CCD sensor placed on the image plane. The sensing step corresponds to integrating the 3-D spectral data cube along the spectral dimension, yielding snapshots that are blurred from dispersion (SD-CASSI) or not (DD-CASSI and 3-D-CASSI) with modulated patterns. For the PMVIS system, the mask is placed on the image plane to obtain uniform sampling, and the sensor is located beyond the image plane by a certain distance to ensure that the dispersive spectral bands of the sampling points fill the sensor without overlapping one another. Figure 6 exhibits snapshot measurements on the sensor for the four systems.

Sensing matrix and spectral sensing coherence
To compare the multispectral sensing ability of these undersampling systems, we analyze their sampling efficiency. Consider the following theorem [2], [34], [35].

For a given signal $f \in \mathbb{R}^n$, suppose that its coefficient sequence x in the orthonormal basis $\mathbf{\Psi}$ is S-parse, i.e., the coefficient sequence x has S nonzero elements. Then with m randomly selected measurements in the $\mathbf{\Phi}$ (sensing matrix) domain, the signal f can be exactly reconstructed through L1 minimization with overwhelming probability if

$$m \geq c \, \mu^2(\mathbf{\Phi}, \mathbf{\Psi}) \, S \, \log n, \quad (8)$$

where $\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle|$ is the coherence between the sensing matrix $\mathbf{\Phi}$ and sparse domain bases $\mathbf{\Psi}$, and c is a known positive constant.

According to this theorem, a smaller coherence $\mu(\mathbf{\Phi}, \mathbf{\Psi})$ indicates that fewer measurements are needed for complete reconstruction, and therefore the sensing system has higher sampling efficiency. In general, a randomly generated measurement matrix would be effective for most signals. However, for a specific task like multispectral imaging where signals exhibit commonalities that allow representation with a sparse basis or dictionary, the measurement matrix $\mathbf{\Phi}$ can be designed to achieve better performance.

In this article, we measure the quality of a sensing matrix as its spectral sensing coherence information with sparse domain bases: $\jmath_m(\mathbf{\Phi}, \mathbf{\Psi}) \overset{\text{def}}{=} \| I - \mathbf{\Psi}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{\Psi} \|_F$, where I is the identity matrix. Minimization of $\jmath_m(\mathbf{\Phi}, \mathbf{\Psi})$ imposes the condition that the Gram matrix $\mathbf{\Psi}^T \mathbf{\Phi}^T \mathbf{\Phi} \mathbf{\Psi}$ be as close as possible to the identity matrix, which provides a good sensing matrix as well.

Evaluation of undersampling systems

For the four types of undersampling systems, we examine their sampling efficiency based on the spectral sensing coherence information of their sensing matrices, and then evaluate their reconstruction accuracy on a diverse multispectral database containing images of various scenes—including indoor scenes, outdoor scenes, various materials and different illuminations—from four online data sets [40]–[43]. A few example images are shown in Figure 7.

Computation of spectral sensing coherence information and image reconstruction

The spectral sensing coherence information is computed with respect to a domain basis in which the signals can be sparsely represented. From the multispectral image database, we learn two kinds of bases $\mathbf{\Psi}$ in which multispectral images have a sparse representation. The first is from principal components analysis (PCA) [45], which is applied to derive an orthonormal bases. The second is from the K-SVD algorithm [39], which is used to obtain an overcomplete dictionary. The bases represent the specific structural characteristics of the multispectral images and video frames, and thus are suitable for computing spectral sensing coherence information $\jmath_m(\mathbf{\Phi}, \mathbf{\Psi})$ and analyzing the sampling efficiency of the undersampling schemes for multispectral acquisition systems.

In computing the PCA bases and the overcomplete dictionary, we use 100,000 multispectral patches of size $10 \times 10 \times 29$ pixels (horizontal $\times$ vertical $\times$ spectral) that are randomly sampled from the database. The size of each basis element is thus $10 \times 10 \times 29$ as well. Since the PCA bases are orthonormal and complete, it has a size of exactly 2,900. For K-SVD, 6,200 atoms are learned as a sparse representation of the natural multispectral images.

We also synthetically test the reconstruction accuracy of the four undersampling multispectral imaging systems on the

| 420 nm | 450 nm | 480 nm | 510 nm | 540 nm | 570 nm | 600 nm | 630 nm | 660 nm | 690 nm |
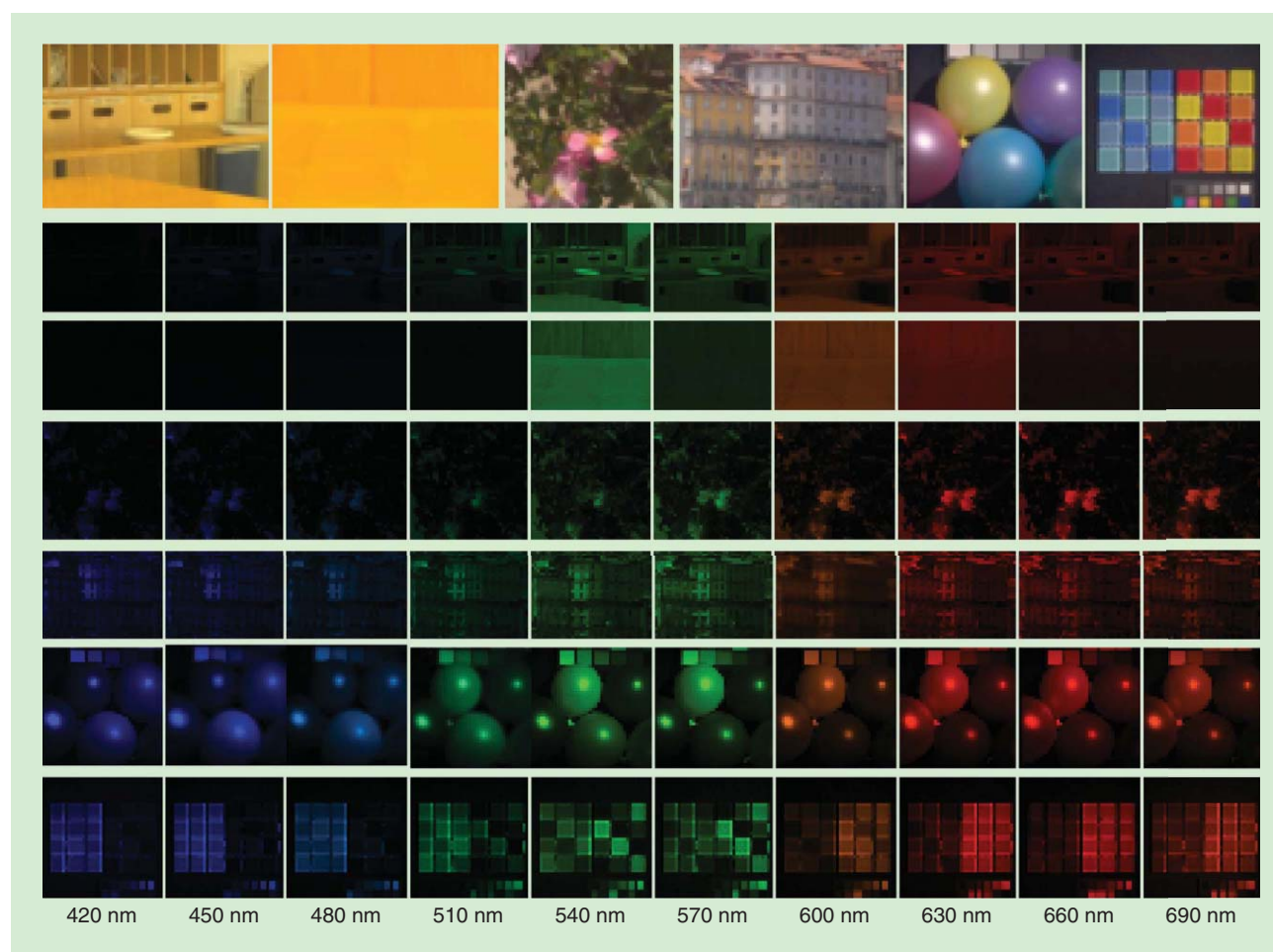
FIGURE 7. Six example images from the multispectral database, including indoor and outdoor scenes, various materials, and different illumination. Ten of the 29 spectral channels (from 420 nm to 700 nm, at 10 nm intervals) are shown. The corresponding RGB images are displayed in the top row.
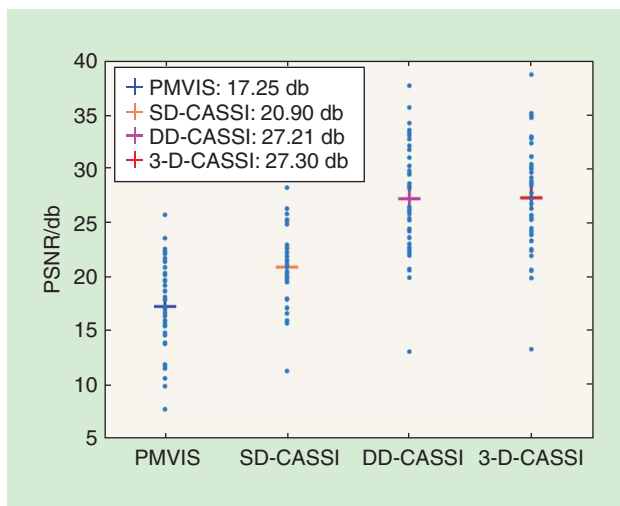
FIGURE 8. PSNR comparison for image reconstruction with the four types of undersampling systems on a set of 50 multispectral images. Crosses of different colors mark the average PSNR for the different methods. The PNSR value for each individual multispectral image is also plotted, as blue dots, to illustrate the statistical distribution of the reconstruction accuracy. Except for PMVIS in which linear interpolation is used for reconstruction, we use the ADMM [46] to compute the reconstruction results.

Table 1. The spectral sensing coherence information between the sensing matrices of different systems and the learned bases.

	SD-CASSI	DD-CASSI	3-D-CASSI	PMVIS
K-SVD	0.7920	0.7787	0.7737	0.8148
PCA	0.7048	0.6432	0.6663	0.7251

Table 1 presents the spectral sensing coherence information values computed between the sensing matrices of the four types of the undersampling systems and the three kinds of bases. Note that since hybrid PMVIS [30], [31] and hybrid CASSI [32] each obtain two snapshots, they are omitted in this analysis for an even comparison. Multiple snapshot systems are discussed in the section "Evaluation of Undersampling Systems." For the coded aperture-based systems, binary codes randomly generated by the Bernoulli distribution, with the same probability $p(x = 1) = 0.5$, are applied.

Specifically, the codes of SD-CASSI and DD-CASSI are derived by shifting and stacking the randomly generated 2-D patterns. As for 3-D-CASSI, the code is generated directly in 3-D space. Both the K-SVD and PCA bases are learned from the database.

The 3-D-CASSI system has the most complex modulation and achieves the best spectral sensing coherence information on the overcomplete dictionary learned by the K-SVD algorithm. However, for the PCA bases, DD-CASSI provides the best spectral sensing coherence information. For both of the bases, the coherences of DD-CASSI and 3-D-CASSI are very close, which indicates comparable quality of their sensing matrices. It is shown in Figure 7 that DD-CASSI and 3-D-CASSI also perform comparably on hyperspectral image reconstruction accuracy, which is consistent with the theorem discussed in the section "Understampling Systems."

Aside from DD-CASSI and 3-D-CASSI, the coherence values of the other systems have a consistent ranking on both the PCA and K-SVD bases, which suggests that the relative quality of sensing matrices is not greatly affected by the bases, if they represent the sparse structure of the data well. This is also indicated by the reconstruction results in Figure 8.

The reconstruction performance of the four undersampling systems is displayed for the 610-nm channel of an example image in Figure 9. The result of PMVIS exhibits blocking

database images. The inputs of the four systems are generated by sampling the multispectral images according to the corresponding sensing matrices described in the section "Undersampling Systems." Image reconstruction is performed using a widely employed algorithm—the alternating direction method of multipliers (ADMM) [46]—except for PMVIS which simply employs linear interpolation (as it cannot be solved by ADMM directly because of its special sampling scheme). ADMM is widely used in image reconstruction and has shown superior performance. It is worth noting that the choice of the algorithm may affect the reconstruction accuracy, but the ranking of the results does not change.

In testing PMVIS, we use an image down-sampling rate of 0.3%, as is the case in the prototype camera [27]. Theoretically, in PMVIS systems, a minimal down-sampling rate of $1/\Omega$ (where Ω is the number of spectral channels) is needed to prevent overlaps between the spectra of different samples. The current prototypes are not well calibrated, so the down-sampling rate may potentially be improved in the future.

FIGURE 9. A comparison of reconstructed results for the four undersampling systems. All of the results are shown at the 610-nm channel. (a) PMVIS PSNR = 16.6845 db, (b) SD-CASSI PSNR = 18.0859 db, (c) DD-CASSI PSNR = 29.8178 db, and (d) 3-D-CASSI PSNR = 32.5659 db.

artifacts due to its low sampling rate in the spatial domain. The results for the other three coded aperture systems are of much higher quality. 3-D-CASSI produces particularly good results in this example.

It is worth noting that the random code may not be optimal for specific data such as multispectral images, which exhibit certain characteristics and strong redundancy. In regard to this, the feasible space of coding patterns for the coded aperture-based systems is constrained by the light paths of the systems. With a larger feasible domain, there is greater potential for a system to achieve higher performance. Since 3-D-CASSI, whose coding space completely encompasses those of the other three systems, does not exhibit much superiority over the other systems in our experiments, we believe that a random code is far from optimal in the multispectral imaging scenario.

Analysis of light throughput and system complexity

Besides the sensing matrix, the light throughput and calibration error also affect the reconstruction accuracy. For discussion of these factors and practical system complexity, we list the light throughput and the number of optical elements in Table 2. For the PMVIS system, its light throughput is determined by its down-sampling rate, which is the reciprocal of the number of spectral channels Ω. For the typical multispectral imaging scenario, with 30 or more spectral channels, the light throughput loss of PMVIS is relatively large. Both SD-CASSI and DD-CASSI have a light throughput of 0.5, while that of 3-D-CASSI is 0.25 because of its two modulators. With regard to system complexity and calibration difficulty, PMVIS and SD-CASSI are relatively simple and easy to calibrate because of their smaller number of optical elements and simpler light paths. Particularly, PMVIS is much more robust to calibration errors (e.g., slight shifts or rotations of the coded aperture) because its reconstruction algorithm is based on simple interpolation, which makes the system highly practical. The number of optical elements also has a strong influence on calibration and light throughput, and thus it affects the signal-to-noise ratio of the captured multispectral images. Although PMVIS and SD-CASSI have lower reconstruction accuracy on synthetic data as shown in Figure 8, this gap is narrowed by taking their practical benefits into consideration.

As shown in Figure 10, it is clear that when the sensing noise increases, the performance gap between DD-CASSI/3-D-CASSI

> **With regard to system complexity and calibration difficulty, PMVIS and SD-CASSI are relatively simple and easy to calibrate because of their smaller number of optical elements and simpler light paths.**

and SD-CASSI/PMVIS decreases rapidly. The reconstruction results of all the systems are degraded with the increase of sensing noise. However, with greater system complexity there is more degradation in performance. Considering the high complexity of DD-CASSI and 3-D-CASSI, which leads to lower light throughput and larger calibration errors, the advantages of the complex coded aperture systems may be counteracted by the effect of sensing noise. Thus, further investigation is needed for reducing the light path complexity of coded aperture-based spectral imaging systems and improving the noisy tolerance of the reconstruction algorithms.

Discussions and future directions

Video-rate multispectral cameras with multiple snapshots

We have focused on multispectral video imagers with single snapshot measurements thus far, but there exist other systems [2], [30]–[32], [48] that acquire two or more snapshot measurements to recover the spectral information with higher accuracy while still at video rates. These systems can be also used for multispectral capture of dynamic scenes.

Multiple snapshots have been acquired in two ways. One is by adding extra cameras into the optical path. Figure 11 shows one design for such an implementation, where the incoming light rays are first directed along two separate paths by a beam splitter, essentially making two copies of the light rays, each with a lower light intensity. One of the paths enters the optical configuration of an undersampling system (e.g., PMVIS or CASSI), while the other light path may lead to an RGB or grayscale camera to record a high spatial resolution image of the scene. This hybrid camera design has been implemented based on PMVIS [30], [31] and CASSI

Table 2. Typical parameters for the four types of undersampling systems.

	SD-CASSI	DD-CASSI	3-D-CASSI	PMVIS
Light throughput	0.5	0.5	0.25	$\frac{1}{\Omega}$
Number of optical elements	6	9	8	6
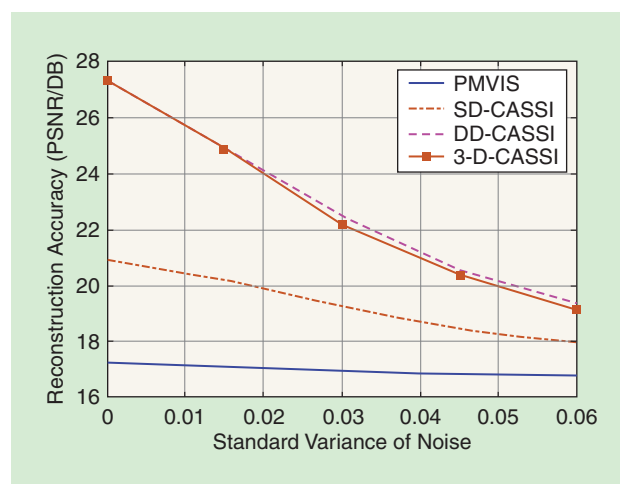
FIGURE 10. Noise tolerance curves of the four kinds of spectral imaging systems. (The image intensity is normalized to 0~1.)
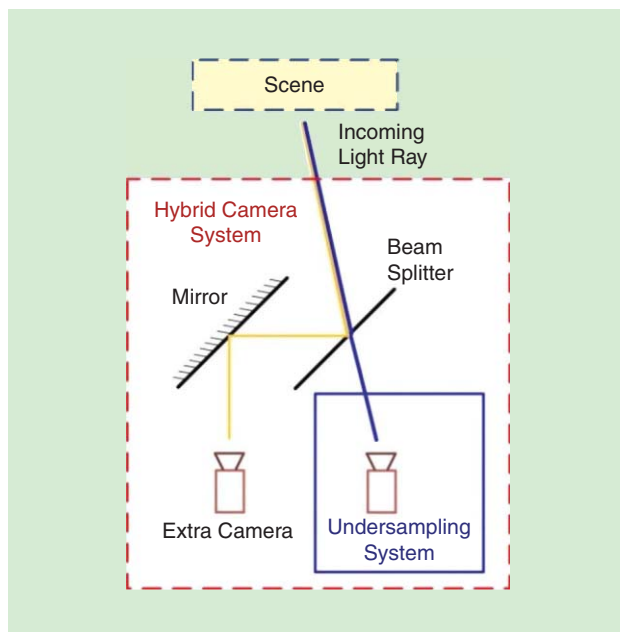
FIGURE 11. A hybrid camera design for multiple snapshot measurements.

Table 3. PSNR comparison of three multisnapshot systems.

Systems	Hybrid PMVIS (+ Extra Camera)		SD-CASSI (Measure Twice)	Hybrid CASSI (+ Extra Camera)
	Simple Algorithm	Complex Algorithm		
PSNR (DB)	25.86	33.19	28.04	32.10

[32], [48]. We also conducted spectral reconstruction experiments using this hybrid design by adding another full spatial resolution image as part of the input (RGB for PMVIS, and gray scale for CASSI). As for the PMVIS system, two kinds of reconstruction algorithms, i.e., a simple bilateral propagation-based method [30] and a more complex learning-based method [49], are applied.

The results in Table 3, which are derived by averaging the results on the aforementioned spectral image database, show that the reconstruction accuracy is increased by about 8 dB and 5 dB on average for PMVIS and CASSI, respectively. This tremendous gain in signal recovery demonstrates the effectiveness of a hybrid camera design that includes an additional basic sensor. In addition, the complex learning-based algorithm achieves about 8 dB improvement over the simple bilateral propagation method for the PMVIS system, which shows the great potential of improving existing reconstruction algorithms.

The other method for acquiring extra snapshot measurements is to use high-frequency optical elements and sensors that allow multiple snapshots to be captured for each multispectral video frame. Spatial light modulators or digital micro-mirror devices operating at 120 Hz or above can be used for this purpose in conjunction with high-speed camera sensors, all of which have become increasingly affordable in recent years. Systems based on this design have been successfully built for multiple snapshot multispectral video capture [2], [29]. Significant gains in reconstruction accuracy (about 5 dB) have been reported in comparison to the single snapshot

> **With the advances in signal processing theory and algorithms and the increasing demand for high-resolution multispectral images/ videos, undersampling schemes for multispectral image acquisition have become a hot topic in computational photography and signal processing.**

CASSI system in [2]. As a result, in practice, by adding an extra sensor or using an ultrafast coded aperture, greater accuracy in multispectral acquisition can be obtained with some increase in system cost.

Advanced theory in reconstruction from undersampled signals

The use of random projections in compressed measurements was originally motivated by the idea that many signals of interest may be represented sparsely in an orthonormal basis, such as the wavelet transform. However, sparsity represents only one class of signal model, and other models may lead to other forms of compressive measurement that may perform better than random projections. For example, it has been demonstrated that many signals of interest may be represented in terms of a union of low-dimensional linear subspaces [35], [37], [44], [48]. From a statistical or signal processing perspective, such a model may be represented as a Gaussian mixture model (GMM), in which the covariance matrix of each mixture component is low rank [36]–[38]. Recent theory has shown that good measurement matrices correspond to projections that are aligned with the signal space [36], [37]. There is already evidence to demonstrate that such a signal model, which may be learned based on the data [38], is well suited to the multispectral data of interest. This model will greatly facilitate the measurement design of novel multispectral video cameras.

With the advances in signal processing theory and algorithms and the increasing demand for high-resolution multispectral images/videos, undersampling schemes for multispectral image acquisition have become a hot topic in computational photography and signal processing. A number of undersampling-based acquisition systems have been proposed, but there had been little analysis of their relative effectiveness. In this article, we have examined existing multispectral video systems based on their sampling efficiency and optical performance, from a signal processing perspective. We introduced the spectral sensing coherence information of the sensing matrix and bases learned from multispectral data as a metric for comparing the sampling efficiency of different systems. From

these analyses, readers may be inspired to design or develop better sampling schemes for multispectral sensing.

Acknowledgments

Authors

Xun Cao (caoxun@nju.edu.cn) received his B.S. degree from Nanjing University, China, in 2006 and his Ph.D. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2012. He is currently an associate professor at the School of Electronic Science and Engineering, Nanjing University, China. He was a visiting researcher at Philips Research, Aachen, Germany, in 2008 and Microsoft Research Asia, Beijing, China, during 2009–2010. He was a visiting scholar at the University of Texas at Austin, Texas, from 2010 to 2011. His research interests include computational photography, image-based modeling, and rendering.

Tao Yue (yuetao@nju.edu.cn) received the B.S. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2009 and the Ph.D. degree from Tsinghua University, Beijing, China, in 2015. He is currently an associate researcher with the School of Electronic Science and Engineering, Nanjing University, China. His research interests mainly include computer vision, image processing, and computational photography.

Xing Lin (xinglin@stanford.edu) received the B.E. degree in electronic science and technology from Xidian University, Xi'an, China, in 2010 and the Ph.D. degree in automation from Tsinghua University, Beijing, China, in 2015. He is a research associate with the Howard Hughes Medical Institute at the Department of Biology, Stanford University, California. His research interests include computational imaging, computer vision, and optical bioimaging.

Stephen Lin (stevelin@microsoft.com) received the B.S.E. degree from Princeton University, New Jersey, and the Ph.D. degree from the University of Michigan, Ann Arbor. He is currently a senior researcher with the Internet Graphics Group of Microsoft Research, Beijing, China. His current research interests include computer vision, image processing, and computer graphics. He served as a program cochair for the International Conference on Computer Vision 2011 and the Pacific-Rim Symposium on Image and Video Technology 2009.

Xin Yuan (x.yuan@alcatel-lucent.com) is currently a video analysis and coding lead researcher at Bell Labs, Murray Hill, New Jersey. He had previosuly been a postdoctoral associate with the Department of Electrical and Computer Engineering, Duke University, from 2012 to 2015, where he was working on compressive sensing and machine learning. Before joining Duke, he obtained his B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, in 2007 and 2009, respectively, and his Ph.D. degree from the Hong Kong Polytechnic University in 2012. His research interests are in signal processing, computational imaging, and machine learning.

Qionghai Dai (qionghaidai@tsinghua.edu.cn) received his Ph.D. degree in automation from Northeastern University, Shenyang, China, in 1996. He has been a faculty member since 1997 and a professor since 2005 of the Department of Automation, Tsinghua University, Beijing, China. He has published more than 120 conference and journal papers, and he holds 67 patents. His current research interests include the areas of computational photography, computational optical sensing, and compressed sensing imaging and vision. His work is motivated by challenging applications in the fields of computer vision, computer graphics, and robotics.

Lawrence Carin (lcarin@duke.edu) received his B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1985, 1986, and 1989, respectively. In 1989, he joined the Department of Electrical Engineering at Polytechnic University, Brooklyn, New York, as an assistant professor, where he became an associate professor in 1994. In 1995, he joined the Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, where he is currently a professor. He was the chair of the Department of Electrical Communication Engineering from 2011 to 2014. Since 2014, he has been the vice provost of research at Duke University. His current research interests include machine learning and statistics. He is a Fellow of the IEEE.

David J. Brady (dbrady@duke.edu) is a professor of electrical and computer engineering at Duke University and Duke Kunshan University. He is a graduate of Macalester College and Caltech and was previously on the faculty of the University of Illinois. He is the author of *Optical Imaging and Spectroscopy*. His research focuses on array cameras and compressive tomography. He is a Fellow of IEEE, OSA, and SPIE, and he won the 2013 SPIE Dennis Gabor Award for work on compressive holography.

References

[1] V. Backman, M. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Müller, Q. Zhang, G. Zonios, E. Kline, and T. Mcgillican, "Detection of preinvasive cancer cells," *Nature*, vol. 406, no. 6791, pp. 35–36, 2000.

[2] G. Arce, D. Brady, L. Carin, H. Arguello, and D. S. Kittle, "An introduction to compressive coded aperture spectral imaging," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[3] S. Delalieux, A. Auwerkerken, W. W. Verstraeten, B. Somers, R. Valcke, S. Lhermitte, J. Keulemans, and P. Coppin, "Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves," *Remote Sens.*, vol. 1, no. 4, pp. 858–874, 2009.

[4] R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," *J. Biomed.*, vol. 16, no. 5, pp. 056005, 2011.

[5] M. Kim, Y. Chen, and P. Mehl, "Hyperspectral reflectance and fluorescence imaging system for food quality and safety," *ASAE Trans.*, vol. 44, no. 3, pp. 721–729, 2001.

[6] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1552–1560, 2003.

[7] D. J. Brady, *Optical Imaging and Spectroscopy*. Hoboken, NJ: Wiley, 2009.

[8] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. Brady, "Coded aperture compressive temporal imaging." *Opt. Express.*, vol. 21, no. 9, pp. 10526–10545, 2013.

[9] X. Yuan, P. Llull, X. Liao, J. Yang, G. Sapiro, D. J. Brady, and L. Carin, "Low-cost compressive sensing for color video and depth," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2014, pp. 3318–3325.

[10] T. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging." *Opt. Lett.*, vol. 40, no. 17, pp. 4054–4057, 2015.

[11] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: Wide field of view multispectral imaging," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 10, pp. 1334–1348, 2002.

[12] N. Gat, "Imaging spectroscopy using tunable filters: A review," *Proc. SPIE*, vol. 4056, pp. 50–64, Apr. 2000.

[13] A. J. Villemaire, S. Fortin, J. Giroux, T. Smithson, and R. J. Oermann, "Imaging Fourier transform spectrometer," *Proc. SPIE*, vol. 2480, pp. 387, Sept. 1995.

[14] H. Morris, C. Hoyt, and P. Treado, "Imaging spectrometers for fluorescence and Raman microscopy: Acousto-optic and liquid crystal tunable filters," *Appl. Spectroscopy*, vol. 48, no. 7, pp. 857–866, 1994.

[15] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the airborne visible infrared imaging spectrometer (AVIRIS)," in *Proc. Remote Sensing Environment Conf.*, vol. 65, pp. 227–248, 1998.

[16] E. Herrala, J. T. Okkonen, T. S. Hyvarinen, M. Aikio, and J. Lammasniemi, "Imaging spectrometer for process industry applications," *Proc. SPIE*, vol. 2248, pp. 33–40, June 1994.

[17] SPECIM. (2011). Spectral Imaging LTD. [Online]. Available: http://www.specim.fi

[18] M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Appl. Opt.*, vol. 34, no. 22, pp. 4817–4826, 1995.

[19] A. Wagadarikar, R. John, R. Willett, and D. J. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[20] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[21] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 2, 2014.

[22] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," *Opt. Lett.*, vol. 39, no. 7, pp. 2044–2047, 2014.

[23] C. V. Correa, H. Arguello, and G. R. Arce. "Compressive spectral imaging with colored patterned detectors." in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing*, Florence Italy, 2014, pp. 7789–7793.

[24] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Processing*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[25] H. Rueda, D. Lau, and G. R. Arce, "Multi-spectral compressive snapshot imaging using RGB image sensors," *Opt. Express*, vol. 23, no. 9, pp. 12207–12221.

[26] H. Du, X. Tong, X. Cao, and S. Lin, A prism-based system for multispectral video acquisition," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, vol. 30, no. 2, pp. 175–182, 2009.

[27] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 33, no. 12, pp. 2423–2435, 2011.

[28] Y. August, C. Vachman, Y. Rivenson, and A. Stern, "Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains," *Appl. Opt*, vol. 52, no. 10, pp. D46–D54, 2013.

[29] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," *Opt. Lett.*, vol. 36, no. 14, pp. 2692–2694, 2011.

[30] X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," *IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 42, no. 7, pp. 297–304, 2011.

[31] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin, "Acquisition of high spatial and spectral resolution video with a hybrid camera system," *Int. J. Comput. Vis*, vol. 110, no. 2, pp. 141–155, 2014.

[32] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 4942–4950.

[33] R. M. Willett, M. F. Duarte, M. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 116–126, 2014.

[34] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.

[35] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[36] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired projection design with application to compressive sensing," *SIAM J. Image Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[37] A. Rajwade, D. Kittle, T. H. Tsai, D. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM J. Image Sci.*, vol. 6, no. 2, pp. 782–812, 2013.

[38] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using gaussian mixture models," *IEEE Trans. Image Processing*, vol. 23, no. 11, pp. 4863–4878, 2014.

[39] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[40] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 8, pp. 1509–1519, 2011.

[41] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum," Department of Computer Science, Columbia Univ., New York, Tech. Rep. CUCS-061-08, 2008.

[42] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *J. Opt. Soc. Amer. A*, vol. 23, no. 10, pp. 2359–2372, 2016.

[43] Daylight Spectra. Spectral colour Research. [Online]. Available: http://www2.uef.fi/en/spectral/daylight-spectra

[44] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[45] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ: Wiley, 2002.

[46] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing." in *Proc. IEEE Int. Conf. Image Processing*, Phoenix, AZ, 2016.

[47] C. V. Correa, H. Arguello, and G. R. Arce, "Snapshot colored compressive spectral imager," *J. Opt. Soc. Amer. A*, vol. 32, no. 10, pp. 1754–1763, 2015.

[48] X. Yuan, T. H. Tsai, R. Zhu, P. Llull, D. J. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE J. Select. Topics Signal Processing*, vol. 9, no. 6, pp. 964–976, Sept. 2015.

[49] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Processing*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[50] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 83, 2008.

[51] J. Bao and M. G. Bawendi, "A colloidal quantum dot spectrometer," *Nature*, vol. 523, no. 7558, pp. 67–70, 2015.

[52] J. Jia, K. J. Barnard, and K. Hirakawa, "Fourier spectral filter array for optimal multispectral imaging," *IEEE Trans. Image Processing*, vol. 25, no. 4, pp. 1530–1543, Apr. 2016.

[53] Y. V. White, M. Parrish, X. Li, L. M. Davis, and W. Hofmeister, "Femtosecond micro-and nano-machining of materials for microfluidic applications," in *Proc. Int. Soc. Optics and Photonics, NanoScience+ Engineering*, 2008, pp. 70390J–70390J.

[54] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform Theory*, vol. 52, no. 2, pp. 489–509, 2006.

SP

Vishal M. Patel, Joseph N. Mait,
Dennis W. Prather, and Abigail S. Hedden

Computational Millimeter Wave Imaging

Problems, progress, and prospects
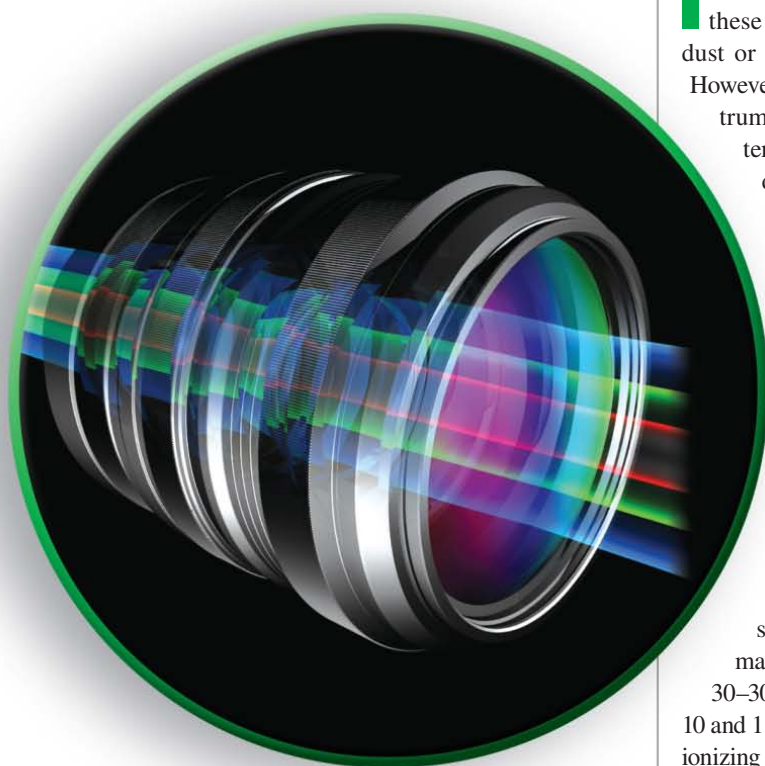
maging using millimeter waves (mmWs) has many advantages and applications in the defense, security, and aviation markets. All terrestrial bodies emit mmW radiation, and these wavelengths are able to penetrate smoke, blowing dust or sand, fog/clouds/marine layers, and even clothing. However, there are many obstacles to imaging in this spectrum that have to be overcome before mmW imaging systems can be successfully realized for surveillance and defense applications. Recent developments in computational imaging have the potential to significantly improve capabilities of mmW imaging systems. Our article provides an overview of computational imaging and its implication to mmW imaging in various operation modes. We discuss the merits and drawbacks of available computational mmW imaging approaches and identify avenues of research in this rapidly evolving field.

Introduction

In the past several years, interest in imaging at millimeter wavelengths has been driven primarily by their ability to penetrate poor weather and other obstacles such as clothes and polymers [1], [2]. Within the electromagnetic spectrum, mmWs are historically defined in the 30–300 GHz range with corresponding wavelengths between 10 and 1 mm, respectively. Radiation at these frequencies is nonionizing and is, therefore, considered safe for human exposure. Applications of this technology include the detection of concealed weapons, explosives, and contraband (see Figure 1). Furthermore, unlike visible and infrared systems, passive mmW imaging systems are not significantly hindered by atmospheric obscurants, such as cloud cover, fog, smoke, rain and dust storms and may reduce or even eliminate the impact of low-visibility atmospheric conditions [3]. Figure 2 shows atmospheric attenuation of naturally emitted black-body radiation through 1 km of fog, illustrating how low-loss bands within the mmW region allow passive imaging in adverse weather conditions.

©ISTOCKPHOTO.COM/YAKOBCHUK
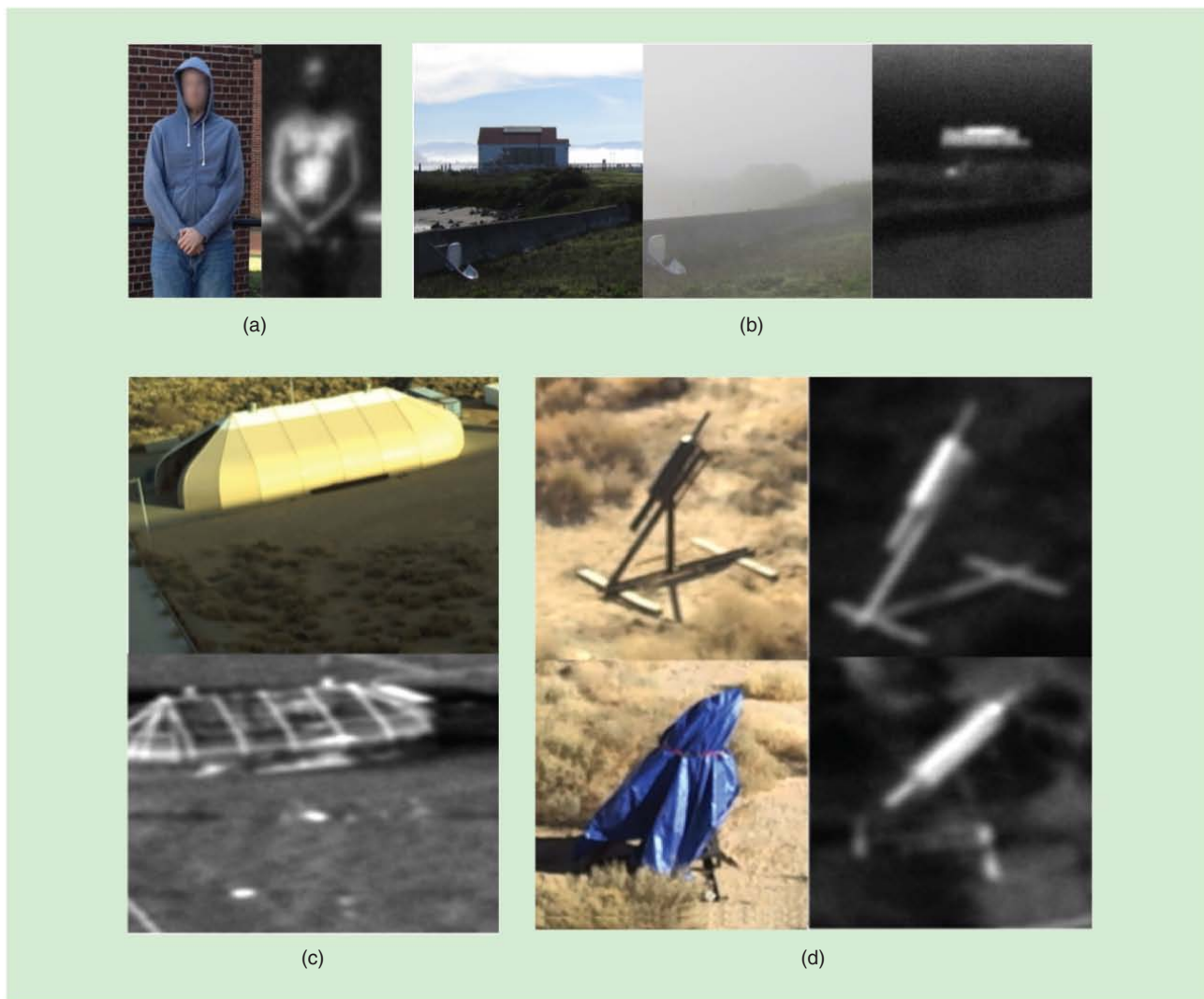
FIGURE 1. Applications of mmW imaging. (a) Imaging through clothing. (b) Imaging through fog. (c) Imaging through tarps and building materials. (d) Imaging IRAM through canvas.
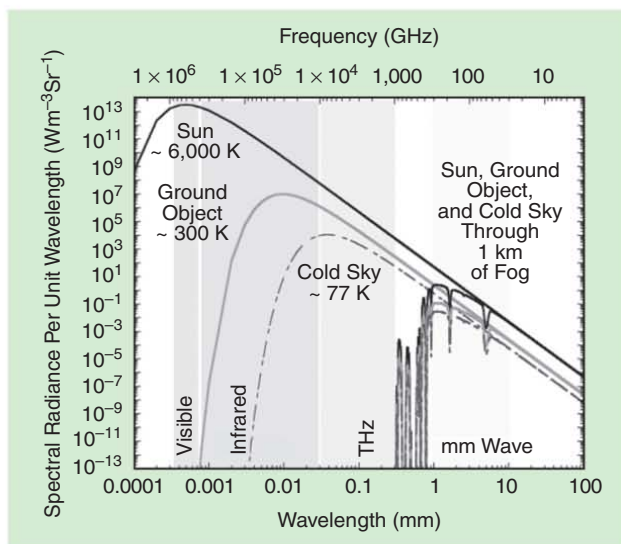
FIGURE 2. Low-loss bands within the mmW region allow passive imaging in adverse weather conditions [4].

Unfortunately, mmW imaging combines the worst of radio-frequency imaging and visible imaging. Consider that measuring phase at radio frequencies and measuring intensity at visible wavelengths are simple and inexpensive. However, measuring phase and measuring intensity are both expensive at mmWs. This is due primarily to the small signal-to-noise ratios that exist for passive mmW sources. Further, no large-scale integrated detector arrays exist for mmWs. At 100 GHz, there are only 200×200 resolution elements across an 60×60-cm^2 aperture. Man-portable mega-pixel imagers at mmWs will be difficult to realize.

In light of the fact that the magnitude of mmW measurements is considerably less than that of visible measurements and that each measurement is expensive, it would be good to increase the information content in the measurements that are made. This requires methods beyond conventional imaging and leads us naturally to consider computational imaging techniques. In computational imaging, the burden of image formation is shared across two domains, the optical
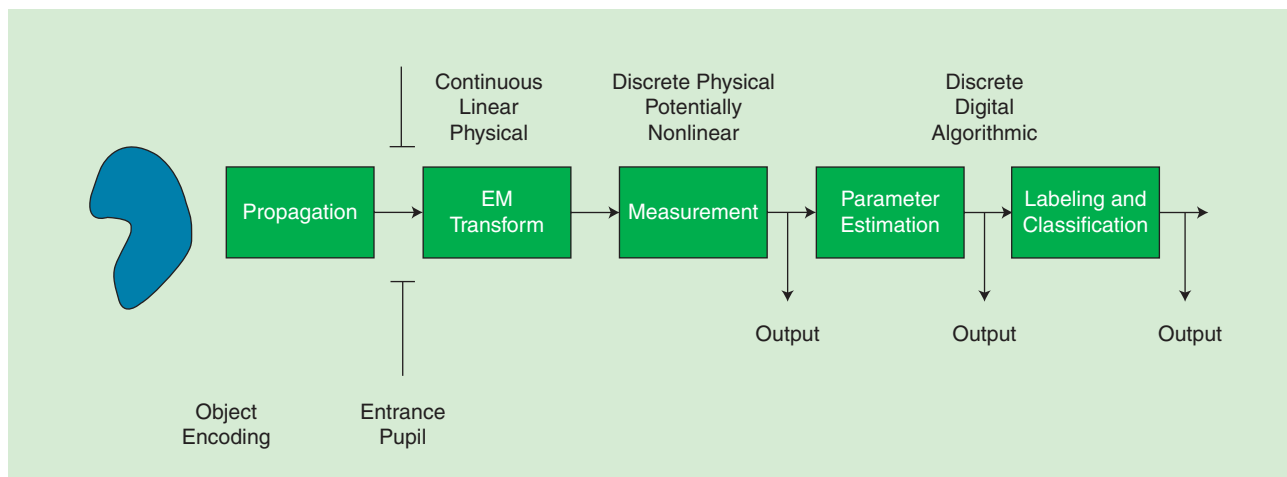
FIGURE 3. A schematic representation of an imaging system.

measurement and the digital postprocessing domains. The opportunities for computational imaging depend upon the architecture of the system [4]–[13]. In this article, we review work that we have done for conventional imagers and for pupil, or Fourier plane, imagers.

Foundations of computational imaging

Figure 3 is a schematic representation of a generic imaging system. To the left of the entrance pupil is a natural scene consisting of self-luminous objects or objects illuminated passively, i.e., we do not have active control over the scene illumination. The electromagnetic field incident upon the entrance pupil exists in three spatial dimensions (x, y, z) and one temporal dimension t and exhibits intrinsic physical properties of wavelength λ and polarization p. The amplitude of the field is represented by $a(x, y, z, t, \lambda, p)$.

All elements to the right of the entrance pupil are under a designer's control and together define the imaging system. The imager's front end contains elements that manipulate the incident wave front. The front-end electromagnetic processing is represented by linear, continuous integral transforms based on physical models.

After manipulation by the front end, the transformed wave front impinges on a detector or transducer of some kind. Transduction is a nonlinear physical process in terms of field amplitude. It is the irradiance of the field that is transduced, $f = |a|^2$. Also, discrete sampling is implicit in transduction. The continuous values (x, y, z, t, λ, p) over which f is defined are now discrete.

A matrix representation can be used to provide a mathematical description of the processes up to and including measurement [14], [15]

$$\mathbf{g} = \mathbf{Hf} + \mathbf{n}, \tag{1}$$

where $\mathbf{g}$ is the measurement, $\mathbf{f}$ is a sampled representation of the scene irradiance in the object domain, $\mathbf{H}$ is the system transfer or measurement matrix, and $\mathbf{n}$ is noise introduced in

the measurement process. The propagation of $\mathbf{f}$ from the object domain to the imager is included in $\mathbf{H}$. Thus, $\mathbf{H}$ consists of both natural and engineered components.

A parameter estimator $\mathbf{T}$ is applied in postdetection either to estimate $\mathbf{f}$ or some property or parameter of $\mathbf{f}$ denoted $\mathbf{\Omega f}$, i.e., either

$$\begin{aligned} \widehat{\mathbf{f}} &= \mathbf{Tg}, \\ &= \mathbf{THf} + \mathbf{Tn}, \end{aligned} \tag{2}$$

or

$$\begin{aligned} \mathbf{\Omega f} &= \mathbf{Tg}, \\ &= \mathbf{THf} + \mathbf{Tn}. \end{aligned} \tag{3}$$

In contrast to the measurement matrix $\mathbf{H}$, which operates linearly on $\mathbf{f}$, $\mathbf{T}$ can be linear or nonlinear.

The last block in Figure 3 represents additional processing beyond estimation, namely, labeling or classification. That is, based upon the properties estimated, elements within the scene are discriminated from one another and assigned to a particular, discrete class of objects. Depending on the application, classification may not be necessary. We combine classification processing with estimation into a single transformation $\mathbf{T}$.

It is important to note the special case of $\mathbf{T} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This is called a direct measurement, where the measurements correspond directly to the parameters of interest. For example, conventional imaging is a case of direct measurement of scene irradiance values in object space. The goal in designing a conventional imaging system is to produce a response that is as close as possible to a δ-function over all expected operating conditions.

Another classic example of direct measurement is optical matched filtering, e.g., [16]. A matched filter is designed to detect an object $\mathbf{o}$ in the object scene. Its ideal performance is such that, wherever $\mathbf{o}$ is present in the scene $\mathbf{f}$, the measurement produces a large response at its location. Locations where $\mathbf{o}$ is not present produce a small response. Thus, $\mathbf{T}$ outputs the location of $\mathbf{o}$ and, by virtue of its design, allows the classification of

points in the scene **f** into two regions, those that likely contain **o** and those that don't.

These examples represent extremes in which all processing is performed in the physical domain. The vast expanse of work in optical design speaks to the difficulty in realizing the former, and the short-lived history of optical pattern recognition in the 1960s and 1970s underscores problems with the latter. Thus, the application space of current interest is the one in which the processing burden is shared.

We delineate computational imaging into three broad applications: enhancing cameras, enhancing images (also known as computational photography), and enhancing human cognition. Reference to cameras in the first application emphasizes the conventional notion of a camera as a device that produces a recognizable representation of a scene. An enhanced camera uses computation to improve some aspect of the camera, for example, reduce its physical depth while maintaining optical performance [17], [18], increase its spatial resolution [19], or expand its dynamic range [20]. Others have considered computation as a hybrid element to reduce or overcome aberrations [21], [22].

Computation has also been used to filter or accentuate information within a scene. For example, combining unique optics with postdetection processing allows one to extend an imager's depth of field (DoF) [23]. Other examples include modulating a shutter during an exposure to reduce motion blur [24]. Computation in combination with new sensing modalities allows humans to "see" polarimetric information [25], spectral information [26], and three-dimensional information [27] in a manner similar to how they "see" through a human body using magnetic resonance.

With regard to enhancing cognition, some within the imaging community [28], [29] seek to extract information from scenes directly using physical means and postdetection processing but in a manner different from pure imaging processing, i.e., image detection followed by image processing, and from pure optical matched filtering. Such task-specific imagers require automatic feedback, dynamic elements, and adaptive processing to realize [29].

Computational mmW imaging approaches

Many mmW imaging systems have practical considerations that limit or preclude their use from surveillance and defense-related applications. In this section, we highlight several examples of computational mmW imaging methods that have been used to enhance imaging capabilities and to address some of these considerations, like size-weight-and-power (SWaP), imaging speed, and limited DoF. These are important considerations for many potential applications, like stand-off imaging and surveillance of moving targets where high angular resolution, high image frame rates, and an extended DoF are keys to mission success. Computational imagers, like the distributed aperture imaging system discussed below, have

> **The goal in designing a conventional imaging system is to produce a response that is as close as possible to a d-function over all expected operating conditions.**

also demonstrated promise in overcoming important SWaP-related issues. This is particularly important at millimeter wavelengths where high image resolution is typically achieved with large apertures and lens-based systems that scale volumetrically and can present challenges from a portability perspective.

Extended depth-of-field imaging

Most mmW imaging systems have a narrow DoF, the distance over which an object is considered in focus. Consider the application of concealed weapon detection by imaging through clothing using mmW imagers. If individuals are moving toward an imager through a corridor, the weapons will be visible only for the brief moment when they were in the DoF. This is one reason individuals are scanned in portals. However, extensions to scanning over a volume could provide scanning without creating bottlenecks, for example, in a public marketplace where security is important but a visible display of security might be counterproductive. Computational imaging methods [23], [30], [31] can be used to extend the DoF of mmW imaging systems. One such method was developed in [5] to extend the DoF of a passive mmW imaging system to allow for operation over a volume. In what follows, we review this computational imaging method for extending the DoF of a passive mmW imager.

In [5], a 94-GHz Stokes-vector radiometer was used to form images by raster scanning the system's single beam. One can model the 94-GHz imaging system as a linear, spatially incoherent, quasi-monochromatic system. The intensity of the detected image can be represented as a convolution between the intensity of the image predicted by the geometrical optics with the system point spread function (PSF) [32]. Under these conditions, (1) is a valid representation with **H** the incoherent PSF. **H** accounts for wave propagation through the aperture and is related to the magnitude square of the inverse Fourier transform of the system pupil function $P(u, v)$.

Displacement of an object from the nominal object plane of the imaging system introduces a phase error in the pupil function that increases the width of a point response and produces an out-of-focus image. For a 94-GHz imager with an aperture diameter $D = 24$ in and object distance $d_o = 180$ in, DoF ≈ 17.4 in, which ranges from 175.2 to 192.6 in (see Figure 4).

The DoF of this imager was extended using a cubic phase element in conjunction with postdetection processing. The cubic phase element $P(u, v)$ is

$$P(u,v) = \exp[\,j\theta(u,v)]\,\mathrm{rect}\!\left(\frac{u}{W_u}, \frac{v}{W_v}\right), \tag{4}$$

where

$$\theta(u,v) = (\pi\gamma)\left[\left(\frac{2u}{W_u}\right)^3 + \left(\frac{2v}{W_v}\right)^3\right]$$

and rect is the rectangular function. The phase function is separable in the u and v spatial frequencies and has spatial extent W_u and W_v along the respective axis. The constant γ represents the strength of the cubic phase. Figure 5 shows the cubic phase element mounted on the antenna.

Figure 6 shows the measured PSFs for conventional imaging and imaging with a cubic phase. The width of the in-focus PSF at 180 in is approximately 2 mm, which is consistent with a 1 mm pixel width. Note that the response of the cubic phase system is relatively unchanged, whereas the response of the conventional system changes considerably. A postdetection signal processing step is necessary to produce a well-defined sharp response [23], [30], [31].

If we assume (3) represents a linear postdetection process, we can implement **T** as a Wiener filter in Fourier space,

$$T(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + \dfrac{K^{-2}\hat{\Phi}_N(u,v)}{\hat{\Phi}_L(u,v)}}, \qquad (5)$$

where $H(u, v)$ is the optical transfer function associated with the cubic phase element, the parameter K is a measure of the signal-to-noise ratio, and the functions $\hat{\Phi}_L$ and $\hat{\Phi}_N$ are the expected power spectra of the object and noise, respectively. The optical transfer function is usually estimated from the experimentally measured point responses. One can view the estimated $i_p(x, y)$ as a diffraction limited response.

The extended object used in the experiments is represented in Figure 7(a). Images of an extended object for conventional imaging system at 113, 146, and 180 in are shown in Figure 8(a)–(c), respectively. Each image is represented by 41×51 measurements or pixels. The object size within the image is a function of optical magnification. Note that the conventional imaging system produces images with significant blurring. In contrast, even without signal processing, the images produced with cubic phase element retain more discernible characteristics of the object than the images from the conventional system, as shown in Figure 8(d)–(f). Figure 8(g)–(i) shows that postprocessing compensates for the effect of the cubic phase element and retains frequency
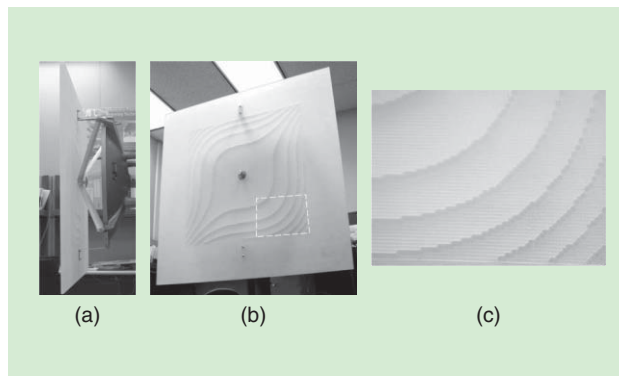
FIGURE 5. The cubic phase element. (a) Side view and (b) front view of the cubic phase element mounted on the antenna. (c) The detail of the fabricated cubic phase element.
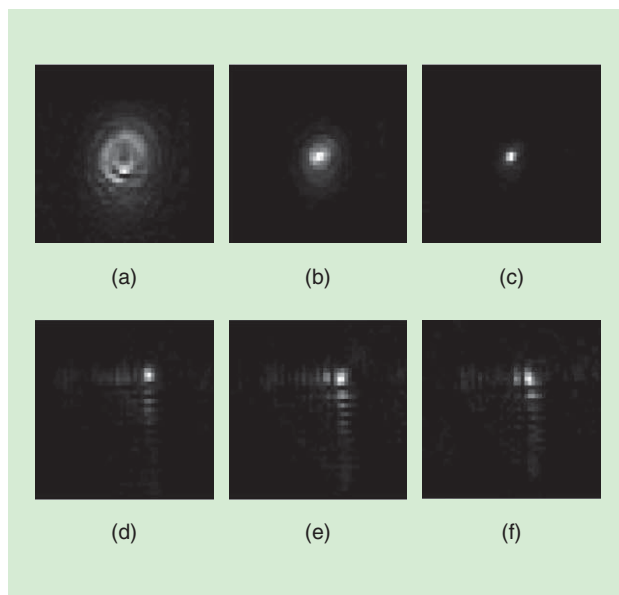
FIGURE 6. The measured PSFs for conventional imaging and imaging with a cubic phase. PSFs for conventional system at (a) 113 in, (b) 146.5 in, and (c) 180 in. (d)–(f) PSFs for a system with cubic phase at the same distances for (a)–(c).
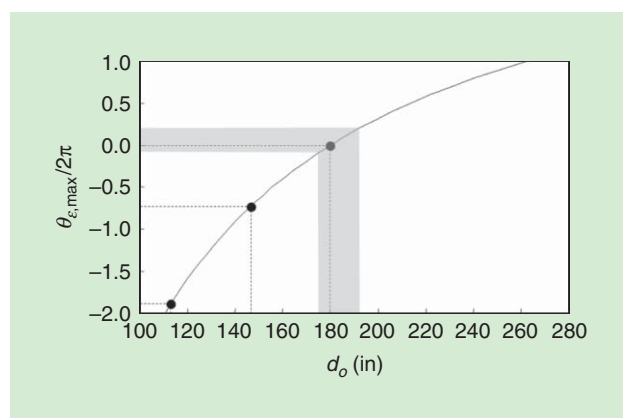
FIGURE 4. The maximum relative pupil phase error as a function of object distance. The shaded region indicates a conventional depth of field.
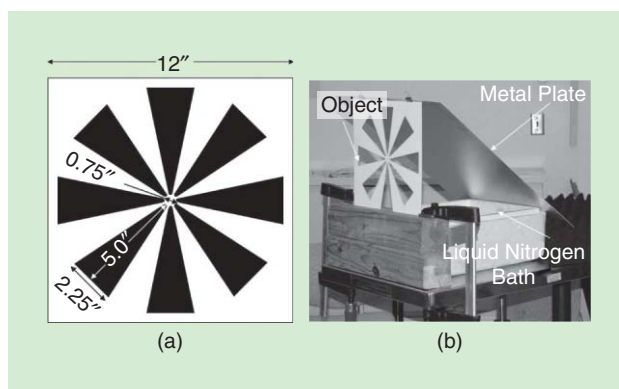
FIGURE 7. (a) A representation of the extended object used to compare conventional and cubic-phase imaging. (b) A schematic of object illumination [5].

content that is otherwise lost in a conventional system. The wider bandwidth, in addition to the noise suppressing characteristics of the Weiner filter, produces images that appear sharper than those produced by a conventional imaging system. Therefore, one can extend the region over which the system generates diffraction limited images. In fact, [5] showed that the DoF of a conventional 94-GHz imaging system can be extended from 17.4 in to more than 68 in.

Distributed-aperture mmW imaging

Recently, a pupil plane, distributed aperture mmW imager was developed by the University of Delaware and Phase Sensitive Innovations [4] shown in Figure 9. As opposed to a continuous aperture over which radiation is collected to form an image, distributed aperture systems sample the incident radiation within subapertures. This is typically done when a continuous aperture is prohibitive due to scale, e.g., for radio telescopy. This approach was particularly taken for mmWs due to the lack of detection technology, such as inexpensive silicon-based detector arrays used for detecting visible radiation. This approach offers important SWAP-related benefits compared with more traditional architectures like compound antenna systems and lens-based imagers, since the overall upconversion system size scales in two dimension (2-D) versus three-dimension (3-D).

Image formation in a distributed array requires recording both magnitude and phase of the incident field at each subaperture and cross-correlating all the recorded, complex information. Typically, systems distribute a local oscillator to downconvert the captured field data to a lower intermediate frequency, where it can be digitally recorded and processed. Although well-suited for imaging at microwave frequencies, the power, size, and space requirements for distributing the local oscillator, the intermediate frequency processing, and construction of the correlation engines present significant design challenges at mmW frequencies, which increase cost.

Martin et al. [4] showed that these challenges can be overcome by upconverting to optical frequencies and taking advantage of existing optical technology for processing and imaging. Electro-optic modulators were used to modulate received millimeter-wave radiation onto the sidebands on an optical carrier [33]. Optical upconversion allows the use of lightweight, flexible fiber optics to route optical energy before and after mmW encoding, which eliminates the need for cables to distribute a local oscillator.

Even more significant, optical upconversion allows the use of an optical lens to perform the necessary correlation required for image formation. Digital reconstruction requires discrete spatial Fourier transforms and correlations, the number of which increases quadratically with the number of subapertures. The phase transformation of a lens combined with propagation over a distance physically generates the correlations necessary for image formation.

> **Optical upconversion allows the use of an optical lens to perform the necessary correlation required for image formation.**
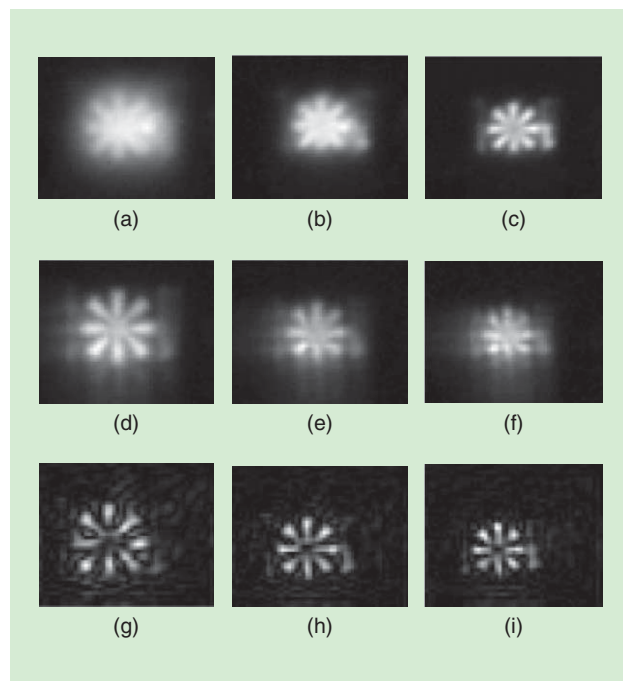
FIGURE 8. Images from a conventional imaging system at (a) 113 in, (b) 146 in, and (c) 180 in. (d)–(f) Images from a system with cubic phase at the same object distances as for (a)–(c). (g)–(f) Processed images from a system with cubic phase at the same object distances as for (a)–(c) [9].
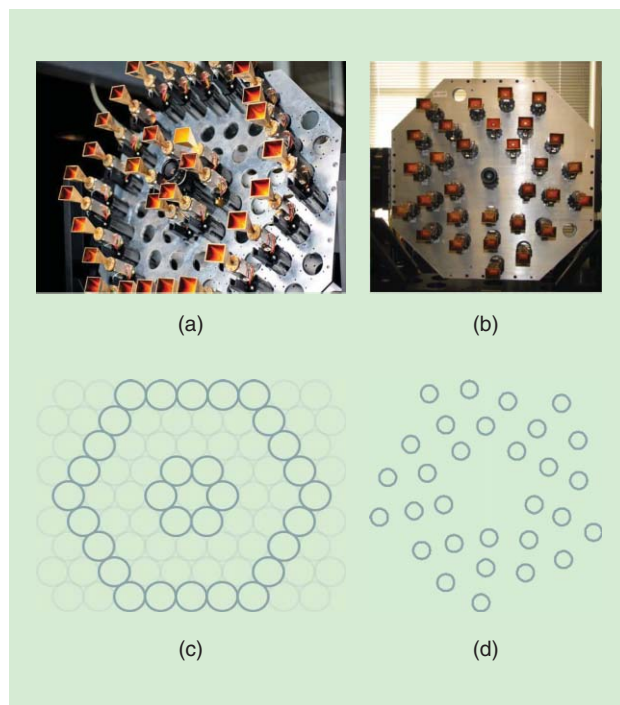
FIGURE 9. A 35-GHz, 30-channel distributed aperture imaging system with (a) hexagonal and (b) nonredundant distributed apertures and distributed aperture geometries for (c) hexagonal and (d) nonredundant apertures.

Therefore, the system proposed in [4] samples discretely the complex wave-amplitude of a mmW signal and converts the mmW signal to an optical one using electro-optical modulators while preserving the spatial distribution of samples using an optical fiber array. The output of the fiber array is spatially Fourier transformed using a lens, and the resulting optical image is captured using an optical detector array or charge-coupled device. One of the important features of this imager is that the imager volume does not scale with the aperture diameter, as the scale of the image-forming elements is fixed.

Another significant feature of this imager is its ability to control the relative phase of each receiving element in the distributed aperture, which provides electronic control of the imager's PSF. It allows multidomain sensing by simultaneous and independent manipulation of both the Fourier and image planes of the system. This is a unique capability that permits the change of the imager's PSF on-the-fly and enables rapid sparse sampling of desired target by electronically steering the beam in a manner similar to a phased array antenna. Analysis of the imager as an incoherent imaging system highlights the link between element phase and PSF [7]. Mait et al. [7] showed that, by modifying the aperture phases of the hexagonal and nonredundant distributed aperture systems, a low-resolution analog image processing can be performed. The simplest approach could be to take the difference between two images of the same object captured using two different pupil functions,

> **Compressive sensing is an important tool that has shown promise in overcoming some of the common limitations associated with mmW imaging.**

$$f(x,y) = f_+(x,y) - f_-(x,y), \tag{6}$$

where $o(x, y)$ is the input object and

$$f_+(x,y) = o(x,y) ** h_+(x,y), \tag{7}$$

$$f_-(x,y) = o(x,y) ** h_-(x,y). \tag{8}$$

For example, one can construct a one-dimensional band-pass filter by manipulating the phase functions. To understand this heuristically, one can model the corresponding PSFs as δ-functions

$$h_+(x,y) = \delta(x,y), \tag{9}$$

$$h_-(x,y) = \frac{1}{2}[\delta(x - x_0, y) + \delta(x + x_0, y)], \tag{10}$$

to approximate the composite transfer function $H(u, v)$ as

$$H(u,v) = H_+(u,v) - H_-(u,v) = 1 - \cos(2\pi x_0 u), \tag{11}$$

which filters low spatial frequencies and passes frequencies centered at $u = 1/x_0$. Such filtering can be useful for edge detection (see Figure 10). Furthermore, [34] showed that phase can also be used to do more complex signal processing

such as reducing noise. Figure 11 illustrates the real-time nature of the imager where a person behind a plywood is imaged at a video rate.

Compressive mmW imagers

Compressive sensing (CS) is an important tool that has shown promise in overcoming some of the common limitations associated with mmW imaging. For example, a wide field of regard and high image frame rates are desired for many applications, like stand-off imaging of moving targets. Because large-format arrays present cost challenges due to technological hurdles like availability of cost-effective and powerful source technology and sensitive, low-cost detectors, many current systems use a single element or a small array in combination with a compound antenna system to scan a larger scene and build up an image (e.g., [35]–[38]). These architectures can present challenges to achieving the video frame rates required for some imaging applications [38]. As highlighted in this section, mmW CS techniques have shown promise in overcoming issues like high frame-rate challenges by potentially reducing the overall number of scene measurements needed to reconstruct an image.

CS can be viewed as a special case of computational imaging in which partial or low-dimensional measurements are obtained by designing a specific sensing modality. In this case, the measurement matrix $\mathbf{H}$ in (1) has more columns than rows, and a nonlinear recovery algorithm is used to reconstruct the
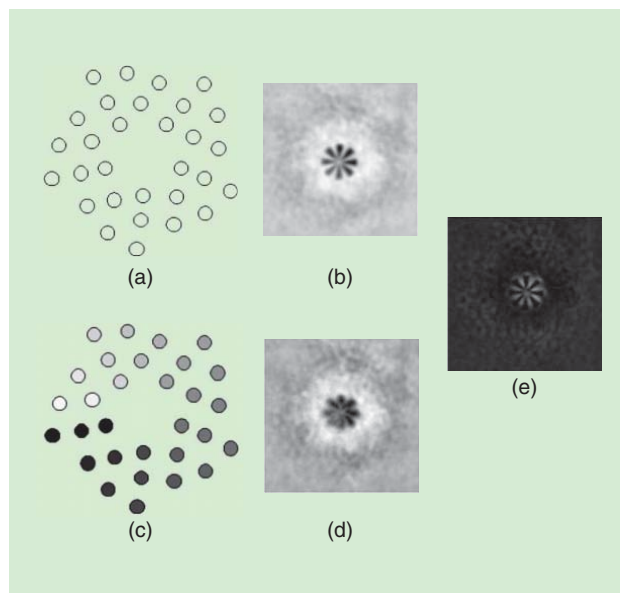
FIGURE 10. Edge detection performed using a nonredundant distributed aperture. (a) Aperture phase and (b) corresponding image of an extended object assuming zero aperture phase. (c) and (d) Same as in (a) and (b) except with circular phase across the aperture. White represents 0-phase, and black represents 2π-phase. Intermediate grey colors represent phases between 0 and 2π. (e) The difference between (b) and (d) [7].
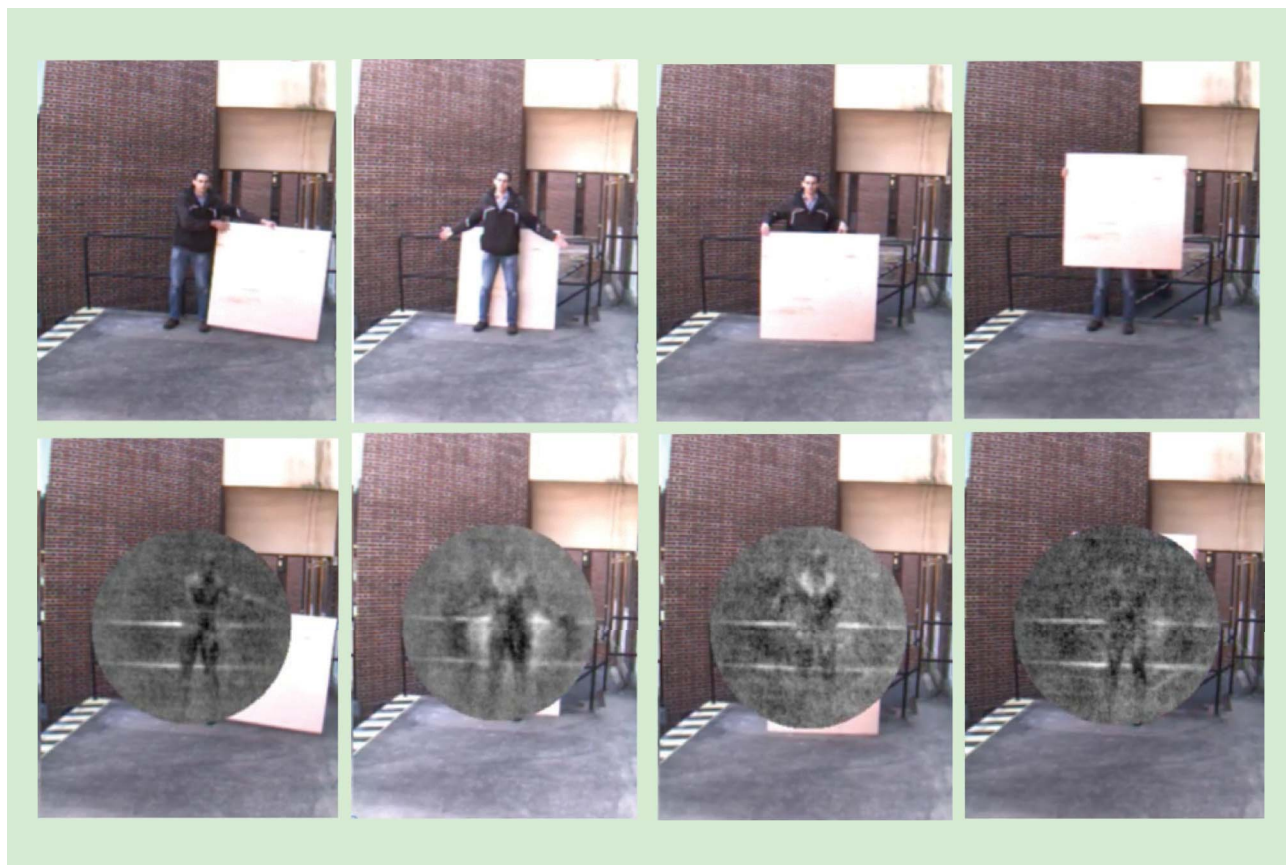
FIGURE 11. Real-time video snapshots seeing through 1/4-in plywood.

scene [39], [40]. A number of CS methods have been developed to reduce the acquisition time of mmW imagers [8]–[11], [41]. For instance, [10] proposed a compressive passive mmW imaging method in which randomly encoded masks are employed at the focal plane of the imager to acquire incoherent measurements of the imaged scene. A Bayesian reconstruction algorithm was developed to estimate the original image from these compressive measurements. This system shows can significantly reduce the number of required measurements for passive mmW imaging. This method was later extended in [42] by constructing a single unified and compact mask such that no mechanical mask exchange is necessary for collecting compressive measurements.

Another method based on CS for terahertz (THz) imaging was proposed in [11]. This method uses a single pixel detector in combination with a series of random masks to enable high-speed image acquisition. This system showed to be capable of producing 32 × 32 images of complex objects with only 300 (approximately 30%) measurements. Rather than using random masks, [43] proposes using Toeplitz matrix-based masks. This method has the advantage of a large number of masks, which can be represented by a single sensing mask. The image acquisition time of this system showed to be only limited to the speed of the THz detector.

A mmW imaging modality with extended DoF with reduced spatial sampling was developed in [8] and [9]. This method essentially uses a cubic phase element at the pupil of the imager while collecting partial measurements. The image is then recovered by using a nonlinear reconstruction algorithm. Using this system, a greater than four-fold increase in DoF can be achieved with a reduction in sampling requirements by a factor of at least two.

In a recent work [44], active matamaterials were introduced as real-time tunable, spectrally sensitive spatial masks for single pixel THz imaging. This method requires no moving parts and can yield improved signal-to-noise ratios over standard raster-scanning techniques for THz imaging. Furthermore, it was demonstrated that the use of this technique in the CS framework can allow one to acquire high-frame-rate and high-fidelity images.

> We expect that derivation of the performance bounds for various computational mmW imaging methods will produce stronger guidance to developing more advanced mmW imaging modalities, which will have a wider spectrum of applications in surveillance, defense, and aviation problems.

Discussion and concluding remarks

This article presented a review of recent developments in mmW imaging based on computational imaging methods for security and surveillance applications. We believe that recent advances in computational imaging have brought substantial opportunities to mmW imaging. We hope that the survey helped guide the interested reader through the extensive literature. It does not cover all the literature on mmW and computational imaging, so we have chosen to focus on a subset of work that reflects some of the most recent progress.

A number of challenges and issues commonly confront mmW imaging technology. Computational imaging methods may prove useful in addressing some of these challenges. The following are several examples:

- *Affordability*. The technology readiness level of mmW devices is immature compared to optical and infrared arrays. The lack of readily available and affordable sources and detection technology has resulted in comparatively small arrays (kilopixels or fewer) and a tradeoff between the number of achievable image pixels and the desire to rapidly image wide fields of regard with high angular resolution. MmW compressed sensing has shown promising results in reducing the overall number of scene observations needed to reconstruct an image. Perhaps these techniques or other computational imaging methods could help curb the cost of mmW systems by requiring fewer detector elements to realize an imaging capability that is more comparable to what could be achieved with a larger-format array.

- *SWaP*. Many mmW imaging systems are not viable for deployment across a broad variety of platforms that would benefit from their use. Compound antenna systems and lens-based imagers, for example, scale volumetrically. To achieve a high resolution and a wide field of view, larger apertures and mechanical scanners, which have important implications for SWaP, are usually used. These solutions do not tend to be man-portable, for example. Additionally, for broad applicability, one also wants platform-agnostic solutions that do not require specific aspects of the platform to form images, like platform motion, for example. Computational imagers may offer some key advantages, like the distributed aperture mmW imaging technology discussed in the "Computational mmW imaging approaches," which scales in 2-D versus 3-D, for example.

- *Surveillance of moving targets*. Imaging of moving targets with high resolution and high frame rates can be challenging with existing systems. At lower frequency, SAR offers excellent atmospheric penetration properties but relatively slow frame rates. MmW imagers can be limited by the speed of mechanical scanners, and electronic beam-scanning technology is immature and costly at millimeter wavelengths. Given challenges like these, perhaps computational imaging techniques could be applied to help compensate for image blur with existing systems.

Computational mmW imaging promises to be an active area of research. However, little is known about the quantitative performance advantage of computational imaging methods for mmW imaging. We expect that derivation of the performance bounds for various computational mmW imaging methods will produce stronger guidance to developing more advanced mmW imaging modalities, which will have a wider spectrum of applications in surveillance, defense, and aviation problems.

Authors

Vishal M. Patel (vishal.m.patel@rutgers.edu) received his B.S. degrees in electrical engineering and applied mathematics (Hons.); his M.S. degree in applied mathematics from North Carolina State University, Raleigh, in 2004 and 2005, respectively; and his Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2010. He is currently an assistant professor in the Department of Electrical and Computer Engineering at Rutgers University. Prior to joining Rutgers, he was a member of the research faculty with the University of Maryland Institute for Advanced Computer Studies, College Park. His current research interests include signal processing, computer vision, and pattern recognition with applications in biometrics and imaging. He is a recipient of the 2016 Office of Naval Research Young Investigator Award and the 2010 Oak Ridge Associated Universities Post-doctoral Fellowship. He is a member of Eta Kappa Nu, Pi Mu Epsilon, and Phi Beta Kappa.

Joseph N. Mait (joseph.n.mait2.civ@mail.mil) received his Ph.D. degree from the Georgia Institute of Technology in 1985. He is the chief scientist of the U.S. Army Research Laboratory. He is a fellow of the Society of Photo-Instrumentation Engineers and the Optical Society of America (OSA) and a Senior Member of the IEEE. He is the immediate past editor-in-chief of OSA's *Applied Optics*. In 2014, he was awarded a Presidential Rank Award for Meritorious Senior Professionals. His research interests include sensors and the application of optics, photonics, and electromagnetics to sensing and sensor signal processing.

Dennis W. Prather (dprather@udel.edu) received his Ph.D. degree from the University of Maryland in 1997. He is an Endowed Professor of Electrical Engineering at the University of Delaware. He is a Senior Member of the IEEE, fellow of the Society of Photo-Instrumentation Engineers, and a fellow of the Optical Society of America. His research focuses on both the theoretical and experimental aspects of RF-photonic elements and their integration into various systems for imaging, communications and radar. He has authored or coauthored more than 400 scientific papers, holds more than 40 patents, and has written ten books/book-chapters.

Abigail S. Hedden (abigail.s.hedden.civ@mail.mil) received her Ph.D. degree in physics from the University of Arizona in 2007. She is a physicist in the RF Technology and Integration Branch of the Sensors and Electron Devices Directorate at the U.S. Army Research Laboratory in Adelphi, Maryland. Her current research interests include development of millimeter-wave instrumentation and radar systems, phenomenology, and experimentation.

References

[1] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, vol. 95, no. 8, pp. 1683–1690, Aug. 2007.

[2] L. Yujiri, M. Shoucri, and P. Moffa, "Passive millimeter wave imaging," *IEEE Microw. Mag.*, vol. 4, no. 3, pp. 39–50, Sept. 2003.

[3] J. P. Wilson, D. G. Mackrides, J. P. Samluk, and D. W. Prather, "Comparison of diurnal contrast changes for millimeter-wave and infrared imagery," *Appl. Optics*, vol. 49, no. 19, pp. E31–E37, July 2010.

[4] R. Martin, C. A. Schuetz, T. E. Dillon, C. Chen, J. Samluk, E. L. Stein, M. Mirotznik, and D. W. Prather, "Design and performance of a distributed aperture millimeter-wave imaging system using optical upconversion," in *Proc. SPIE: Passive Millimeter-Wave Imaging Technology XII*, 2009, Orlando, FL, pp. 7309.

[5] J. N. Mait, D. A. Wikner, M. S. Mirotznik, J. van der Gracht, G. P. Behrmann, B. L. Good, and S. A. Mathews, "94-ghz imager with extended depth of field," *IEEE Trans. Antennas Propagat.*, vol. 57, no. 6, pp. 1713–1719, June 2009.

[6] A. S. Hedden, C. R. Dietlein, and D. A. Wikner, "Design of 220-Ghz electronically scanned reflectarrays for confocal imaging systems," *Opt. Eng.*, vol. 51, no. 9, pp. 091 611–1–091 611–10, Sept. 2012.

[7] J. N. Mait, R. D. Martin, C. A. Schuetz, and D. W. Prather, "Millimeter wave imaging with engineered point spread functions," *Opt. Eng.*, vol. 51, no. 9, pp. 091 606–1–091 606–9, Sept. 2012.

[8] V. M. Patel and J. N. Mait, "Passive millimeter-wave imaging with extended depth of field and sparse data," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2012, Kyoto, Japan, pp. 2521–2524.

[9] V. M. Patel and J. N. Mait, "Compressive passive millimeter-wave imaging with extended depth of field," *Opt. Eng.*, vol. 51, no. 9, pp. 091 610–1–091 610–7, Sept. 2012.

[10] S. D. Babacan, M. Luessi, L. Spinoulas, A. K. Katsaggelos, N. Gopalsami, T. Elmer, R. Ahern, S. Liao, and A. Raptis, "Compressive passive millimeter-wave imaging," in *Proc. IEEE Int. Conf. Image Processing*, 2011, Brussels, Belgium, pp. 2705–2708.

[11] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel Terahertz imaging system based on compressed sensing," *Appl. Phys. Lett.*, vol. 93, no. 12, pp. 121105, 2008.

[12] J. P. Wilson, C. A. Schuetz, T. E. Dillon, P. Yao, C. E. Harrity, and D. W. Prather, "Passive 77 Ghz millimeter-wave sensor based on optical upconversion," *Appl. Optics*, vol. 51, no. 18, pp. 4157–4167, June 2012.

[13] J. Hunt, J. Gollub, T. Driscoll, G. Lipworth, A. Mrozack, M. S. Reynolds, D. J. Brady, and D. R. Smith, "Metamaterial microwave holographic imaging system," *J. Optic. Soc. Am. A*, vol. 31, no. 10, pp. 2109–2119, Oct. 2014.

[14] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Bristol, UK: Institute of Physics Pub. Cop., 1998.

[15] O. Cossairt, M. Gupta, and S. K. Nayar, "When does computational imaging improve performance?," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 447–458, Feb. 2013.

[16] A. V. Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inform. Theory*, vol. 10, no. 2, pp. 139–145, Apr. 1964.

[17] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by bound optics (tombo): Concept and experimental verification," *Appl. Opt*, vol. 40, no. 11, pp. 1806–1813, May 2001.

[18] M. Shankar, R. Willett, N. Pitsianis, T. Schulz, R. Gibbons, R. T. Kolste, J. Carriere, C. Chen, D. Prather, and D. Brady, "Thin infrared imaging systems through multichannel sampling," *Appl. Opt*, vol. 47, no. 10, pp. B1–B10, Apr. 2008.

[19] S. S. Young and R. G. Driggers, "Superresolution image reconstruction from a sequence of aliased imagery," *Appl. Opt*, vol. 45, no. 21, pp. 5073–5085, July 2006.

[20] M. Mirotznik, S. Mathews, R. Plemmons, P. Pauca, T. Torgersen, R. Barnard, B. Gray, T. Guy, Q. Zhang, J. van der Gracht, C. Petersen, M. Bodnar, and S. Prasad, "A practical enhanced-resolution integrated optical-digital imaging camera (periodic)," in *Proc. SPIE, Modeling and Simulation for Military Operations IV*, 2009, vol. 7348, Orlando, FL, p. 734806.

[21] E. R. Dowski and K. S. Kubala, "Reducing size, weight, and cost in a lwir imaging system with wavefront coding," in *Proc. SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XV*, 2004, vol. 5407, pp. 66–73.

[22] D. G. Stork and M. D. Robinson, "Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems," *Appl. Opt.*, vol. 47, no. 10, pp. B64–B75, 2008.

[23] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.*, vol. 34, no. 11, pp. 1859–1866, Apr. 1995.

[24] A. Agrawal, Y. Xu, and R. Raskar, "Invertible motion blur in video," *ACM Trans. Graph.,* vol. 28, no. 3, pp. 95, 2009.

[25] J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, "Review of passive imaging polarimetry for remote sensing applications," *Appl. Opt.*, vol. 45, no. 22, pp. 5453–5469, Aug. 2006.

[26] M. Gehm, R. John, D. J. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no., 21, pp. 14013–14027, Oct. 2007.

[27] T. C. Poon, B. Lee, H. Yoshikawa, and J. Rosen, "Digital holography and 3-D imaging: Feature introduction," *Appl. Opt*, vol. 48, no. 34, pp. DH2, 2009.

[28] M. A. Neifeld and P. Shankar, "Feature-specific imaging," *Appl. Opt*, vol. 42, no. 17, pp. 3379–3389, June 2003.

[29] M. P. Christensen, V. Bhakta, D. Rajan, T. Mirani, S. C. Douglas, S. L. Wood, and M. W. Haney, "Adaptive flat multiresolution multiplexed computational imaging architecture utilizing micromirror arrays to steer subimager fields of view," *Appl. Opt.*, vol. 45, no. 13, pp. 2884–2892, May 2006.

[30] W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," *Appl. Opt.*, vol. 41, no. 29, pp. 6080–6092, Oct. 2002.

[31] S. Bradburn, W. T. Cathey, and E. R. Dowski, "Realizations of focus invariance in optical–digital systems with wave-front coding," *Appl. Opt.*, vol. 36, no. 35, pp. 9157–9166, Dec. 1997.

[32] J. W. Goodman, *Introduction to Fourier Optics*. Englewood, CO: Roberts and Company, 2005.

[33] C. A. Schuetz and D. W. Prather, "Optical upconversion techniques for high-sensitivity millimeter-wave detection," in *Proc. SPIE: Passive Millimetre-Wave and Terahertz Imaging and Technology*, 2004, pp. 166–174.

[34] J. N. Mait, R. D. Martin, C. A. Schuetz, S. Shi, D. W. Prather, P. F. Curt, and J. Bonnett, "Minimum bias design for a distributed aperture millimeter wave imager," in *Proc. IEEE Global Conf. Signal and Information Processing*, Austin, TX, 2013, pp. 711–714.

[35] K. B. Cooper, "Imaging, doppler, and spectroscopic radars from 95 to 700 Ghz," in *Proc. SPIE: Passive and Active Millimeter-Wave Imaging XIX*, 2016, vol. 9830, Baltimore, MD, pp. 983005.

[36] E. Gandini and N. Llombart, "Toward a real time stand-off submillimeter-wave imaging system with large field of view: quasi-optical system design considerations," in *Proc. SPIE: Passive and Active Millimeter-Wave Imaging XVII*, vol. 9462, pp. 946205, Baltimore, MD, 2015.

[37] K. B. Cooper, R. J. Dengler, N. Llombart, B. Thomas, G. Chattopadhyay, and P. H. Siegel, "Thz imaging radar for standoff personnel screening," *IEEE Trans. Thz Sci. Technol.*, vol. 1, no. 1, pp. 169–182, Sept. 2011.

[38] N. Llombart, R. J. Dengler, and K. B. Cooper, "Terahertz antenna system for a near-video-rate radar imager [antenna applications]," *IEEE Antennas Propagat. Mag.*, vol. 52, no. 5, pp. 251–259, Oct. 2010.

[39] Y. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. [Online]. Cambridge, UK: Cambridge Univ. Press, 2012.

[40] V. M. Patel and R. Chellappa, *Sparse Representations and Compressive Sensing for Imaging and Vision*. New York: SpringerBriefs, 2013.

[41] C. Fernandez-Cull, D. A. Wikner, J. N. Mait, M. Mattheiss, and D. J. Brady, "Millimeter-wave compressive holography," *Appl. Opt.*, vol. 49, no. 19, pp. E67–E82, July 2010.

[42] L. Spinoulas, J. Qi, A. K. Katsaggelos, T. W. Elmer, N. Gopalsami, and A. C. Raptis, "Optimized compressive sampling for passive millimeter-wave imaging," *Appl. Opt.*, vol. 51, no. 26, pp. 6335–6342, Sept. 2012.

[43] A. Heidari and D. Saeedkia, "A 2D camera design with a single-pixel detector," in *Proc. Int. Conf. Infrared, Millimeter, and Terahertz Waves*, 2009, Busan, South Korea, pp. 1–2.

[44] C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla, "Terahertz compressive imaging with metamaterial spatial light modulators," *Nat. Photon.*, vol. 8, no. 8, pp. 605–609, June 2014.

SP

Gordon Wetzstein and Douglas Lanman

Factored Displays

*Improving resolution, dynamic range, color reproduction,
and light field characteristics with advanced signal processing*
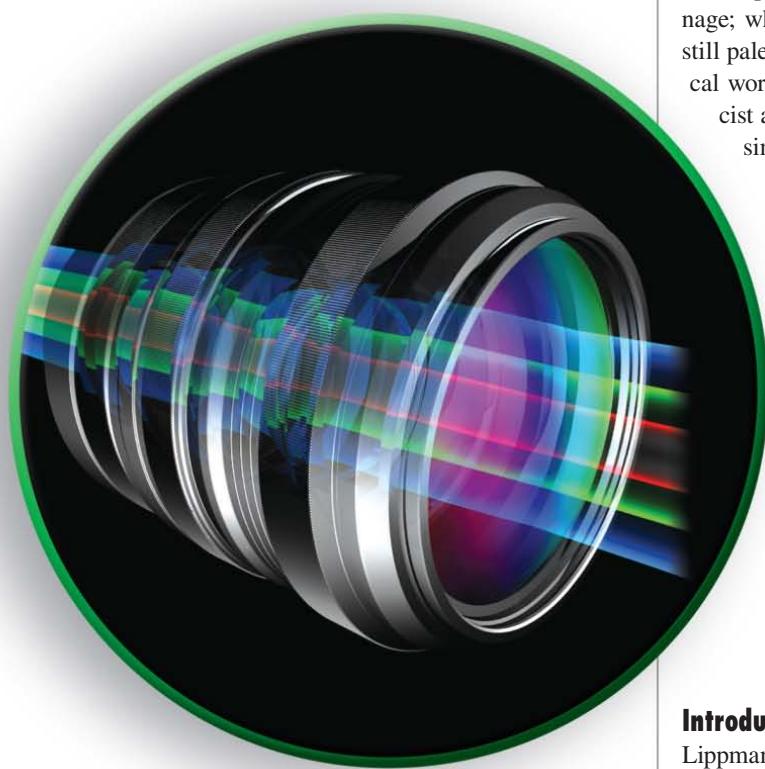
oday, direct-view displays constitute the primary means by which humans visually interact with computers. These displays come in many forms: mobile phones, laptops, workstations, digital projectors, and digital signage; while ubiquitous, our two-dimensional (2-D) displays still pale in comparison to the three-dimensional (3-D) physical world. Gabriel Lippmann, a Nobel Prize-winning physicist and an early photographic innovator, lamented on the similar limitations of photography:

> Can we request that Photography renders the full variety offered by the direct observation of objects? Is it possible to create a photographic print in such a manner that it represents the exterior world framed, in appearance, between the boundaries of the print, as if those boundaries were that of a window opened on reality? It appears that yes, we can request from Photography infinitely more than from the human hand.
>
> —Gabriel Lippmann,
> *Épreuves réversibles donnant la
> sensation du relief*, 1908

©ISTOCKPHOTO.COM/YAKOBCHUK

Introduction

Lippmann's goal of creating a faultless window into a recorded reality is the visual equivalent of Alan Turing's eponymous imitation game. To this end, Lippmann himself introduced the notion of integral imaging: a photographic process by which parallax views are recorded and depicted using microlens arrays placed in close proximity to film. Today, Lippmann's invention continues to be one of the predominant technologies underlying glasses-free 3-D displays. However, displays must advance on many other axes to achieve Lippmann's window, particularly in terms of resolution, dynamic range, and color fidelity.

Is this a case where we simply must wait for the incremental progression of display technology to carry us to the conclusion of the visual Turing test? Mobile phones certainly make a case that such Moore's law progression will occur, with mobile display pixel pitches routinely satisfying 20/20 visual acuity at typical viewing distances. However, emerging display technologies, particularly glasses-free 3-D displays and head-mounted displays (HMDs), will require at least an order of magnitude finer pixel pitches to satisfy human visual acuity. Looking beyond this immediate issue of resolution, the dynamic range of current-generation displays are far from the 1,000,000:1 contrast ratio resolved by the human eye.

Is there another path to constructing Lippmann's window without simply waiting for enough brute force to be applied? We believe so. In this article we review recent progress in factored displays comprising multiple spatial light modulators (SLMs), such as liquid crystal displays (LCDs), digital micromirror devices (DMDs), and liquid crystal on silicon (LCoS) panels. Unlike prior tiled display architectures, factored displays utilize multiple SLMs in a cascaded configuration, such that light from one modulator illuminates the next in the series. Existing multiview displays, including the parallax barrier light field display introduced by Frederic Ives in 1903 [1], can be viewed as an early precursor to factored displays. However, rather than using heuristic procedures to decompose target imagery into multilayer attenuation patterns, emerging factored displays apply principled optimization algorithms to well-defined visual objective functions. As will be emphasized throughout this article, such optimization typically reduces to solving a nonnegative matrix factorization (NMF) or nonnegative tensor factorization (NTF) problem. This unifying problem arises from the cascaded nature of the construction of factored displays: light transmitted through one pixel on one SLM layer is attenuated (i.e., multiplied) by the amplitude of a pixel on a subsequent layer. Additionally, temporal multiplexing can be applied when the pixel states can be altered at a more rapid refresh rate than the critical flicker fusion threshold of the human visual system. A unified image formation model for all factored displays is therefore given by the following equation:

$$\underset{\{f_m^{(n)}\}}{\text{minimize}} \left\| l(x,y,u,v) - \frac{1}{M}\sum_{m=1}^{M}\prod_{n=1}^{N} f_m^{(n)}\left(\phi^{(n)}(x,y,u,v)\right) \right\|_F^2,$$
$$\text{subject to} \quad 0 \le f_m^{(n)} \le 1 \tag{1}$$

where M is number of time-multiplexed images that the visual system perceptually averages, N is the number of attenuating display layers, $l(x,y,u,v)$ is the target light field—a generic representation that encodes any 2-D, 3-D, or four-dimensional (4-D) image content—and $\phi^{(n)}:\mathbb{R}^4\to\mathbb{R}^2$ is a mapping function from the 4-D light field ray space to the 2-D pixel coordinate on SLM layer n. The task of the factorization routine is then to factor $l(x,y,u,v)$ into the best set of temporally varying, nonnegative pixel states $f_m^{(n)}$ that can be addressed by the electro-optical display. Figure 1 illustrates this image formation model and shows a prototype display for a three-layer LCD display.

Superresolution displays

The development of high-resolution displays is of central importance to the display industry. Leading mobile displays recently transitioned from pixel densities of less than 50 pixels/cm (ppcm) and now approach 150 ppcm. Similarly, the consumer electronics industry now offers "4K ultra-high definition (UHD)" displays, having a horizontal resolution approaching 4,000 pixels, as the successor to high-definition television (HDTV). Furthermore, 8K UHD standards already exist for enhanced digital cinema. Achieving such high-resolution displays currently hinges on advances that enable spatial light modulators with increased pixel counts.

Beyond these larger market trends, several emerging display technologies necessitate even greater resolutions than 4K/8K UHD standards will provide. For example, wide-field-of-view HMDs, such as the Oculus Rift, incorporate high-pixel-density mobile displays. Such displays already approach or exceed 20/20 visual acuity when viewed at the distance of a phone or tablet computer; however, they appear pixelated when viewed through magnifying HMD optics, which dramatically expand the field of view. Similarly, glasses-free 3-D displays, including parallax barrier [1] and integral imaging [2] designs, require at least an order of magnitude higher resolution than today's displays.

Factored displays may have a role in addressing the demand for such high resolution displays. Rather than directly fabricating finer pixel pitches, factored multilayer LCDs have been shown to reliably quadruple spatial resolution—effectively repurposing multilayer light field displays for the distinct task of superresolution display. In recent work, Heide et al. [3] introduced cascaded displays: stacking two or more spatial light modulators (SLMs) on top of one another, without spacing, subject to a lateral offset of half a pixel or less along each axis. As shown in Figure 2, lateral offsets are necessary so that each pixel on one layer modulates multiple pixels on another; in this manner, the intensity of each subpixel fragment—defined by the geometric intersection of a pixel on one display layer with one on another layer—can be controlled, thereby increasing the effective display resolution. The target light field in this case is a view-independent 2-D image, i.e., $l(x,y,u,v) = i(x,y)$. The key insight is that cascaded displays may operate as factored displays: utilizing fewer independently addressable pixels than apparent in the displayed image. This outcome is achieved by decomposing target imagery into multilayer, multiframe attenuation patterns (see Figure 1). Similar methods may be adopted to increase the temporal resolution of stacks of two or more SLMs, refreshed in staggered intervals. As such, factored displays with offset pixel layers may be applied for spatiotemporal superresolution, in addition to their more common use as light field displays.

Cascaded displays achieve thin form factors without moving parts. Most significantly, such displays offer an operation mode that eliminates the need for temporal multiplexing of factorized imagery, which is typical of most factored light field displays; as a result, videos can be presented without the appearance of artifacts characteristic of prior methods or the requirement for
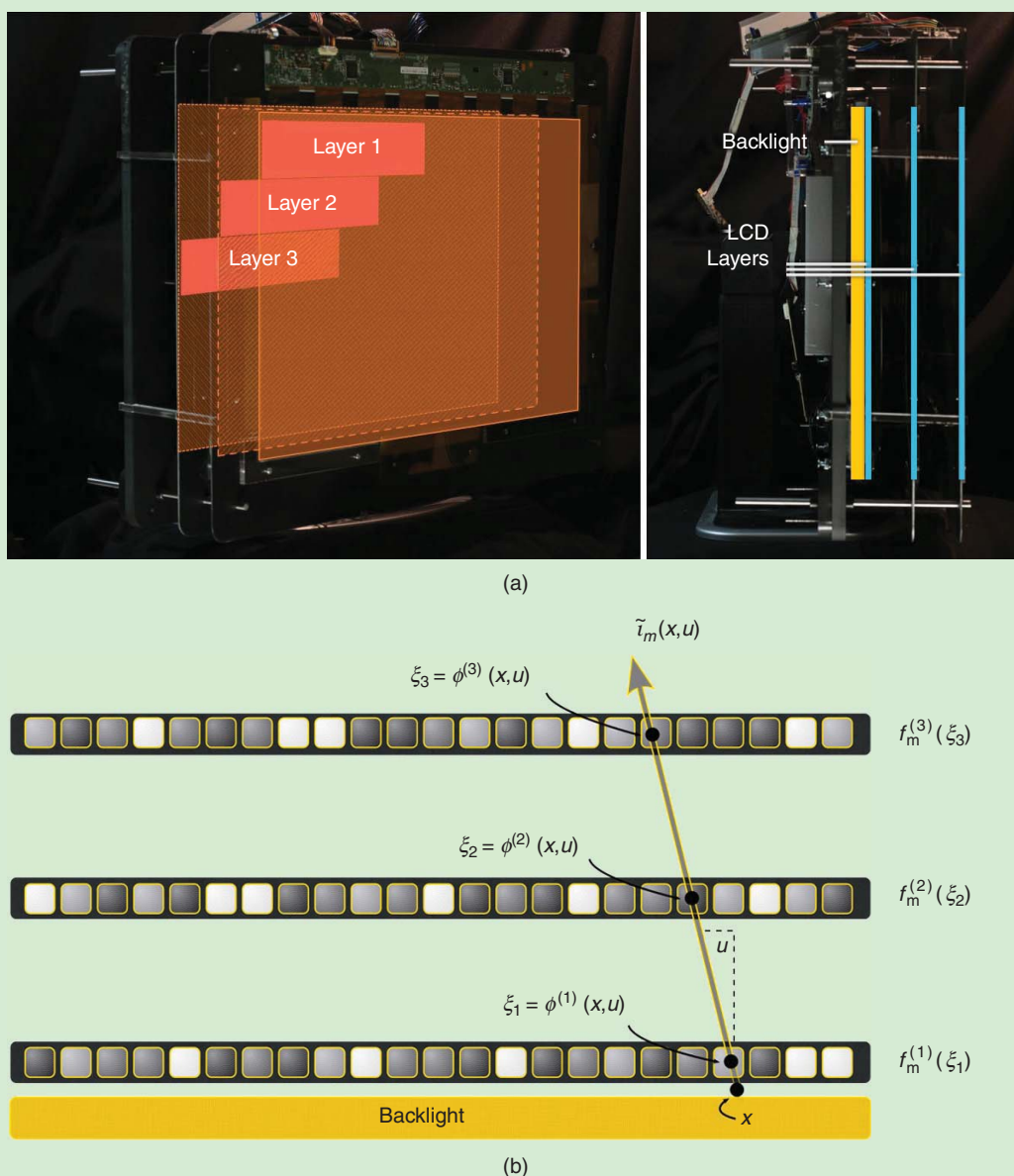
FIGURE 1. (a) Factored light field prototype and (b) image formation. The prototype uses three stacked layers of LCDs that are rear-illuminated by a single backlight. In this particular case, there are separations between the LCDs to allow for light field synthesis, but the same prototype with the inter-layer spacing reduced allows for superresolution image generation. Each emitted light ray $\tilde{l}(x, y, u, v)$ (here illustrated in 1-D) at some time m is defined as the product of all pixel states along its optical path.

high-refresh-rate SLMs. As will be discussed in the section "Light Field Displays," related methods for achieving superresolution displays have also been demonstrated in factored projectors by [4].

Factored spectral displays

Factored displays have the potential to rapidly accelerate resolution enhancement using unconventional, stacked display setups driven by factored image synthesis. It is widely anticipated, however, that a higher resolution alone will not significantly enhance user experiences unless the dynamic range (contrast) and color gamut of the displays are also improved. High-dynamic-range (HDR) display can be achieved via dual modulation [5] and improved with factored displays (see the section "High-Dynamic-Range Displays"). One of the most significant remaining challenges in display design is that of improving the color gamut while maintaining high light throughput and a sufficient bit depth to encode high-quality images. Whereas conventional display design strategies determine the tradeoffs between color fidelity, brightness, resolution, bit depth, and other characteristics of a device before it is fabricated, factored spectral displays advocate for a fundamentally
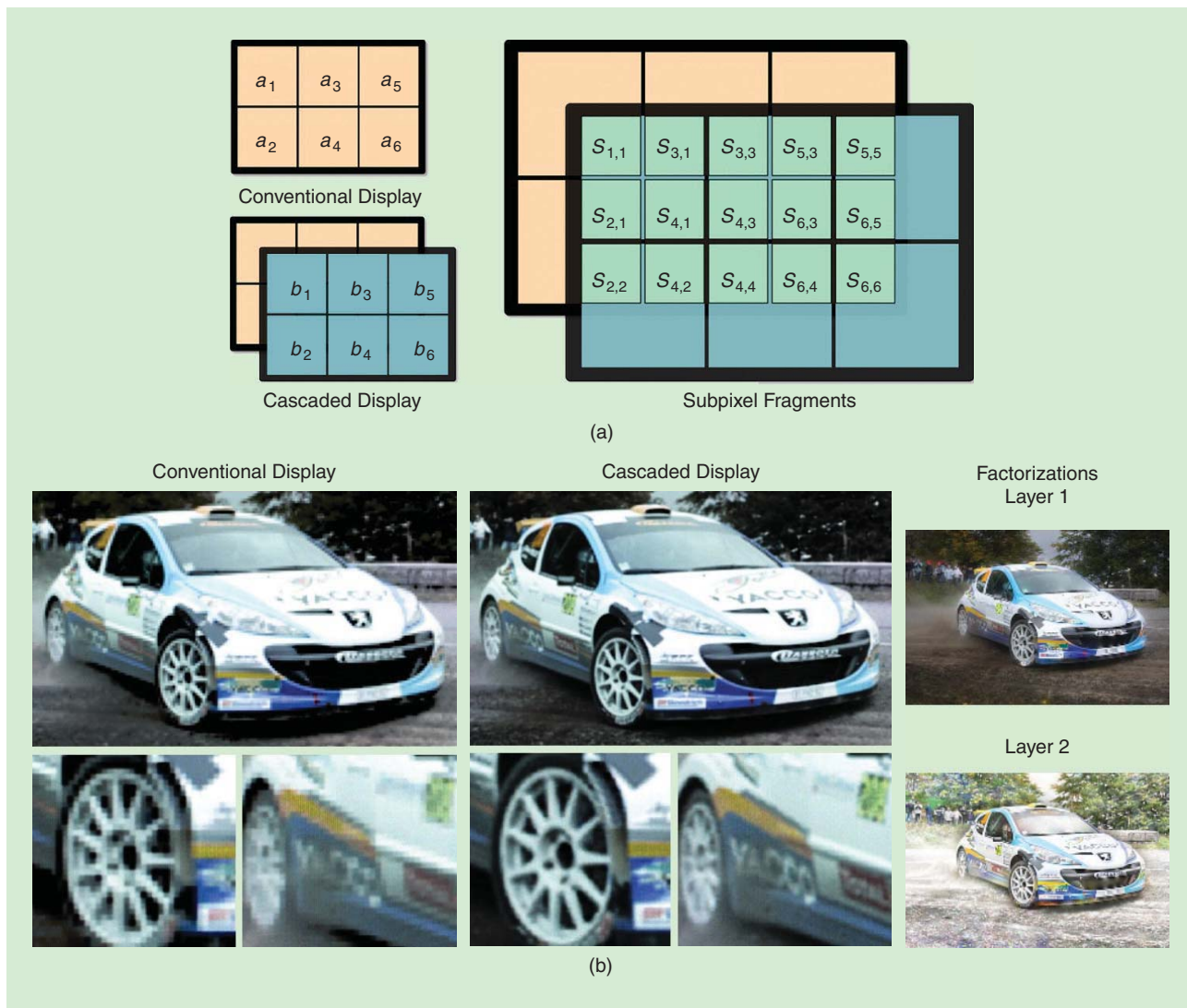
FIGURE 2. (a) Creation of subpixel fragments by cascaded displays. A cascaded display is constructed by layering a shifted pair of conventional displays. The overlap of offset pixel layers creates an array of subpixel fragments (green). (b) Image reconstruction using a single-layer LCD and a cascaded dual-layer LCD. Factorization results are shown on the far right of (b). (Motorsport image courtesy Aurelien Vialatte.)

different strategy: adaptive color display. Through the codesign of display optics, electronics and factored image processing [6] have demonstrated how optimal tradeoffs can be made dynamically in a content adaptive and user centric manner. For example, the spectral power distributions observed in many natural images often do not contain all perceivable colors at once— adapting the display gamut to a specific target image or video clip allows for an optimal tradeoff between brightness and color fidelity to be made. Further, some wide gamut footage may be impossible to be displayed accurately, as it may contain a wide range of colors distributed over the entire perceivable visible color spectrum. In such cases, the perceptually closest approximation of the target should be presented. Human color vision, however, is a complex and nonlinear process; finding a perceptually optimal solution for the color reproduction problem is therefore challenging. Figure 3 demonstrates the concept of adaptive color reproduction for a hyperspectral target image

processed for a three- and a four-primary display with fixed and adaptive gamuts.

Factored spectral displays employ a perceptually driven factorization algorithm that decomposes a target wide-gamut image into the best set of adaptive primaries and corresponding pixel values. The classical problems of primary selection and gamut mapping are thus solved simultaneously and are robustly optimized in a perceptually uniform color space: CIELAB. One of the main benefits of factored spectral displays is its flexibility. Several different display modes are supported by the same device—without mechanically moving parts—simply by switching the software driver. One could imagine a mode that supports very bright monochrome images, for example, text or technical slides, whereas another display mode would support extremely high color fidelity at a lower peak brightness using factorization algorithm; optimal tradeoffs between brightness and color fidelity are dynamically made by the software.

When presenting legacy sRGB content, the device is operated in a conventional three primary mode without any adjustments. Therefore, the display hardware is flexible enough to support existing content without any tradeoffs while also supporting emerging color spaces.

Trends in the display industry are clear: higher resolutions and wider color gamuts will become commonplace in the near future. Increasing resolution will be provided by advanced nano-fabrication technology and potentially driven by factored image generation algorithms. Factored spectral displays, on the other hand, provide a viable solution for adaptive, wide color gamut display that could be seamlessly integrated into cinematic projection systems, home theaters, and office projectors.

Light field displays

Here we discuss emerging light field display technology and a number of unconventional applications. The physical world contains light rays of diverse intensities and colors running in various directions. The light field is a function that describes the amount and color of light flowing in every direction through every point in space. A binocular viewer moving through the light field perceives various depth cues. Binocular cues are created because the two eyes receive different light rays. Motion parallax cues are created because the eyes receive different rays as the viewer's head moves. Accommodation (focus) cues are created because different parts

FIGURE 3. Adaptive color display with factored spectral projector. The device uses a custom light engine with (a) six LEDs. The total addressable gamut of the display spans most of the CIE xy space, but the refresh rate of the digital micromirror device (DMD) only allows for up to four primaries to be used for any projected image. Factored image synthesis allows for the content-specific optimal color gamut spanned by four primaries to be computed while simultaneously solving (b) the gamut mapping problem.

of the viewer's pupil receive different light rays. Light field displays are intended to recreate those cues with sufficient accuracy to enable a high-fidelity 3-D viewing experience.

A light field display emits a 4-D distribution of light rays, which varies over the two spatial dimensions of a display surface but also over the horizontal and vertical viewing angle of each pixel. The display primitives of conventional displays are 2-D pixels (picture elements), those of volumetric or multiplane displays are 3-D voxels (volume elements), and those of 4-D light field displays are light rays, each carrying radiance at some location into a specific direction. Figure 4 illustrates the common two-plane parameterization of a light field: a plane x is located on the physical display screen and another plane u coincides with the pupils of the viewer. To pass our Turing test for displays, i.e., to create a sufficiently persuasive 3-D experience, a 4-D light field display would have to provide appropriate stereoscopic, motion parallax, and focus cues. No such display exists today,

but different tradeoffs can be made to create reasonable approximations of natural light fields.

Over a century ago, Frederic Ives conceived of parallax barriers [1]. A barrier mask consisting of an array of pinholes or slits would be mounted at a slight offset in front of a display such that a viewer would perceive only a subset of the display pixels from any given perspective. The display would render an image that contains the corresponding, interlaced perspectives of the light field. Soon after, Gabriel Lippmann built the first light field camera and display using integral imaging [2]. Instead of pinhole arrays, he mounted microlens arrays on photographic plates, exposed and developed these plates with the lens arrays in place, such that they could be viewed as a light field or glasses-free 3-D image after the fact. The main drawback of parallax barriers and integral imaging is the spatioangular resolution tradeoff: adding more light field viewing zones comes at the cost of reduced spatial display resolution.
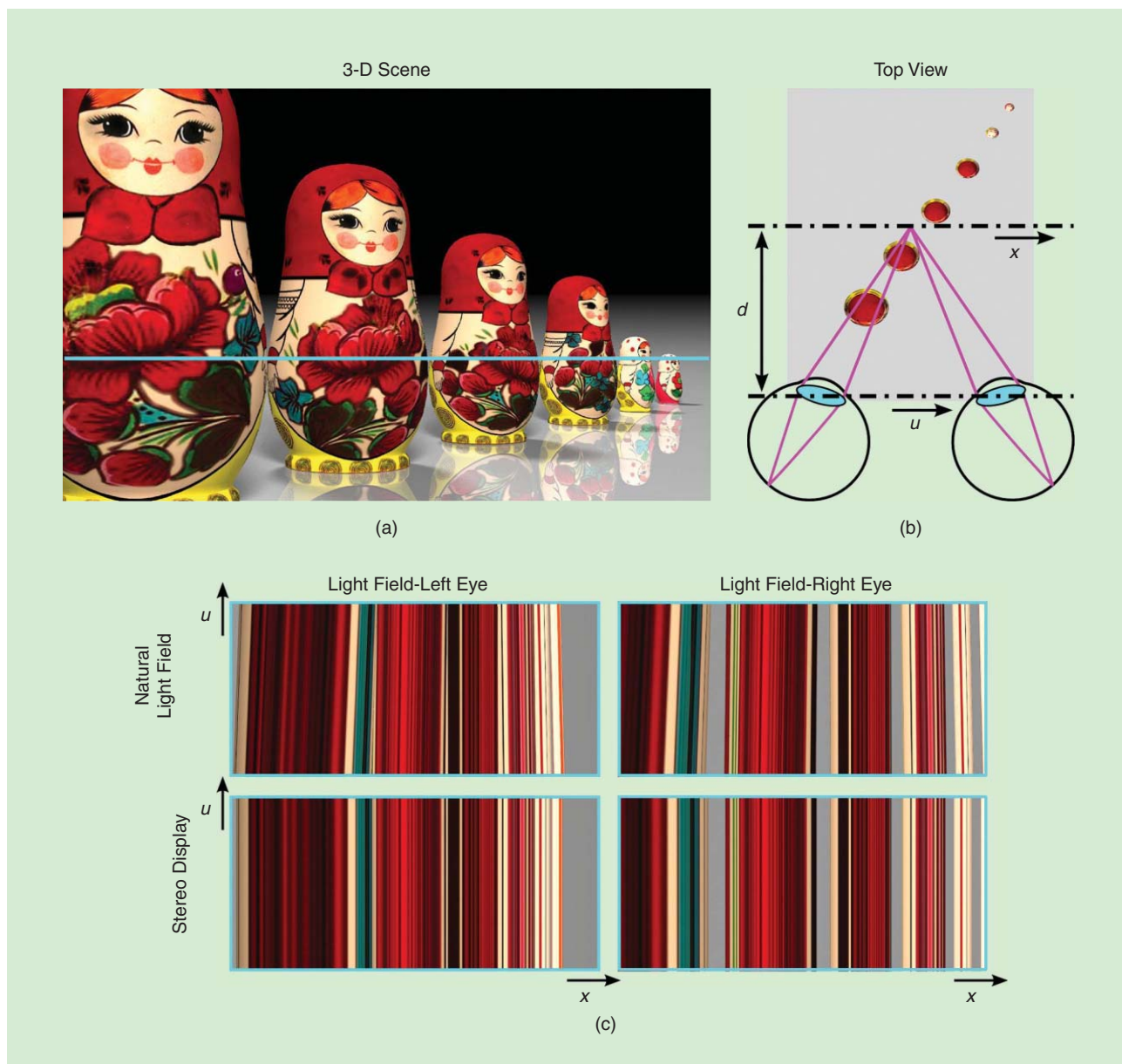
FIGURE 4. The light field of (a) a natural scene is a collection of rays that are parameterized by their coordinates of intersection with (b) two planes x and u. All rays on a horizontal scanline [left (a)] observed in the centers of the viewer's two pupils are shown in (c) (stereo display) and all rays on the same scanline across the viewer's pupils are shown on the upper right (natural light field). The two eyes observe the scene from different vantage points so the left- and right-eye rays differ. Conventional stereoscopic displays do not provide parallax across either pupil and therefore do not support focus cues. The natural light field does provide parallax across each pupil and thereby provides focus cues.

Additionally, parallax barriers are usually dim because most of the emitted light is blocked. To overcome these limitations, many alternative technologies have emerged over the last century to deliver high-resolution, glasses-free 3-D experiences. Yet, none of these can deliver experiences that yet pass our Turing test for displays.

With an ever-increasing demand on image resolution, one of the major bottlenecks in the light field display pipeline is computation. Consider the example of a high-quality, light field display with 100×100 views, each having high-definition (HD) resolution, streamed at 60 Hz. More than one trillion light rays have to be rendered per second

requiring more than 100 Terabytes of floating point RGB ray data to be stored and processed. Further, with conventional integral imaging or parallax barriers, one would need a display panel that has a resolution 10,000 times higher than available HD panels. To relax requirements on display hardware, compressive light field displays and related signal processing algorithms [7] have recently been introduced as a modern version of Ives' vision. These displays exploit two simple properties: 1) light fields of natural imagery are highly redundant, high-dimensional visual signals; and 2) the human visual system has limitations that can be exploited for visual signal compression.
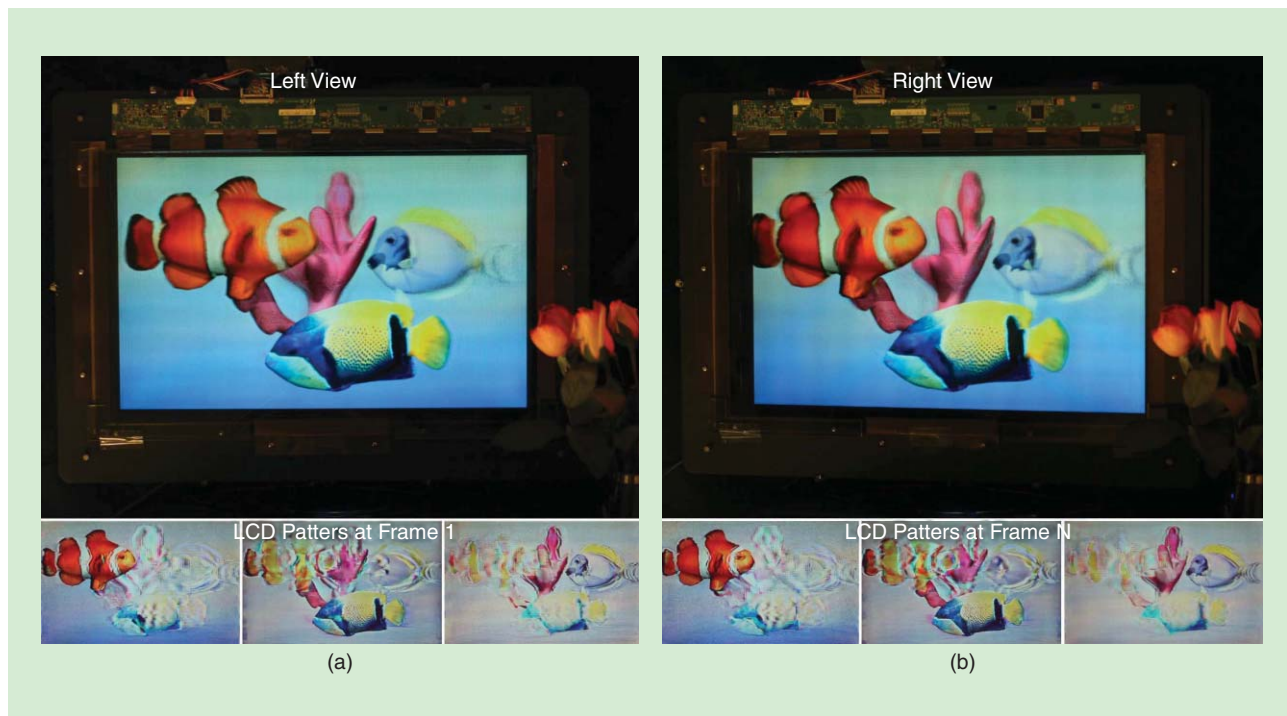
FIGURE 5. A compressive light-field prototype. The prototype uses three stacked layers of liquid crystal displays that are rear-illuminated by a single backlight. A light-field factorization algorithm computes time-multiplexed patterns for all LCD layers that are displayed at a speed exceeding the critical flicker fusion threshold of the human visual system. Perceptually, these patterns fuse into a consistent, high-resolution light field that supports stereo cues and parallax without the need for glasses.

In particular, multiplexing methods (e.g., temporal, spatial, polarization, etc.) can be adopted to optimize the tradeoff between spatial and angular resolution, brightness, etc. in a content-adaptive manner. For example, the refresh rate of modern displays often exceeds the critical flicker fusion (CFF) threshold of human vision. A parallax barrier display implemented with fast LCD panels would allow for the optimal layout of time-multiplexed pinholes to be determined for each target light field. Further relaxing the requirement that the barrier mask is constrained to showing only pinholes leads to the concept of content-adaptive parallax barriers that optimize the time-multiplexed patterns for both display and barrier mask [8]. Such a content-adaptive optimization not only allows adaptive tradeoffs between spatial and angular resolution to be made, but it also allows for display brightness to be optimized with respect to pinhole-based barriers. The light field $\tilde{l}(x,y)$ generated by a time-multiplexed, content-adaptive parallax barrier with two LCDs is given by (1). The corresponding inverse problem is usually formulated as a numerical optimization problem, which can be efficiently solved with NMF approaches.

Compressive light field displays generalize the idea of content-adaptive parallax barriers to a variety of display architectures, including multiple stacked layers of LCDs (Figures 1 and 5), a thin "sandwich" of two LCDs enclosing a microlens array or, in general, any combination of stacked, programmable light modulators and refractive optical elements [9]. Similar to parallax barriers, cascading LCDs usually have a multiplicative effect on the incident light that can selectively attenuate light in some directions [8]–[12]. The aforementioned outlined of light field factorization generalizes to all of these display architectures. Their nonlinear, multiplicative image formation is fundamentally different from the linear, additive image formation provided by multifocal plane displays, volumetric displays, and many other time-multiplexed displays. In general, a nonlinear image formation has the potential to provide more degrees of freedom for the image generation algorithm than an additive, linear image formation [9], [11], [13].

With a factorization framework for generalized parallax barriers in hand, applications to a variety of displays other than television-type systems can be explored. For example, light field projection systems supporting parallax and stereo cues have emerged over the last decade [14], [15]. These types of display systems are most suitable for collaborative experiences and provide impressive image quality over large depths of field. Unfortunately, dozens of projectors have to be employed, making multiprojector light field displays expensive, difficult to calibrate, power hungry, and bulky. The compressive light field methodology has been shown to also apply to projection systems [4]. In this case, the goal is to "compress" the number of required devices, thereby improving power efficiency, form factor, and cost of the system. Hirsch et al. demonstrated that this is possible by generating a light field inside a single projection device, via content-adaptive parallax barriers, and then optically amplifying the limited field of view of the emitted light field using a screen comprising an array of microscopic Keplerian telescopes: one in each screen pixel.

With the re-emergence of near-eye displays for virtual reality (VR) and augmented reality (AR), focus cues have recently attracted a lot of attention in these applications. Most existing near-eye displays naturally provide stereoscopic cues because either two separate microdisplays provide an image for each eye or a single screen is optically split with two lenses. The design principles of near-eye displays available today are very similar to the stereoscopes widely used in Victorian times [16]. In this context, light field displays offer the possibility of providing focus cues. The authors of [17] were the first to demonstrate an integral-imaging-type near-eye display that allowed for a wide accommodation range of the observer with the common caveat of reduced display resolution. (The range was wide enough that correction of refractive errors was supported, allowing the user's prescription eyewear to be removed and digitally correctly by altering the imagery depicted by the near-eye light field display.) More recently, [13] investigated compressive light-field synthesis via two stacked LCDs. This is shown in Figure 6. The device design is inspired by common stereoscopes but it employs two LCD panels spaced at about 1 cm in the display housing. Using light field factorization algorithms similar to those employed by the content-adaptive parallax barriers described previously, a 4-D light-field is emitted independently to each eye, providing parallax over the eye box.

Vision-correcting light field displays

Perhaps one of the most unconventional applications of light field displays is correction of visual aberrations for a human observer [12], [17], [18]. Instead of correcting vision with eyeglasses or contact lenses, the same can potentially be done directly in the screen, allowing for myopia, hyperopia, astigmatism, and even higher-order aberrations to be corrected. For such an application, the light field display presents a distorted light field to the eyes of the viewer such that their natural aberrations optically undistort the light rays, resulting in the desired image (Figure 7). This idea is somewhat similar to wavefront correction with adaptive optics. Implemented with light field displays, the requirement on the angular resolution (density of emitted viewing zones) is similar to those for light field displays supporting focus cues: multiple different viewing zones have to be displayed into the same pupil. Assuming that the prescription of a viewer is known and that their pupil location and diameter can be tracked by a camera, light field displays attempting vision-correction constrain the light field synthesis to the pupil locations. For example, simple implementations using parallax barriers were shown by [12] and, using integral imaging, by [18] and [17]. As shown by Huang et al., the pupils are tracked and the light field is dynamically rendered and predistorted for the viewer such that only viewing zones that actually enter the pupil are considered. Using a 4-D light field display for vision correction, as opposed to a conventional 2-D display, has the advantage that the additional degrees of freedom afforded by the four dimensions allow for the inverse problem of synthesizing

patterns that result in a sharp target image when observed with a defocused eye to be well-posed or invertible. One could attempt to preprocess an image for a 2-D display, such that when the eye is accommodated away from the physical screen it looks like the target image [19]. However, it turns out that this approach is ill-posed and therefore not invertible. But even when the image formation is invertible, vision-correcting displays require the prescription of the viewer to be known (no changes to the hardware are necessary, different prescriptions can be corrected in software) and the pupil positions and diameters to be fixed or tracked. Additionally, this technology has not yet been shown to simultaneously correct for visual aberrations and provide stereoscopic cues in direct-view displays; however, near-eye light field displays, which employ a separate display per eye, have been demonstrated to simultaneously correct visual aberrations while presenting binocular images [17].

> **With an ever-increasing demand on image resolution, one of the major bottlenecks in the light field display pipeline is computation.**

High-dynamic-range displays

The dynamic range of a display usually refers to its contrast, i.e., the ratio of brightest possible display state and the darkest. The dominant display technology for achieving an HDR is a combination of low-resolution light emitting diodes (LEDs) and a high-resolution LCD [5]. The rationale for this design is that the LEDs can be turned off completely, thereby producing a truly black image, or they can be boosted to produce extremely high peak brightness. The main challenge is that LED arrays cannot yet be easily produced at a high resolution and small pixel size. Thus, the HDR display uses a low-contrast, but high-resolution LCD to deliver high image resolution while the low-resolution, high-contrast LED array serves as a programmable backlight. Just over a decade after publication, HDR displays are now widely available in the consumer market, usually under the name "microdimming" or "local dimming."

In addition to providing degrees of freedom for glasses-free 3-D image synthesis, stacked LCDs also decrease the black level of the produced image compared to any individual LCD. Although the peak brightness of a multilayer display is usually reduced compared to each of the respective layers, the observed contrast is increased. This makes multilayer displays ideal display systems for HDR image generation. However, with nonnegligible separations between attenuators, multilayer HDR decomposition becomes a 3-D display problem similar to light field image synthesis, since all viewpoints must produce an accurate rendition of the 2-D image within the target field of view. The factorization framework described by [9], [10] inherently accounts for the limited contrast of each layer, thereby allowing simultaneous optimization of dynamic range and accurate multiview imagery. In a 2-D HDR display mode, the target light field encodes a single plane (e.g., coincident with the front layer), with a texture given by the desired HDR image. Figure 8 shows the result from a parallax-free 2-D HDR display prototype. The optimized layers in (b)
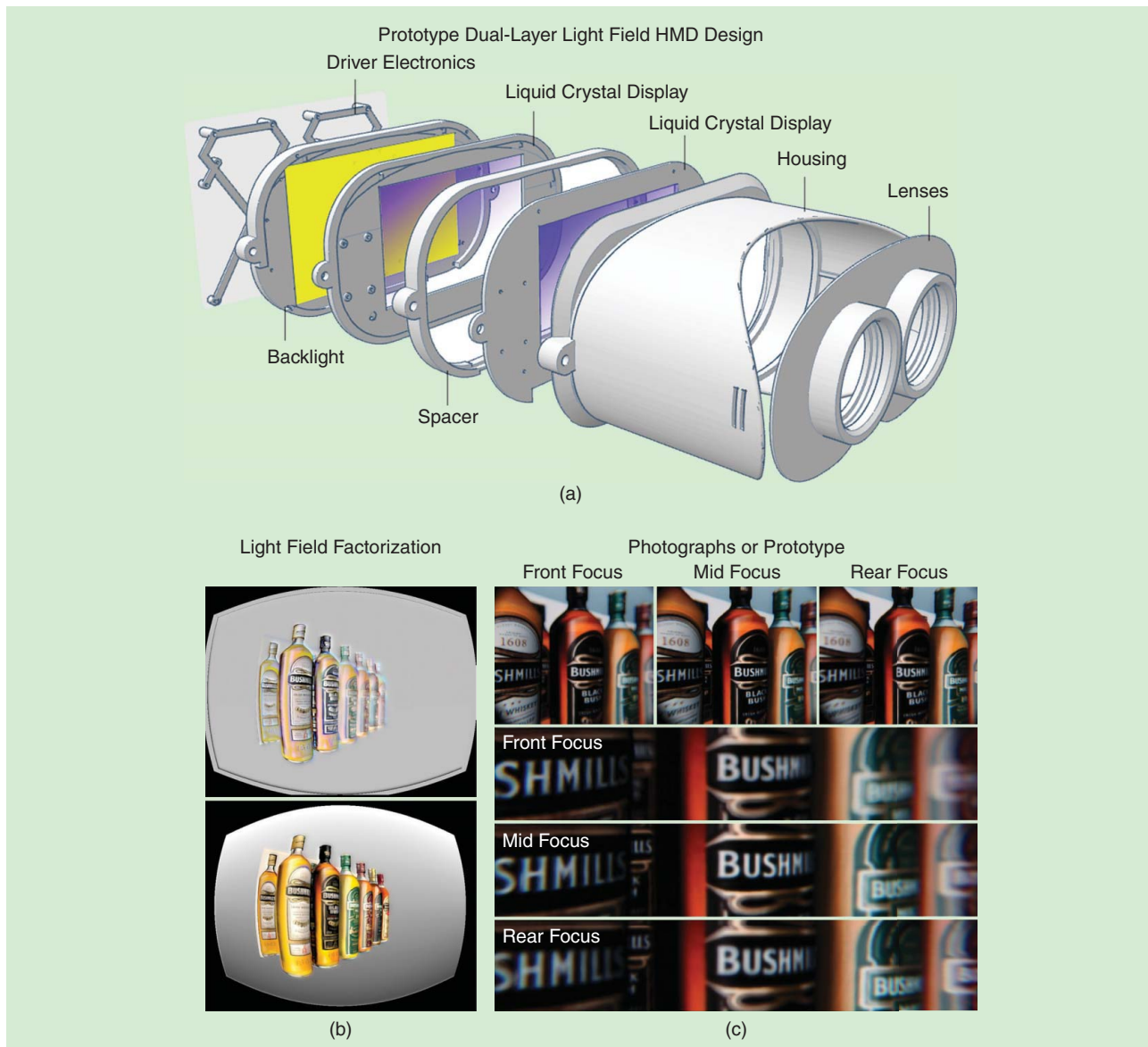
FIGURE 6. Near-eye light field display with support for focus cues. The design is based on the stereoscope but uses two stacked LCDs inside the device to generate (a) a separate light field for each eye. (b) Light-field factorization is employed to computer-generate patterns shown on the front and rear panel. (c) The light field provides sufficient angular resolution for focus cues to be supported, such that a viewer can accommodate within the virtual scene.
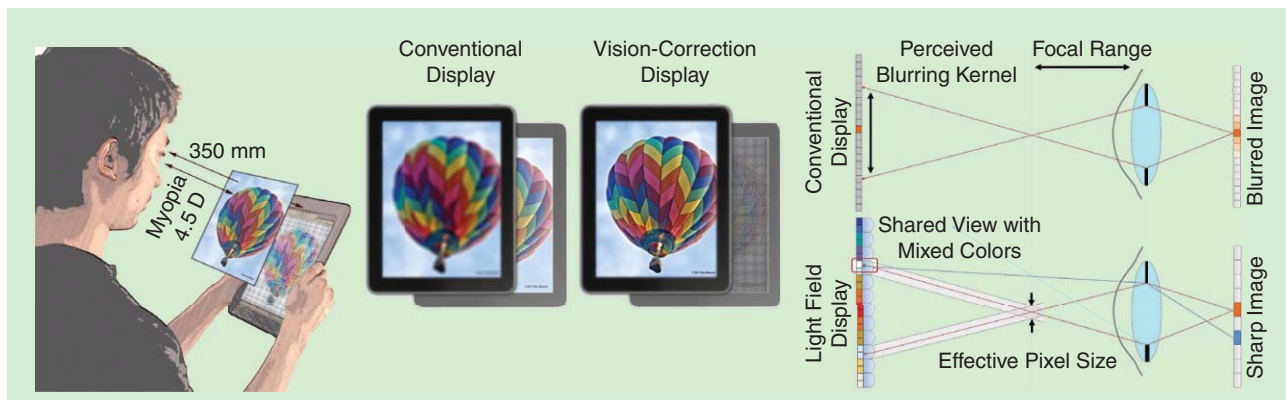
FIGURE 7. A Vision-correcting light field display. Illustration of conventional and vision-corrected display. The device is outside the accommodation range of the viewer and usually looks blurred. A light field display with a sufficient angular resolution to image multiple views into the same pupil provides degrees of freedom for correcting myopia, hyperopia, or higher-order aberrations.
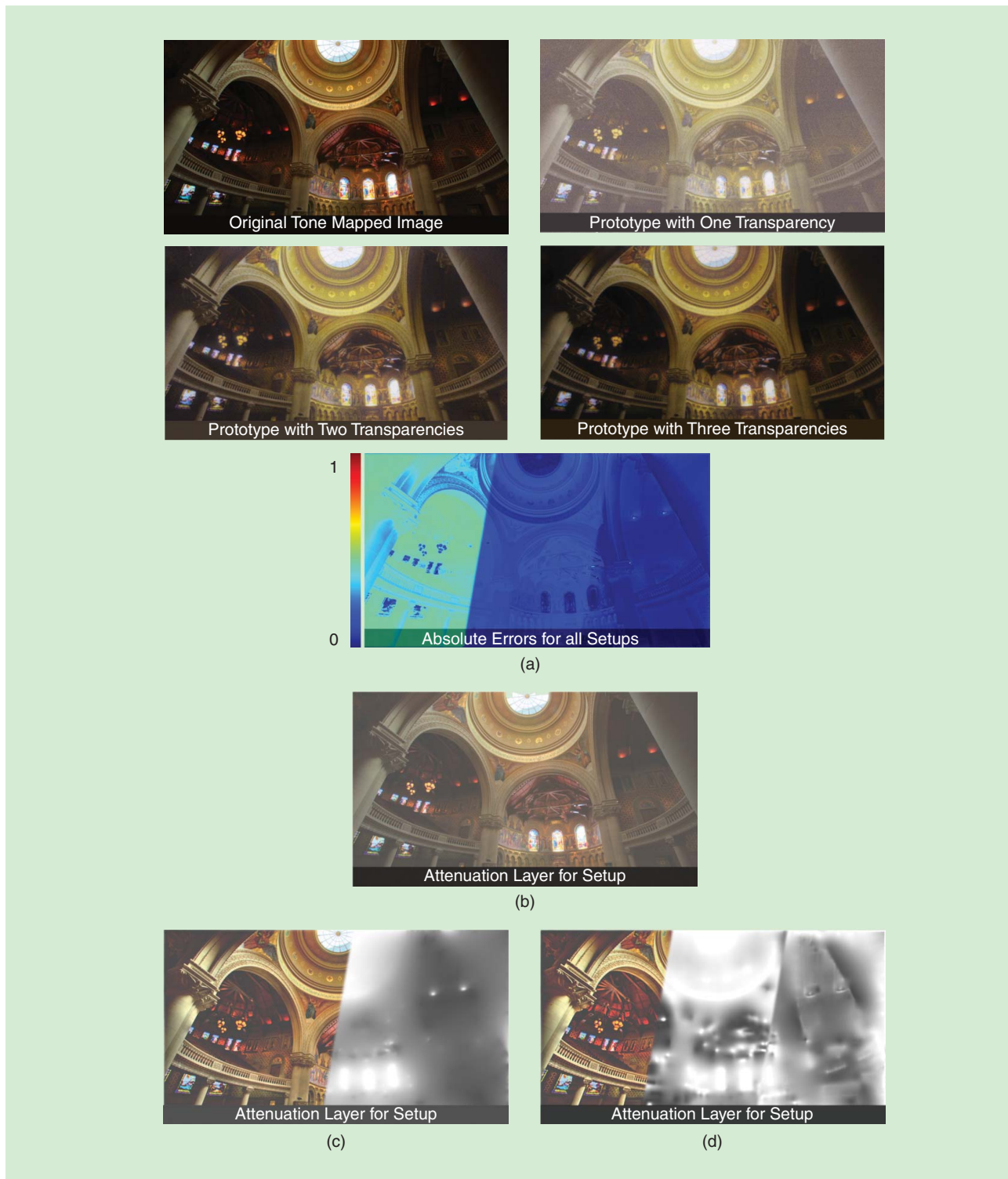
FIGURE 8. A multilayer, parallax-free HDR image display. The target HDR image is shown with (a) tone mapping. Photographs of a multilayer HDR display prototype, incorporating (b) one, (c) two, and (d) three layers (each with a contrast of 3.3:1), are shown; note how image contrast is improved by applying tomographic image synthesis method together with multiple, physically disjoint attenuation layers. The absolute errors between the simulated reconstructions and the target image are shown in (d).

account for the nonzero black level of the printing process and are scaled appropriately before printing.

Although Figure 8 demonstrates improved contrast with static, layered transparencies, the same concept applies to stacks of programmable LCDs, as recently proposed by [20]. If high-speed LCDs are available, the time-multiplexing aspect described in the previous sections can be exploited as it offers increased degrees of freedom compared to static displays.

A variant of this idea was also proposed by [4], who designed and built an HDR light field projector using stacked micro-displays in a projection system along with a custom screen.

Discussion

The dream of "Lippmann's window" continues to appear beyond our reach. Factored displays alone cannot bridge the gap to enable faultless displays with today's electro-optical components and optimization algorithms. However, as outlined in this article, they may provide a shortcut to passing the visual Turing test, allowing for significant enhancements in resolution, dynamic range, color reproduction, and light-field characteristics. Factored displays achieve this by exploiting two key properties: the visual world is structured and the human visual system is flawed. As such, the design of displays should be reconsidered as a compression problem—not just as applied to image capture, storage, and transmission, as is typically done, but also as applied to the modulation of photons upon display by one or more spatial light modulators. Yet, unlike related compressive imaging systems, compressive displays must contend with a major obstacle: the ultimate decoding algorithm is immutable, being that instantiated by the human visual system itself. As a result, all interventions must be made before a ray of light impinges on the retina of a viewer, but under models of its perception by the viewer thereafter. In this sense, factored displays offer a unique opportunity to merge advanced signal processing and optimization methods with our understanding of the limitations of the human vision system and how to exploit them computationally.

The primary limitations of factored displays today include reduced light transmission, optical diffraction effects that are observed as an additional blur that the presented, ray-based image formation model does not account for, and increased computation. With increasing computational resources and emerging electro-optical light modulators, we believe that factored displays will soon demonstrate their full potential in a range of computational displays. Further research on adapting these spatial light modulators and improved factorization methods may place displays that successfully pass our visual Turing test within reach.

Authors

Gordon Wetzstein (gordon.wetzstein@stanford.edu) is an assistant professor of electrical engineering and, by courtesy, of computer science at Stanford University. He is the leader of the Stanford Computational Imaging Group, an interdisciplinary research group focused on advancing imaging, microscopy, and display systems. At the intersection of computer graphics, machine vision, optics, scientific computing, and perception, his research has a wide range of applications in next-generation consumer electronics, scientific imaging, human-computer interaction, remote sensing, and other areas.

Douglas Lanman (dlanman@gmail.com) received his B.S. degree in applied physics with honors from the California Institute of Technology in 2002 and his M.S. and Ph.D. degrees in electrical engineering from Brown University in 2006 and 2010, respectively. He is a research scientist at Oculus VR R&D. His research is focused on computational displays and imaging systems, emphasizing compact optics for head-mounted displays, glasses-free three-dimensional (3-D) displays, light field cameras, and active illumination for 3-D reconstruction and interaction.

References

[1] F. E. Ives, "Parallax stereogram and process of making same," U.S. Patent 725,567, Apr. 14, 1903.

[2] G. Lippmann, "La Photographie Intégrale," *Academie des Sci.*, vol. 146, pp. 446–451, Nov. 1908.

[3] F. Heide, D. Lanman, D. Reddy, J. Kautz, K. Pulli, and D. Luebke, "Cascaded displays: Spatiotemporal superresolution using offset pixel layers," *ACM Trans. Graphics*, vol. 33, no. 4, pp. 60, July 2014.

[4] M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Trans. Graphics*, vol. 33, no. 4, pp. 58, July 2014.

[5] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 760–768, Aug. 2004.

[6] I. Kauvar, S. Yang, L. Shi, I. McDowall, and G. Wetzstein, "Adaptive color display via perceptually-driven factored spectral projection," *ACM Trans. Graphics*, vol. 34, no. 6, pp. 165, Nov. 2015.

[7] F. Heide, G. Wetzstein, R. Raskar, and W. Heidrich, "Adaptive image synthesis for compressive displays," *ACM Trans. Graphics*, vol. 32, no. 4, pp. 132, July 2013.

[8] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: Optimizing dual-layer 3-D displays using low-rank light field factorization," *ACM Trans. Graphics*, vol. 26, no. 6, pp. 163, Dec. 2010.

[9] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive fight field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graphics*, vol. 31, no. 4, pp. 80, July 2012.

[10] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3-D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Trans. Graphics*, vol. 30, no. 4, pp. 95, July 2011.

[11] A. Maimone, G. Wetzstein, D. Lanman, M. Hirsch, R. Raskar, and H. Fuchs, "Focus 3-D: Compressive accommodation display," *ACM Trans. Graphics*, vol. 32, no. 5, pp. 153:1–153:13, 2013.

[12] F.-C. Huang, G. Wetzstein, B. Barsky, and R. Raskar, "Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays," *ACM Trans. Graphics*, vol. 33, no. 4, pp. 59, July 2014.

[13] F.-C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graphics*, vol. 34, no. 4, pp. 60, Aug. 2015.

[14] T. Balogh, "The HoloVizio System," *Proc. SPIE,* vol. 6055, pp. 1–4, Jan. 2006.

[15] A. Jones, J. Liu, J. Busch, P. Debevec, M. Bolas, and X. Yu, "An autostereoscopic projector array optimized for 3-D facial display," in *Proc. ACM SIGGRAPH Emerging Technologies*, New York, 2013, pp. 3.

[16] C. Wheatstone, "Contributions to the physiology of vision.–Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philos. Trans. R. Soc. London*, vol. 128, no. 1838, pp. 371–394, Jan. 1838.

[17] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graphics*, vol.32, no.6, pp. 220, Nov. 2013.

[18] V. Pamplona, M. Oliveira, D. Aliaga, and R. Raskar, "Tailored displays to compensate for visual aberrations," *ACM Trans. Graphics*, vol. 31, no. 4, pp. 81, July 2012.

[19] J. I. Yellott and J. W. Yellott, "Correcting spurious resolution in defocused images," *Proc. SPIE*, vol. 6492, pp. 12, Feb. 2007.

[20] F. Heide, J. Gregson, G. Wetzstein, R. Raskar, and W. Heidrich, "Compressive multi-mode superresolution display," *OSA Optics Express*, vol. 22, no. 12, pp. 14981–14992, 2014.

SP

Xiang Huang, Erich Uffelman, Oliver Cossairt,
Marc Walton, and Aggelos K. Katsaggelos

Computational Imaging for Cultural Heritage

*Recent developments in spectral imaging, 3-D surface measurement,
image relighting, and X-ray mapping*
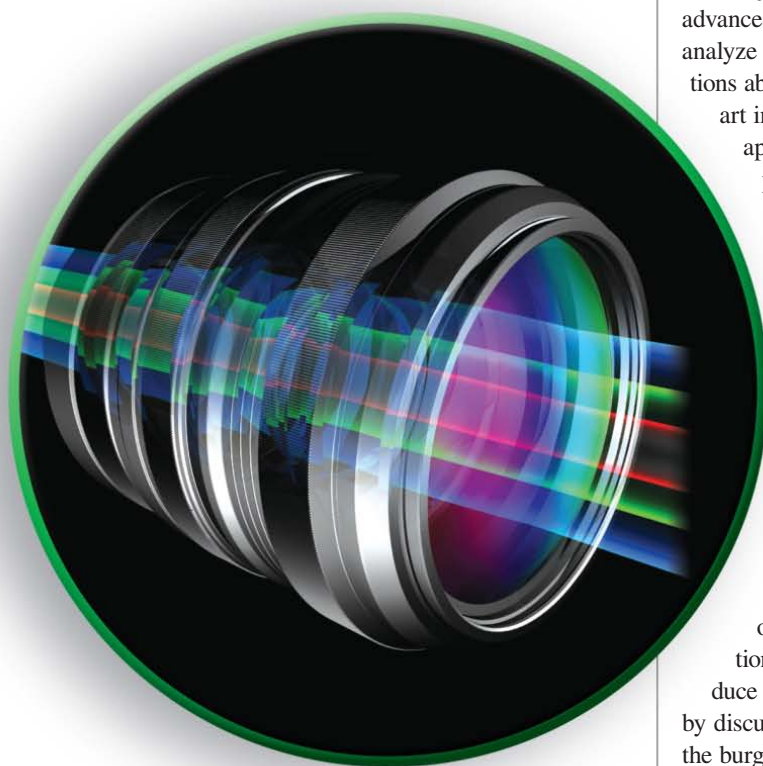
B ecause art is inherently visual, the use of imaging has long been an important way to understand its structure, form, and history. Recently, new ways of engaging with objects from our shared cultural heritage are possible with advances in computation and imaging that allow scientists to analyze art noninvasively, historians to pose new social questions about the art, and the public to explore and interact with art in ways never before possible. There is a rich history in applying image processing techniques to conventional photographic images of works of art, many of which have been highlighted in previous special issues of *IEEE Signal Processing Magazine* (e.g., the 2008 and 2015 July issues). Building on these contributions, this article comprises a survey of techniques where computation is central to the image acquisition process. Known as *computational imaging*, the methods being pioneered in this field are increasingly relevant to cultural heritage applications because they leverage advances in image processing, acquisition, and display technologies that make scientific data readily comprehensible to a broad cohort of nontechnical researchers interested in understanding the visual content of art. Presently, only a small research community undertakes computational imaging of cultural heritage. Here we aim to introduce this growing new field to a larger research community by discussing: 1) the historic background of imaging of art, 2) the burgeoning present day community of researchers interested in computational imaging in the arts, and finally, 3) our vision for the future of this new field.

©ISTOCKPHOTO.COM/YAKOBCHUK

Introduction

The use of electromagnetic radiation, beyond the limits of eyesight, to visualize artworks may be traced to 1895 when Roentgen made his first X-ray shadowgraphs, one of which happened to have been a painted surface. However, it was not until the 1930s, when X-radiography first entered into museums, that a new art history formed around the ability to

assess style and attribution of an artwork from aspects of the painted surface not visible to the naked eye [1]. This trend continued with other wavelengths of illumination. Specifically, ultraviolet (UV)-induced visible fluorescence helped reveal areas of loss/repair and provided a general sense of chemical composition [2]. By the late 1960s, infrared (IR) reflectography was in routine use in museums to reveal hidden underdrawings and preparatory marking in paintings [3]. More specialized techniques, such as autoradiography achieved by neutron bombardment of a work of art, opened up the possibility of combining elemental composition together with imaging for the first time [4]—a technique that would inspire further developments in X-ray fluorescence imaging several decades later [5]. By the 1980s, new three-dimensional (3-D) acquisition techniques were being explored for the 3-D documentation and display of cultural objects, both of which remain relevant subjects of investigation to this day [6].

The recent explosion in imaging of cultural heritage has grown mainly out of the fields of remote sensing and color science. Of particular note is the use of hyperspectral and multispectral imaging instruments for pixel-by-pixel material characterization [7]. A parallel development has been the use of synchrotron-based X-ray fluorescence and diffraction imaging that has grown in conjunction with the diversification of users of these large-scale facilities from all research disciplines. Around the same time, computational illumination techniques were developed to dynamically relight works of art in postcapture [8], [9]. With the advent of inexpensive digital scanners, several researchers have focused on digitization of existing X-radiographs of canvas paintings, enabling recent advances in image processing algorithms to be applied to these historical works, such as in the canvas weave project initiated at Cornell University [10], [11]. The proliferation of inexpensive digital sensors have been allowing museums to capture large amounts of high-resolution photographs in multiple modalities that were then computationally stitched together to provide seamless image mosaics with unprecedented detail [12]. Optical coherence tomography [13] and THz imaging [14] provide in-situ 3-D reconstructions of microscopically thin layers of paint comprising pictures and drawings. At the other extremes of scale, the popularity of both light detection and ranging (LiDAR) [15] and structure-from-motion (SfM) techniques [16] have allowed us to search for ancient cities and document the historic landscapes of modern ones.

In this article, recent developments are discussed in four core areas that have served to advance the field of cultural heritage into new territory: multispectral and hyperspectral imaging, 3-D shape scanning and recovery, image relighting, and macro X-ray imaging. Key developments in each of these areas have dramatically changed the landscape of how one noninvasively

New ways of engaging with objects from our shared cultural heritage are possible with advances in computation and imaging that allow scientists to analyze art noninvasively, historians to pose new social questions about the art, and the public to explore and interact with art in ways never before possible.

documents, assesses, interprets, and conserves culturally significant artifacts housed in museums around the world.

Multispectral and hyperspectral imaging

Human eyes only perceives visible light (380 nm~750 nm) with three types of color-sensitive cones: "red," "green," and "blue." Multispectral and hyperspectral techniques extend the measurable spectrum from visible light to UV (10 nm~380 nm) and IR (750 nm~1 mm) lights with increased resolution: typically multispectral imagery has three to ten bands, while hyperspectral imagery could have hundreds or even thousands of narrower (e.g., 10 nm) bands. Multispectral and hyperspectral imaging provide a wealth of information across space and wavelengths comprising large swaths of the electromagnetic spectrum. The techniques are also flexible since they can be scaled from the imaging of landscapes, when used on satellites and telescopes, down to the microscopic. Also, importantly, these imaging spectroscopies are nondestructive under the normal conditions of their implementation. Liang's recent review [17] should be consulted for developments in the field through 2012, but a brief introduction is provided here.

There are typically three principal ways of obtaining multi/ hyperspectral data sets:

1) imaging the entire object at once through a series of different filters (or through a single filter whose bandpass characteristics may be controlled), e.g., in 2004, Lumiere Technology (http://www.lumiere-technology.com/Pages/Services/services2.htm) used 13 filters from UV to IR and a 240-megapixel camera to image the famous "Mona Lisa" in the Louvre Museum

2) scanning a linear slit view of the object through a grating that spreads the relevant spectral region onto a two-dimensional (2-D) sensing array

3) scanning the entire object point by point across its x-y surface [18].

Aspects of cost, time, and instrumental design parameters will dictate the choice of image acquisition method.

The resulting x-y surface images are "stacked" as a function of wavelength thus creating an image cube that may be interrogated in two ways. If the cube is "sliced" parallel to the x-y image face, one can analyze each of the images taken at each wavelength. Such an analysis at infrared wavelengths might readily provide an image of an underdrawing beneath the surface of a painting. If the stack is rotated by 90° degrees to x-y image face of the cube, one can obtain the detected spectrum at every pixel. Because chemical components have distinguishable spectral responses, multivariate statistical methods such as principal component analysis (PCA) can provide information on the spatial distribution of different materials. Examining the PCA images, spectral angle maps of end members, or

mapping regions of interest provide a wealth of information on the composition, execution, and condition of the art object even when an artist never employed a pure pigment in their composition and the consequential endmember spectra do not correspond to pure pigments.

Traditionally, these methodologies deal with spectroscopic data in the near UV, visible, and near IR. Recently, however, macro X-ray fluorescence (XRF) scanning and macro X-ray diffraction (XRD) scanning have moved these applications into a revolutionary new area of analysis (see the section "Macro X-ray Methods" for more details). In addition, recent work involving mid-IR imaging [19] promises a wealth of new opportunities in cultural heritage analysis; i.e., using the fingerprint region of the IR spectrum (a region typically thought of as ranging from roughly 400 to 1,500 cm^{-1}, a range not completely available yet in commercial mid-IR scanners) enables mapping of a variety of pigments and binders, and comparison of the hyperspectral mid-IR data to point spectra obtained by a conventional IR spectrometer in reflectance mode demonstrates the power of this new technique, which should expand as the accessible mid-IR range of the instrumentation increases.

> **The recent explosion in imaging of cultural heritage has grown mainly out of the fields of remote sensing and color science, such as spectral imaging and macro X-ray scanning.**

Currently, powerful combinations of multispectral and hyperspectral imaging with other imaging and analytical modalities are revealing the rich information that can be gleaned from the synergy of combined methodologies (e.g., Raman spectroscopy, fiber optic reflectance spectroscopy) [7], [20]. While multi/hyperspectral data cubes contain rich material information, they are challenging to acquire and analyze due to the sheer size of these data sets. Traditionally, dimension reduction and feature exaction techniques such as PCA and end-member analysis were used for hyperspectral data [18]. Recently, compressive sensing has been used in hyperspectral imaging for sensing, reconstruction, and material classification [21]. The technique exploits the sparsity of signals by solving the following optimization problem:

$$\min_f \| f \|_p, \text{ st. } \| g - Af \|_2 \le \epsilon, \tag{1}$$

where ϵ controls the tolerate approximation due to noise, and $p = 0$ or 1 describe the sparsity of the signal as L_0 norm (total number of nonzeros) or L_1 norm (sum of absolute value).

The optimization framework from (1) can be used to decompose measured reflectance spectra into pure spectral components, with can be used to identify and "unmix" heterogeneous pigment combinations on painting surfaces [22]. In this approach, g is defined as the spectral vector $g(\lambda)$ at a given pixel, and each column of A is a spectral vector of known material from a predetermined dictionary of pigment spectra. Solving the optimization problem then reconstructs the sparse coefficient f, which tells us the material components of that pixel, along with their relative concentrations. Note that these spectral decomposition methods are entirely linear, and therefore cannot accurately model nonlinear effects such as wavelength-dependent scattering, self-absorption, etc., which may be common in a real painting material.

We conclude this section by highlighting a recent computational advance that leverages tremendous power from combined imaging modalities, i.e., advances in registration software that enable the "stacking" of images or especially the stacking of full data cubes from different regions of the electromagnetic spectrum [12].

The entire reason for manipulating data cubes from different regions of the electromagnetic spectrum—visualization of different and often complementary data—brings with it a concomitant challenge: Is it possible to register images in which the features below the immediate surface have been moved, painted over, or scraped away? Artists often experimented with multiple underdrawings on the same painting and then overpainted those underdrawings with further alterations in the paint layers. Revealing and spatially registering these pentimenti can provide significant insights to artistic process and intent.

One solution [12] has utilized image fusion methods in which the modulus of the wavelet transform is determined and allows for the identification of "candidate control points," common features in the different images that can be used for alignment. The true functionality of the algorithm comes from how it assesses the statistical quality of these control points and seeks a wide enough spatial distribution of them so that a function may be calculated to register a variety of different sized images. Registration often requires a couple hours of computational time on a desktop PC. In the case of Figure 1, a rotated IR image has been registered with an X-radiograph followed by an adjustment of their relative intensities to clarify the legibility of the underlying portrait. This legibility enhances the confidence in assigning the underlying portrait to an artist other than Vermeer. (Readers are encouraged to view the movies of registered images in the supplementary materials of Conover, et al. [12] at http://link.springer.com/article/10.1007/s00339-015-9140-1.) This software is not only exceptionally powerful, but it is also readily implemented; one of the authors of this article routinely trains 18-year-old first-year college students how to obtain multiple multispectral image cubes and to register them.

Three-dimensional shape scanning and recovery

Since the 1990s, 3-D laser scanning has made the shape capture of 3-D cultural-heritage objects possible. In one of the first efforts to capture sculpture in the round, the Digital Michelangelo project, researchers scanned several Michelangelo statues, including his masterpiece David [6]. While the project produced spectacular geometries of statues meters in size at millimeter resolutions from a combination of three expensive laser scanners, the results still fails with in specular/shiny areas of
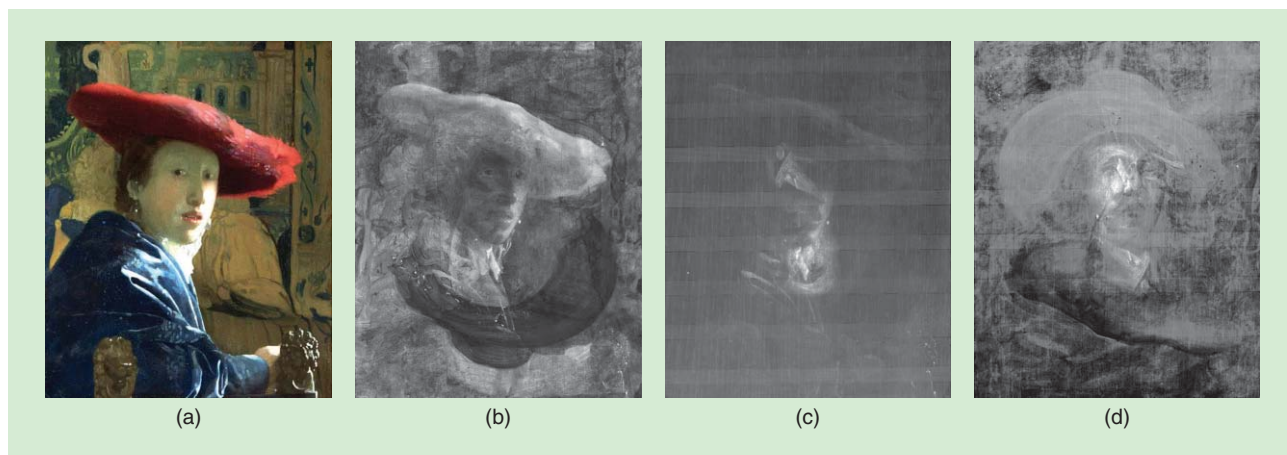
FIGURE 1. (a) A color image of Johannes Vermeer's *Girl with the Red Hat* (1665/1666). Andrew W. Mellon Collection, 1937.1.53, National Gallery of Art, Washington, D.C., (b) infrared reflectance (2,100-2,400 nm), (c) X-radiograph, and (d) summation of the rotated X-radiograph and the intensity-inverted and rotated infrared reflectance image. (Images and figure caption used with permission from [12].)

objects. Processing these data is also a nontrivial task since gaps in the scanned area must be filled and the different scans must be aligned and registered in X, Y, Z space. Also, further processing to map color information on the acquired 3-D meshes is needed to produce a fully rendered result. The whole process was expensive (US$2 million) and took 32 people years (1997–2004) to plan, scan, and model ten statues. Thus, laser scanning poses many challenges that limit its widespread use. Beside academy, Aicon3D (http://www.aicon3d.com/) and Metis (http://metis-digital.com/) companies currently offer high-resolution 3-D scanner for artworks, and Artmyn (http://www.artmyn.com) provides Web solution for 3-D imaging, using the one of the oldest New Testament papyri (Papyrus 66) as an example. Due to the limitation of laser scanning, new methods are still needed for quick and cost-effective ways for digitally archiving art.

Another effective approach to the imaging of extremely large structures has been airborne LiDAR remote sensing techniques, which make it possible to record shapes on the extreme landscape scale. For instance, the 200 km^2 area of the ancient Maya landscape at Caracol, Belize, was scanned with a resolution that could resolve structures of roughly 25-cm height [15]. The data obtained helped researchers understand that the ancient Maya could radically modify their landscape to create a sustainable urban environment. On the other hand, Google Earth (https://earth.google.com/) has made 3-D buildings of cities (e.g., Chicago, Illinois) and 3-D historical sites (e.g., Rome's Colosseum) easily accessible to the public, though in low resolution.

To overcome some of the limitations of terrestrial and airborne laser scanning due to the high operational, researchers

> To overcome some of the limitations of terrestrial and airborne laser scanning due to the high operational, researchers have more recently used a more convenient and purely image-based method, SfM, a photogrammetry stereo technique, to recover the shape of historical sites.

have more recently used a more convenient and purely image-based method, SfM, a photogrammetry stereo technique, to recover the shape of historical sites. In 2006, the 3-D structure of the Colosseum in Rome was generated from a large collection of consumer photos taken at different viewpoints [16]. These photos were gathered from an Internet-sharing website. The photo explorer uses image-based rendering techniques to create smooth transition between different viewpoints, so the user can comfortably and virtually tour historic locations. SfM may also be used to great effect on smaller moveable objects, however, the depth accuracy of the photogrammetry stereo method is limited to only textured surfaces and fails on feature-less surfaces, hence, the depth resolution is typically lower than the lateral resolution at each pixel.

Another image-based method, photometric stereo, recovers the 3-D shape of an object by taking multiple images at fixed view but varies lighting positions. Photometric stereo models the image intensity as a function of surface normal, reflectance, and lighting/viewing angle. The surface normal is recovered by solving an optimization problem, and the 3-D surface of the object can be recovered by integrating the surface normal across the field of view. Unlike photogrammetry, photometric stereo works extremely well on textureless surfaces and can produce high-resolution normal maps. Classic photometric stereo methods assume a point light source placed at infinitely far away from direction L, and assume the material is Lambertian with albedo reflectivity of k, so that the reflected light intensities at an object point with surface normal n becomes

$$I = k(n \cdot L). \qquad (2)$$

By taking a series of measurements I, with different, but known lighting direction L, the surface normal n, and albedo k can be estimated from a system of linear equations using, for instance, a least squares method.

Classical photometric stereo assumes a distant-light model. This considerably simplifies the problem, as it produces constant lighting angle and incident radiance across the object surface. However, distant-light sources are impractical due to finite space and energy constraints. As a result, for a typical photometric stereo capture setup, lighting angle and incident radiance vary across the object surface. Using the simplified far-light model from (2) with such a setup produces a 3-D shape with large global error [25]. Recently, researchers have explored the near-light photometric stereo method to recover millimeter to submillimeter scale markings on the surfaces of Paul Gauguin's paintings (Figure 2) [9]. The depth maps acquired achieve a depth precision of fewer than 100 μm for a field of view as large as 300 mm. These depth maps have revealed new details of how Gauguin produced his paintings using his unique drawing transfer techniques.

Classical photometric stereo also assumes Lambertian surfaces with perfect diffuse reflection. However, this assumption is invalid for a large class of real materials such as metals, plastics, and glass, which exhibit different combinations of diffuse and specular reflections. The most accurate way to model how light is reflected from an opaque surface uses the bidirectional reflectance distribution function (BRDF), which is a four-dimensional function $f_r(\omega_i, \omega_o)$, which depends on the incoming light direction ω_i and outgoing light direction ω_o. The BRDF is the most general way to model surface reflection (not considering subsurface scattering), but it also severely complicates the photometric stereo problem. As a result, several researchers have investigated lower-dimensional reflectance models for

> **While photometric stereo can produce submillimeter precision surface measurements with a large field of view, other methods can be used to measure surface detail on the microscopic scale.**

use with photometric stereo algorithms. Ikehata et al. [23] models the non-Lambertian, specularities and shadows as additive corruption E, so that the observed image intensity is $I = k(n \cdot L) + E$. Assuming the corruption E is spatially sparse, the problem can be solved by compressive sensing algorithms by modeling the optimization similar to the Lagrange form of (1) as:

$$\min_{k,n,E} \| I - k(n \cdot L) - E \|_2 + \lambda \| E \|_0, \qquad (3)$$

where λ is an nonnegative parameter controls the balance between data fit and sparsity.

While photometric stereo can produce submillimeter precision surface measurements with a large field of view, other methods can be used to measure surface detail on the microscopic scale. Optical coherence tomography (OCT) has recently been employed for examining the layer structure of paintings [13]. High-resolution 3-D images at a micron scale can be reconstructed thus revealing the underlayers of paintings and their corresponding depth positions. Originally proposed for biomedical imaging of structures such as the eye, OCT can produce high-resolution contrast depth maps. OCT presents challenges in that the instrumentation is expensive and can only scan centimeter-sized areas. The depth maps obtained are also not linked to material color information, so interpreting these data is not immediately intuitive.

Image relighting for cultural heritage

In addition to the 3-D geometry, characterizing surface appearance under different lighting conditions is also critical for cultural heritage. The appearance of an artwork is the sum result of how its material and microstructure interact with all possible incoming light rays and all the possible subsequent measured outgoing light rays that may have been reflected, absorbed, scattered, refracted, and transmitted from the artwork's surface. This compressive light-transport function combines each possible incident light location, wavelength, direction, polarization with how this incident electromagnetic radiation scatters underneath the object's surface, and global illumination effects such as self-shadowing and interreflection. It is an immense totality of measurements that is only theoretically possible to collect completely. Consequently, the light-transport function at a fixed viewpoint may be easier to gather by capturing images of artwork lit from various light directions.

FIGURE 2. (a) The setup for capturing photometric stereo of Gauguin's *Nativity*: a color checker for color calibration, a 3-D calibration target for 3-D surface calibration, a reflective sphere for calibrating light direction, and the work of art. (b) Several frames from an animation visualizing the 3-D surface shape at the location of the lines drawn in *Nativity*. The 3-D reconstruction shows clear evidence of protrusions on the page where ink has been deposited. This is solid evidence for the ink being transferred from a matrix such as that in a monotype transfer process. (Images and figure caption used with permission from [9].)

Reflectance transformation imaging (RTI), originated from polynomial texture mapping (PTM), is one such approximation method. Malzbender [8] first discussed RTI as a method for examining an artwork using interactively changeable lighting conditions with a set of digital images. By interpolating multiple images of a work, each with different illumination angles from a fixed camera position, an "active photo" may be produced with easy controls that encourage exploration to see vanishingly subtle features, including self-shadowing and interreflection. The PTM typically stores six coefficients $c = [c_0, c_1, c_2, c_3, c_4, c_5]$ for each pixel, and computes the pixel intensity I from a novel illumination direction $l = [l_x, l_y, l_z]^T$ as a biquadratic function:

$$I = c_0 l_x^2 + c_1 l_y^2 + c_2 l_x l_y + c_3 l_x + c_4 l_y + c_5. \quad (4)$$

The RTI either uses polynomial basis of order six or higher (e.g., $h(l) \triangleq [l_x^2, l_y^2, l_x l_y, l_x, l_y, 1]^T$ above) or uses hemispherical harmonics (HSH) basis $h(l)$ to generate a novel image from a new illumination direction l interactively specified by a user. For both cases, the pixel intensity is universally given by $I = h(l)^T c$. While the basis $h(l)$ is the same for all the image pixels, the coefficients c are pixel dependent. The coefficients can be computed from a set of precaptured K images under different known lighting directions, by least squares of an overdetermined (assume c has less order than K) linear system:

$$\begin{bmatrix} h(l_1) \\ h(l_2) \\ \vdots \\ h(l_K) \end{bmatrix} c = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_K \end{bmatrix}. \quad (5)$$

Over the past several years, the art conservation community has adopted RTI for close digital examination of artworks through relighting. RTI provides visually compelling ways to interactively explore surface relief and discover subtle surface features otherwise missing or indiscernible in ordinary photos or by direct visual inspection [24]. The free viewer software from Cultural Heritage Imaging (CHI) has been a boon to the field since it can exaggerate surfaces, pixel-by-pixel, to depict the topography more clearly and to compute estimates of surface normal vectors via photometric stereo or from the PTM interpolation equation itself.

However, these methods assume that lighting is infinitely far away from the object, a condition that cannot be easily achieved in practice due to power limitations of the light source and limited space around the object. The obvious solution is to capture images using near lighting, but these conditions result in nonuniform illumination artifacts made
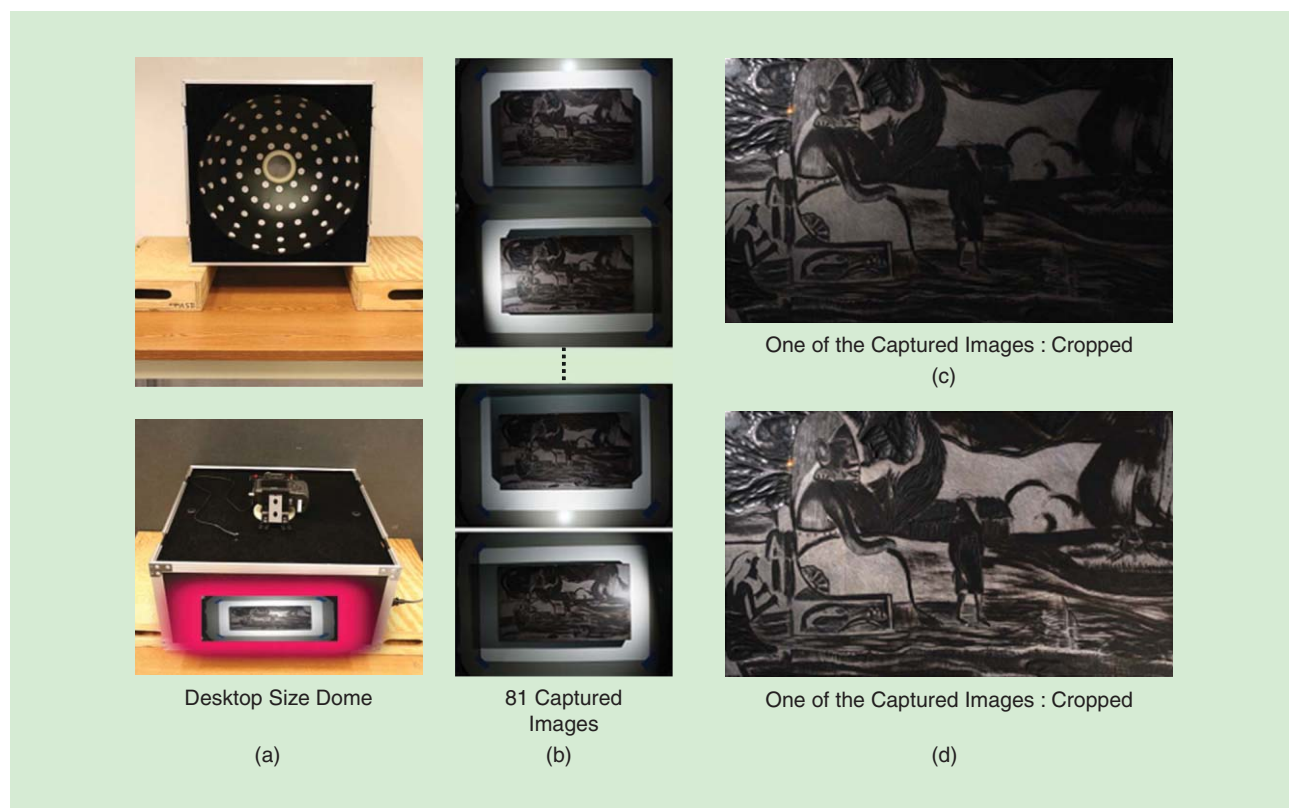
FIGURE 3. Relighting comparisons for a woodblock by Paul Gauguin (accession number 1940–91) housed at the Art Institute of Chicago. (a) The woodblock was inserted under a dome to capture (b) 81 images, each under different lighting. The light position used to compute light attenuation due to the distance squared fall-off. (c) The inverse of this attenuation was used to produce relit images with even illumination. (d) The corrected images look uniformly lit and more visually pleasing. (Images and figure caption used with permission from [25].)

worse as the distance between the light source and scene narrows. Not only does uneven illumination produce poor visualizations for relighting, but when these data are used for photometric stereo, systematic errors introduce curl to the surface normal estimations and make quantitative surface reconstructions difficult. An algorithm to correct for captures that violate far-light assumptions has recently become available [25] that creates uniformly lit images (Figure 3) and, moreover, accurate photometric stereo calculations for estimation of surface normals.

Macro X-ray methods

All macro X-ray methods in cultural heritage stem from Roentgen's discovery of X-rays in the late 19th century. X-radiography has been a staple of the field for decades and is still valuable in its original form: a contrast image formed from the absorption of X-rays by high-Z contrast elements, such as the lead associated with the pigment lead white. X-radiographs are routinely used by conservators and curators to characterize the method and style of painting and can be indicative of the artist's thought process when pentimenti are observed. Within the last decade, major advances have been made in interpreting X-radiographs by computational methods. Also, there has been an increasing use of macro tomographic methods, as well as the development of macro XRF scanning and macro XRD scanning that have transformed our views of cultural heritage objects.

> Some of the most exciting recent developments in cultural heritage analysis have involved XRF.

Computational processing of X-radiographs

Computational processing of X-radiographs has revolutionized the area of thread count and thread direction analysis for paintings on fabric supports [11] and now the chain lines are impressed into the paper by the wire mesh of the molds during fabrication [26]. The development of the method in [11] hinged on realizing that a Fourier transform to the observed alternating light and dark X-ray contrast patterns of a canvas could provide both thread count and thread direction data. Prior to this insight, threads were painstakingly counted by hand under magnification, and those counts were limited to only a few centimeters of a painting.

These new computational methods permit global analysis of the entire work. The overall pattern of threads has been shown to be very diagnostic for matching paintings to a single bolt of fabric and is now being used to date paintings. Furthermore, primary and secondary cusping in the canvas weave (scallop patterns caused by the stretching methods used to prepare canvases for old master paintings) becomes obvious after employing the computational algorithm, and not only can these patterns be used to match paintings to proximal regions of a bolt of cloth, their absence can be used to infer that a painting has been trimmed. More importantly, this method provides a way to match paintings at approximately the same period to a single bolt of cloth [11].

Macro XRF scanning

Some of the most exciting recent developments in cultural heritage analysis have involved XRF. The method involves using an X-ray source to ionize core electrons from atoms or ions. After the generation of inner-shell electron "holes," higher energy electrons "fall" into those holes, leading to the fluorescence of an X-ray. Because electron energy levels are quantized, the fluoresced X-rays are characteristic of the elements involved. Because inner-shell electrons are involved in these processes, the technique gives only elemental rather than chemical information. Therefore, for better and for worse, the spectra are simplified by their lack of chemical information. X-rays of different energies are attenuated different amounts when passing through a given material from emitter to detector. As a result, it is possible to make some statements regarding the depth of materials relative to one another in the layers of a painting, particularly when a model of that layered material can be computer simulated [27]. For example, the difference in intensity for an element's spectral response compared to theory can indicate how close to the surface of the object that element is, given information from the spectrum about which elements might be on top of it. Highly portable, rugged XRF point analyzers have made it possible to do qualitative (and under favorable conditions, semiquantitative) elemental analysis nondestructively on cultural heritage objects in a matter of minutes.

The true revolution in the field has resulted from taking XRF scanning methodologies and repurposing them with transportable macro XRF scanners [5], [28]. These scanners acquire a hyperspectral XRF data cube by scanning point by point in the x-y plane—each point in the x-y plane contains a full XRF spectrum. As with single point analysis, depth information can often be inferred based on relative X-ray intensities. As one might imagine, the amount of data involved in these cubes has demanded computational methods that can handle and mine this wealth of information [29]. Sometimes scanning a painting on a canvas support from behind can provide a better data set, due to different X-ray absorption characteristics, than scanning a painting from the front. The resulting information about elemental composition can be used to infer pigment maps and inferred information about relative depth can be used in combination with those maps to reconstruct paintings underneath overpaint [28]. For an artist such as van Gogh, whose work sold so poorly during his lifetime that he frequently reused his canvases and was supported by his brother, this XRF scanning technique has opened vast new areas of research (Figure 4).

Macro XRD scanning

As with point XRF analysis, point powder XRD has historically been invaluable in the characterization of artists' pigments. When performed in situ, the method does not require a sample and is considered nondestructive. Because the diffraction of X-rays requires a regular repeating array
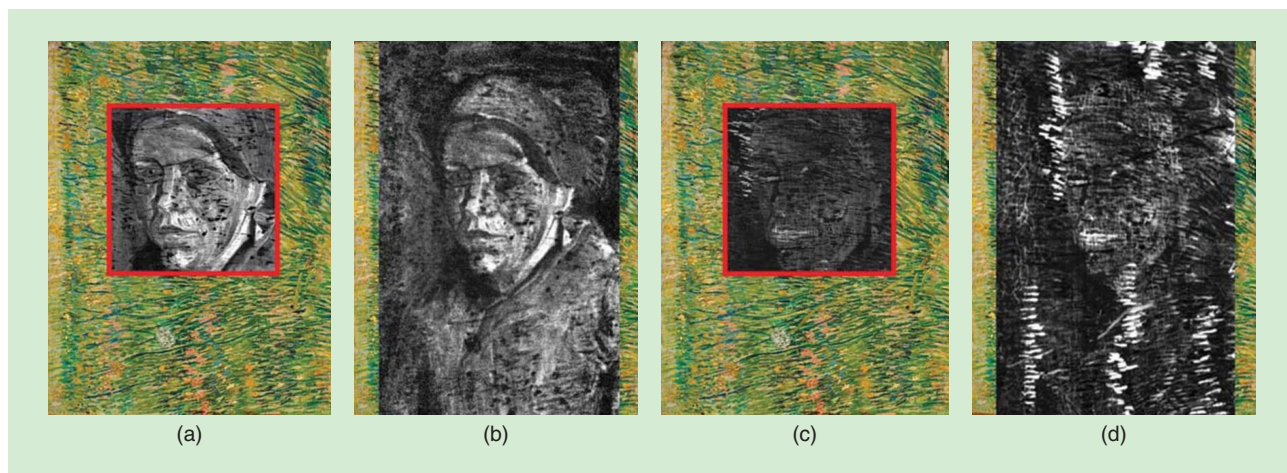
FIGURE 4. Elemental maps, obtained on Vincent van Gogh's *Patch of Grass*, showing the hidden portrait of a woman. (a) and (b) show the Sb distribution, while (c) and (d) show the Hg distribution. (a) and (c) were acquired with macro-XRF at a synchrotron source, while (b) and (d) are results of in situ measurements by means of Instrument B. (a) and (c) were acquired with a step size of 0.5 mm and 2 seconds dwell time in two days, while (b) and (d) were acquired with a step size of 1 mm and a dwell time of 5.1 seconds in six days. (Images and figure caption used with permission from [28].)

of electron density, the method requires microcrystallinity in the analyte. Thus, the technique cannot be used on amorphous materials or materials that do not diffract X-rays well, and in this regard it is inferior to XRF. However, because the diffraction pattern of a crystalline substance is essentially a fingerprint, it provides direct chemical information about the analyte, and in that regard it is superior to XRF. Because XRD typically requires greater photon flux than XRF, the method was more resistant to migration from synchrotrons to transportable scanning methodology. Fortunately, those problems are being solved [5]. In addition to providing positive chemical identification of materials present, XRD also offers the advantage that, because it requires higher energy X-rays, it provides greater depth penetration. Combined with XRF macro scanning data cubes and hyperspectral imaging cubes from the UV, vis, and IR, these techniques, operating synergistically, allow unprecedented insights into the composition of cultural heritage objects, with all of the attendant implications for art history and art conservation.

Conclusions

In this article, we surveyed how computational imaging has impacted five key areas of cultural heritage science. There are three key features that have resulted in these techniques making a significant impact on the cultural heritage community. The first is the proliferation in recent years of image sensing technology, which has spawned technological advances in new imaging modalities such as XRF, XRD, hyperspectral, etc. The second feature is that recent advances in these new imaging modalities has given accessibility to entirely new types of information latent within the artworks held by museums. The third feature is the ability to visualize information about artifacts intuitively in the form of images, which has made this information much more accessible and comprehensible to nonexperts.

Computational imaging of cultural heritage is opening up many new avenues for investigating the technical art history of objects and to assess the condition of works of art that will aid in their long-term preservation. There are several areas of computational imaging that have not been thoroughly explored on cultural heritage objects. Also compressive sensing and sparse imaging could significantly improve sensitivities especially for conditions where low light is necessary for light-sensitive materials and when increased imaging speeds are necessary for experiments that cannot be conducted in the public spaces of museums over days (as in macro X-ray scanning). Improved material databases with bidirectional reflectance distribution function data [30] could lead to advances in reconstruction algorithms that produce more accurate image archives and renderings. Scalability is another principal obstacle. For example, a comprehensive measurement of the chemical composition and spatial structure of layers of paint in an entire work of art could provide new and valuable tools for art historians and conservators.

Another important direction that has not been covered in this article is the dissemination, visualization, and display of the great body of visual information now being captured by museums and galleries around the world. For instance, augmented reality is projected to strongly impact the museum visitor's experience in coming decades. Finally, for the computational imaging field, it is important to note that artworks provide fantastic test scenes that can inspire researchers to push the envelope by providing new imaging and display techniques that can probe the complex light-material interactions inherent in so many works of art. In this regard, it is the hope that cultural heritage can serve as a catalyst for novel research in computational imaging.

Authors

Xiang Huang (xianghuang@gmail.com) is a postdoctoral researcher in the Mathematics and Computer Science Division of Argonne National Laboratory. His interestes are solving

challenging problems in computational photography, image processing, machine learning, and computer vision.

Erich Uffelman (uffelmane@wlu.edu) is the Bentley Professor of Chemistry at Washington and Lee (W&L) University. He obtained his Ph.D. degree from the California Institute of Technology as a National Science Foundation Predoctoral Fellow in 1991 and performed work as a National Institutes of Health postdoctoral fellow at Stanford until he joined the W&L faculty in 1993. He has had numerous grants and publications and was a winner of the 2009 State Council of Higher Education of Virginia Outstanding Faculty Award.

Oliver Cossair (ollie@eecs.northwestern.edu) is an assistant professor at Northwestern University within the Electrical Engineering and Computer Science Department. He directs the Computational Photography Laboratory, with research consisting of a diverse portfolio, ranging in topics from optics/photonics, computer graphics, computer vision, and image processing.

Marc Walton (marc.walton@northwestern.edu) currently runs the Northwestern University/Art Institute of Chicago Center for Scientific Studies in the Arts as its senior scientist, and he holds a joint appoint as a research associate professor in the Department of Materials Science and Engineering at Northwestern University. His research has focused primarily on the trade and manufacture of ancient objects, and he has published extensively in the fields of archaeological science and technical art history.

Aggelos K. Katsaggelos (aggk@eecs.northwestern.edu) is currently a professor and holds the Joseph Cummings chair in the Department of Electrical Engineering and Computer Science at Northwestern University. He has published extensively in the areas of multimedia signal processing and communications, computational imaging, and machine learning (including seven books, more than 250 journal papers, 500 conference papers, and 40 book chapters), and he is the holder of 25 international patents.

References

[1] A. Burroughs, *Art Criticism From a Laboratory*. Boston, MA: Little, Brown and Company, 1938.

[2] J. J. Rorimer, *Ultra-violet Rays and Their Use in the Examination of Works of Art*. New York: Metropolitan Museum of Art, 1931.

[3] J. R. J. van Asperen de Boer, "Reflectography of paintings using an infrared vidicon television system," *Stud. Conserv.*, vol. 14, no. 3, pp. 96–118, 1969.

[4] E. V. Sayre and H. N. Lechtman, "Neutron activation autoradiography of oil paintings," *Stud. Conserv.*, vol. 13, no. 4, pp. 161–185, Nov. 1968.

[5] S. Legrand, F. Vanmeert, G. V. D. Snickt, M. Alfeld, W. D. Nolf, J. Dik, and K. Janssens, "Examination of historical paintings by state-of-the-art hyperspectral imaging methods: from scanning infra-red spectroscopy to computed X-ray laminography," *Heritage Sci.*, vol. 2, no. 13, pp. 2–13, 2014.

[6] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D.Fulk, "The digital Michelangelo project: 3D scanning of large statues," in *Proc. 27th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '00*, New York, 2000, pp. 131–144.

[7] J. K. Delaney, P. Ricciardi, L. D. Glinsman, M. Facini, M. Thoury, M. Palmer, and E. R. de la Rie, "Use of imaging spectroscopy, fiber optic reflectance spectroscopy, and X-ray fluorescence to map and identify pigments in illuminated manuscripts," *Stud. Conserv.*, vol. 59, no. 2, pp. 91–101, Jan. 2014.

[8] T. Malzbender, D. Gelb, and H. Wolters, "Polynomial texture maps" in *Proc. 28th Annu. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH'01*, New York, 2001, pp. 519–528.

[9] O. Cossairt, X. Huang, N. Matsuda, H. Stratis, M. Broadway, J. Tumblin, G. Bearman, E. Doehne, A. Katsaggelos, and M. Walton, "Surface shape studies of the art of Paul Gauguin" in *Proc. Digital Heritage Granada Conf.*, Granada, 2015, pp. 13–20.

[10] C. Richard Johnson, J. Hendriks, E. Berezhnoy, I. J. Eugene Brevdo, S. M. Hughes, I. Daubechies, E. P. Jia Li, and J. Z. Wang, "Image processing for artist identification," *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 37–48, July 2008.

[11] D. H. Johnson, C. R. Johnson, Jr., and R. G. Erdmann, "Weave analysis of paintings on canvas from radiographs," *Signal Process.*, vol. 93, no. 3, pp. 527–540, Mar. 2013.

[12] D. M. Conover, J. K. Delaney, and M. H. Loew, "Automatic registration and mosaicking of technical images of old master paintings," *Appl. Phys. A: Mater. Sci. Process.*, vol. 119, no. 4, pp. 1567–1575, 2015.

[13] H. Liang, M. G. Cid, R. G. Cucu, G. M. Dobre, A. G. Podoleanu, J. Pedro, and D. Saunders, "En-face optical coherence tomography: A novel application of non-invasive imaging to art conservation," *Opt. Express*, vol. 13, no. 16, pp. 6133–6144, Aug. 2005.

[14] J. B. Jackson, M. Mourou, J. F. Whitaker, I. N. Duling III, S. L. Williamson, M. Menu, and G. A. Mourou, "Terahertz imaging for non-destructive evaluation of mural paintings," *Opt. Commun.*, vol. 281, no. 4, pp. 527–532, Feb. 2008.

[15] A. F. Chase, D. Z. Chase, J. F. Weishampel, J. B. Drake, R. L. Shrestha, K. C. Slatton, J. J. Awe, and W. E. Carter, "Airborne LiDAR, archaeology, and the ancient Maya landscape at Caracol, Belize," *J. Archaeol. Sci.*, vol. 38, no. 2, pp. 387–398, Feb. 2011.

[16] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," in *Proc. ACM SIGGRAPH*, New York, 2006, pp. 835–846.

[17] H. Liang, "Advances in multispectral and hyperspectral imaging for archaeology and art conservation," *Appl. Phys. A: Mater. Sci. Process.*, vol. 106, no. 2, pp. 309–323, Feb. 2012.

[18] M. T. Eismann, *Hyperspectral Remote Sensing*. Bellingham, WA: SPIE, 2012.

[19] F. Rosi, C. Miliani, R. Braun, R. Harig, D. Sali, B. G. Brunetti, and A. Sgamellotti, "Noninvasive analysis of paintings by mid-infrared hyperspectral imaging," *Angew. Chem. Int. Ed.*, vol. 52, no. 20, pp. 5258–5261, May 2013.

[20] A. Cesaratto, A. Nevin, G. Valentini, L. Brambilla, C. Castiglioni, L. Toniolo, M. Fratelli, and D. Comelli, "A novel classification method for multispectral imaging combined with portable Raman spectroscopy for the analysis of a painting by Vincent Van Gogh," *Appl. Spectrosc.*, vol. 67, no. 11, pp. 1234–1241, Nov. 2013.

[21] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.

[22] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.

[23] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Robust photometric stereo using sparse regression," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 318–325.

[24] M. Mudge, T. Malzbender, C. Schroer, and M. Lum, "New reflection transformation methods for rock arts and multiple viewpoint display," presented at the 7th Int. Symp. Virtual Reality, Archaeology and Cultural Heritage, Nicosia, Cyprus, Nov. 2006.

[25] X. Huang, M. Walton, G. Bearman, and O. Cossairt, "Near light correction for image relighting and 3D shape recovery," in *Proc. Digital Heritage Granada Conf.*, Granada, 2015, pp. 215–222.

[26] C. Richard Johnson, W. A. Sethares, M. H. Ellis, and S. Haqqi, "Hunting for paper moldmates among Rembrandt's prints," *IEEE Signal Processing Mag.*, vol. 32, no. 4, pp. 28–37, July 2015.

[27] L. de Viguerie, P. Walter, E. Laval, B. Mottin, and V. A. Sole, "Revealing the sfumato technique of Leonardo da Vinci by X-ray fluorescence spectroscopy," *Angew. Chem., Int. Ed.*, vol. 49, no. 35, pp. 6125–6128, Aug. 2010.

[28] M. Alfeld, K. Janssens, J. Dik, W. de Nolf, and G. van der Snickt, "Optimization of mobile scanning macro-XRF systems for the in situ investigation of historical paintings," *J. Anal. Atom. Spectrometry*, vol. 26, no. 5, pp. 899–909, 2011.

[29] M. Alfeld and K. Janssens, "Strategies for processing mega-pixel X-ray fluorescence hyperspectral data: A case study on a version of Caravaggio's painting Supper at Emmaus," *J. Anal. At. Spectrom.*, vol. 30, no. 3, pp. 777–789, 2015.

[30] J. C. Ashbaugh, R. S. Berns, B. A. Darling, and L. A. Taplin, "Artist material BRDF database for computer graphics rendering," in *Proc. IS and T/SID Color Imaging Conf.*, 2009, pp. 62–68.

SP

Michael Stengel and Marcus Magnor

Gaze-Contingent Computational Displays

Boosting perceptual fidelity
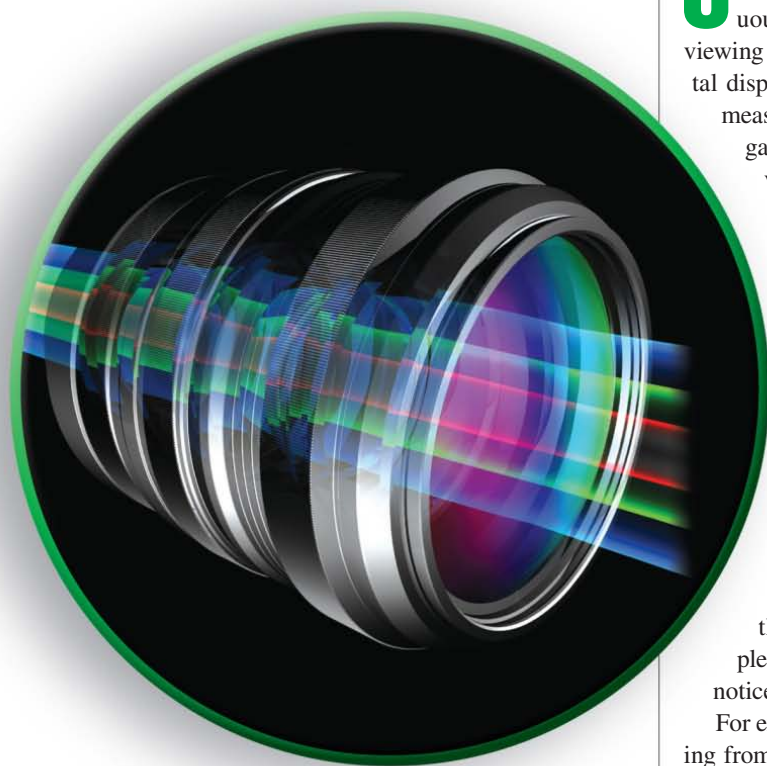
Contemporary digital displays feature multimillion pixels at ever-increasing refresh rates. Reality, on the other hand, provides us with a view of the world that is continuous in space and in time. The discrepancy between viewing the physical world and its sampled depiction on digital displays gives rise to perceptual quality degradation. By measuring or estimating where we look, a new breed of gaze-contingent algorithms aims to exploit the way we visually perceive digital images and videos to remedy visible artifacts. In this article, we provide an overview of recent developments in computational display algorithms that enhance perceived visual quality of conventional video footage when viewed on commodity monitors, projectors, or head-mounted displays (HMDs).

Introduction

Display technology is advancing at a breathtaking pace. Driven by consumer demand, screen size, resolution, contrast, and refresh rates are growing bigger, faster, and higher almost on a weekly basis. Nevertheless, the fundamental difference between the continuous physical world and its digitally sampled and displayed image still gives rise to perceptually noticeable quality degradation.

For example, while TV screen resolution has been increasing from National Television System Committee/phase alternating line to high-definition (HD) and on to full-HD today, physical screen size has grown in step, leaving the number of pixels per inch (PPI) almost unchanged on commodity hardware. Common full-HD screens result in a narrow vertical field-of-view of 18° when matched to eye acuity, so that single pixels can no longer be perceived (see Table 1). High-end screens with 2,880p (5k) or 4,320p (8k) vertical resolution are required to widen the field of view for a more immersive viewing experience. Exceptions are smartphone displays that do feature high pixel densities of 500–800 PPI to enable clear

©ISTOCKPHOTO.COM/YAKOBCHUK

IEEE SIGNAL PROCESSING MAGAZINE | September 2016 | 139

Table 1. An overview of common display device types.

Display Type	Refresh Rate	Vertical Resolution	Pixels per inch	Lossless Viewing Distance	Lossless FOV
Smartphone	60 Hz	1,920p/5.5 in–3,840p/5.5 in	400.5–801.0	> 8.5–4.3 in	< 32–64°
Tablet	60 Hz	1,200p/7 in–1,824p/12.3 in	323.5–267.0	> 10.6–12.9 in	< 20–30.4°
Monitor	60–120 Hz	1,080p/23 in–2,880p/27 in	123.1–217.6	> 35.9–15.8 in	< 18–48°
TV set	60–600 Hz	1,080p/55 in–4,320p/85 in	40.0–103.6	> 85.8–33.2 in	< 18–72°
Projector	24–1,000 Hz	1,080p/125 in–2,160p/125 in	17.6–35.2	> 195.1–97.5 in	< 18–36°
HMD	60–90 Hz	800p/7 in–1,200p/7 in	215.6–352.9	> 15.9–9.79 in	< 13.3–20°

Each column contains values for a middle-class device and a high-end device. Perceptually lossless minimum viewing distance and maximum lossless field of view (FOV) are given for a person with average vision (20/20 Snellen) so that the perceived resolution matches foveal acuity.

readability of small-scale text and playback of high-resolution videos. This implies, however, that more and more pixels must be rendered on resource-strapped mobile devices. Incongruities also exist in the temporal domain between digital video recording and display capabilities. While TV display refresh rates today commonly match or exceed 60 Hz, standard video acquisition frame rates still hover between 24 and 30 frames/second (fps). The discrepancies between the physical world and its digital representation, as well as the mismatch in acquisition versus display capabilities and displays versus human visual perception, lead to noticeable artifacts.

If it is known, or can be reliably estimated, how human vision perceives digital images at any one time, gaze-contingent display methods are able to make use of a number of perceptual strategies to improve perceived visual quality. In addition, gaze contingency allows allocating computational resources on the fly to image regions that are perceptually relevant for the current gaze direction. While gaze-contingent display approaches have been proposed before, only recently have eye-tracking hardware, saliency estimation methods, and graphics hardware become sufficiently fast, robust and affordable to allow for incorporating advanced gaze-aware methods in mass-market devices. This article highlights recent examples of gaze-contingent computational display approaches that enhance perceived visual fidelity of common, consumer-market display technologies.

Modeling human vision

Gaze-contingent displays exploit abilities and limitations of the human visual system (HVS). Although correlations between the eye and the visual cortex are not yet fully understood, models of the HVS enable us to conservatively express some important features of human vision.

Visual acuity provides an estimate of the smallest visual detail that the HVS is spatially able to perceive. The acuity follows approximately the distribution of cones and rods in the retina. It reaches its highest value, therefore, in the foveal region, which can be modeled by a central disc with a radius of two degrees visual angle. Acuity falls off rapidly with eccentricity in the periphery. Therefore, visual acuity is estimated by

a function over the eccentricity or, in other words, the distance to the fovea given in degrees visual angle.

The acuity value is commonly given either as the Snellen value or a minimum angle of resolution (MAR) value [1]. For healthy young adults, the common highest visual acuity is defined as 20/20 Snellen or, equivalently, 1 minute of arc in terms of MAR for the foveal region [1]. The acuity limit is primarily reasoned by the spacing of photoreceptors in the retina. However, due to the large variability between human eyes, estimating the distribution of rods and cones across the retina is difficult. Additionally, studies have shown that, at eccentricities greater than two degrees, the acuity is worse than what can be predicted from cone spacing [1]. Consequently, visual acuity cannot be determined from the distribution of the photoreceptors on the retina only.

The psychophysical model of Aubert and Foerster from 1857 is a well-established model for low-level vision tasks [1]. It states that the minimum discernible angular size increases roughly linearly with eccentricity for the first 20−30°; it then rises more rapidly [1]. Due to its simplicity and conservative approximation of the real acuity, the model of Aubert and Foerster is still commonly used. However, the linear model of Aubert and Foerster does not do justice to the full complexity of the HVS, as peripheral vision is not a scaled-down version of the foveal area [1]. Eye adaptations in very bright and dark areas as well as motions of the eye also influence the amount of detail perceived. Therefore, sophisticated acuity models must include means to deal with additional vision features. Color vision is another aspect affected by the distribution of rods and cones. Although color can be still perceived in nonfoveal vision if the stimulus is large enough, retinal performance falls off linearly until 20–30° in periphery for color discrimination and many other visual tasks as well [1].

Vision models for gaze-contingent displays may not only consider visual acuity, brightness adaptation, and color vision but also the measureable movements of the human eyeball. The most important motion abilities of the eye are saccades, the motion when jumping from one object of interest to another, and fixations, which give humans the ability to directly gaze intentionally into a certain direction. Both eyes are commonly

directed at the same scene object, but, reasoned by the physical disparity of the eyes, the difference of the per-eye gaze vectors can be quite large, e.g., when one is looking at an object close-by. This vergence motion allows humans to perceive depth in nature and in stereoscopic images.

The unconsciously triggered tracking reflex when a moving object attracts gaze is called *smooth-pursuit eye motion*. On a screen, smooth-pursuit eye motion may affect the acuity function, assuming that the frame render duration is above the retinal integration time. This brightness-level dependent value describes the average duration of collecting photons before transmitting the visual information to the higher levels of the visual system [2].

Accomodation is another form of eye motion and describes the mechanical ability of the eye to compress or relax the lens of the eye so that objects in the foveal spotlight are projected clearly. In other words, accomodation describes the natural counterpart of adjusting a camera lens so that an object in the scene gets into focus. In-focus objects can be described by a depth-of-field plane within the focus distance to the eye and are clearly projected onto the retina, whereby out-of-focus objects appear blurred.

Adaptation is the time-dependent process that adjusts our eye to the environmental lighting situation. Although adaptation enables us to see over seven orders of magnitude of brightness intensities, we are not able to see equally well at all intensity levels. The visual acuity for details and color perception is reduced with dimming of light. In other words acuity is lower at scotopic levels (cones) of illumination than at photopic levels (rods). It is, therefore, difficult to read a book in a dim room without additional light. During daytime, contrast sensitivity is lower, but sharp vision and color vision work very well.

Gaze detection and prediction

Gaze-contingent approaches require knowledge of where on the screen the user is looking at any time. Depending on the application, gaze direction must be known to about one degree accuracy and updated with at least 50 Hz, while total latency may not exceed 60 milliseconds [3]. In modern eye tracking systems, video cameras are frequently mounted either close to the eyes or at a distance to record the pupils and corneal reflections. If the user is allowed to move the head, it, too, needs to be tracked. The cameras must be calibrated and the video streams processed in real time to continuously determine gaze direction. A comprehensive guide to eye tracking methods has recently been compiled by Holmqvist et al. [4].

Active eye-tracking methods require additional hardware and calibration of the hardware for each user. The calibration process may hamper deployment of gaze-contingent applications. Just recently, researchers began to loosen calibration

While gaze-contingent display approaches have been proposed before, only recently have eye-tracking hardware, saliency estimation methods, and graphics hardware become sufficiently fast, robust and affordable to allow for incorporating advanced gaze-aware methods in mass-market devices.

constraints to enable eye tracking in more general scenarios. Cazzato et al. derive pupil information and head pose at the same time to enable gaze tracking while both the eye-tracking camera and the tracked person can move independently [5].

Most image and video content, on the other hand, is being viewed in quite a predictable fashion. Our visual attention, i.e., our gaze, is subconsciously drawn to certain image or video regions while other parts of an image are merely being glanced over [6].

In the past two decades, numerous algorithms have been devised to predict visual saliency of local image features. In 1998, Itti et al. proposed a visual attention estimator inspired by the early vision processes of our visual cortex [7]. Based on a set of predefined center-surround operations and linear filters, image features are computed on multiple image resolution scales and combined into one saliency map, from which a neural net selects and ranks the most prominent features. The predictor of Itti et al. is still competitive to other state-of-the-art methods today.

Over the ensuing years numerous improvements, as well as alternative approaches, to visual saliency prediction have been proposed [8]. One of the most successful current methods is based on image decompositing and contrast measures akin to high-dimensional Gaussian filtering [9]. The resulting saliency map is claimed to successfully separate fore- and background information on a per-pixel basis. Alternatively, visual saliency may be learned directly from large amounts of eye-tracking data [10]. Benchmark tests of various computational models of visual saliency indicate, however, that there is no single method equally suitable for all types of scenes and situations [11].

In comparison to work on image saliency, there exists significantly less research into video saliency, although motion and moving objects are known to be strong attractors of visual attention. Little research has also been invested so far into saliency prediction in the outer peripheral field of view, i.e., into the kind of visual stimulus that is able to elicit long-range saccades.

Because local image saliency determines to a large extent where people look, innumerable painters throughout the centuries have experimented with visually salient cues to actively guide the eye of the beholder. On today's gaze-contingent displays, gaze can be actively steered around using imperceptible local contrast modulations in the peripheral field of view [12]. By turning the modulation stimulus off immediately after the onset of the elicited saccade, saccadic masking prevents the user from being consciously aware of the stimulus. The modulation stimulus is never perceived by the user's foveal vision.

Gaze-contingent techniques

The notion of gaze-contingent display devices dates back at least two decades. Excellent earlier review articles on gaze-contingent techniques and applications include those of Reingold [13] and Duchowski [14].
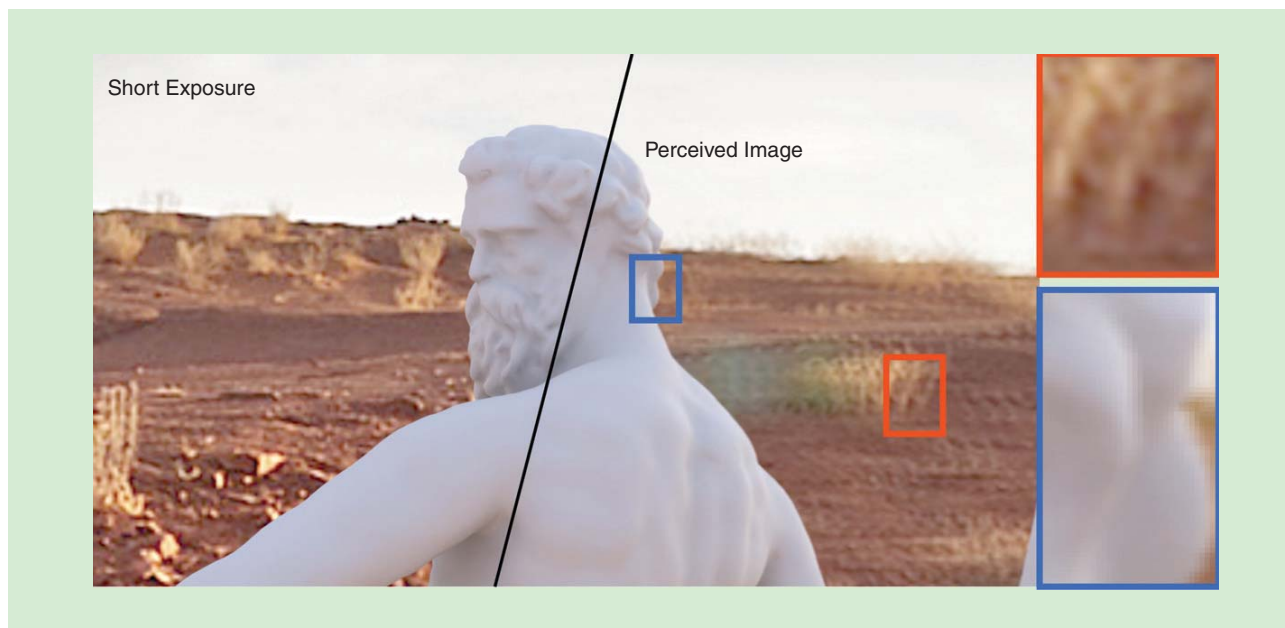
FIGURE 1. Judder due to camera motion. If exposure times are short and frame rates too low for image-plane motion, judder (ghosting) artifacts are perceived in the background on the screen.

As early as 1990, Levoy and Whittaker proposed a gaze-contingent approach to render volume data sets according to view direction [15]. Motivated by the limited computational resources of the time, a ray tracer was described, the local ray density of which varies depending on the angle between volume region and gaze direction while an eye tracker continuously measures gaze direction in real time.

In their 1996 article, Ohshima et al. employ level-of-detail rendering depending on gaze to interact with multiple objects in a virtual environment [16]. Besides angular distance from gaze direction, the authors take into account additional perceptual clues from kinetic and binocular vision to adapt the rendered level of detail to what can and cannot be perceived. In contrast, Luebke et al. simplify three-dimensional (3-D) geometry meshes directly in accordance with gaze [17]. To remain visually imperceptible, the degree of mesh simplification is controlled by gaze direction as determined via eye tracking. Along similar lines, Murphy and Duchowski propose a nonisotropic, level-of-detail-rendering approach for geometry meshes based on a user-study-derived 3-D spatial degradation function [18]. Their eye-tracking system was probably also the first designed for an HMD to be used in virtual reality applications.

A real-time simulator of glaucoma and other ophthalmic degradations of the field of view has been presented by Perry and Geisler [19]. Their system accepts conventional video footage as input and filters it with a predefined low-pass kernel centered on current gaze direction at 60 fps. Parkhurst and Niebur investigate how gaze-contingent level-of-detail rendering affects our ability to detect and localize objects [20]. Their experiments demonstrate that object detail significantly influences the speed with which we are able to perform different tasks. Based on a user study, Levi et al. predict that, by using gaze-contingent

approaches, rendering cost needs to increase only linearly (instead of quadratically) with increasing display field of view, as long as angular resolution remains constant [21]. By putting the main emphasis on rendering acceleration, significant speed ups are achieved using a gaze-contingent rendering approach [22].

Boosting temporal fidelity

In the physical world, our individual gaze determines if, how, and where we perceive blur. Our blur perception in the real world can differ distinctly from camera-recorded motion blur. While watching live-action shots on screen, we may notice annoying ghosting, judder, or edge-banding artifacts. When observing a dynamic scene in real life, our eyes automatically follow and track moving features to hold the object steady on the foveal region of our retina, allowing us to resolve details. As noted earlier, this automatic, involuntary, and effortless behavior of our human visual system is known as *smooth-pursuit eye movement*.

By visually tracking the moving foreground with our gaze, our peripheral vision perceives the nontracked background as blurred by motion. A camera tracking the moving foreground also records the background in motion-blurred fashion. When displayed on screen, however, the background can appear distinctly different and annoyingly distracting from the foreground. If frame rate and exposure time are not correctly adapted to apparent scene motion in the image plane, aliasing is introduced in the background, referred to as *judder* (Figure 1). These judder artifacts become especially apparent when viewed on wide-angle displays because spurious high-frequency details in the moving background lead to the perception of discontinuous, jumpy motion by our peripheral vision. User studies indicate that missing or insufficient motion blur affects gaze behavior in such a way that our visual attention is involuntarily drawn

toward incorrectly sharp and detailed image regions. The discontinous motion of the background distracts our visual attention from the tracked foreground because our peripheral vision is especially sensitive to movements and expects consistent, smooth motion (Figure 2).

To remedy the effects of temporal aliasing, Nvidia recently released its G-Sync technique, which adapts display-refresh rate to the processing time of each rendered frame (www.geforce.com/hardware/technology/g-sync). The technique reduces some stutter motion artifacts by avoiding repeated display of the same frame, allowing for smoother motion perception in games.

Similar to judder, hold-type blur is another form of aliasing that occurs when objects move faster in image space than they can be captured by the camera. Motivated by the Nyquist theorem, object velocities above one pixel per frame in the image space introduce aliasing. As a result, at 24-Hz capture frame rate and HD resolution, the sampling criterion is violated already for slow-moving objects. Higher frame-rate videos with a typical frame rate of 48 or 60 Hz reduce recorded motion blur and hold-type blur, yielding sharper perceived images. High frame-rate movie cameras are currently being explored by movie directors (e.g., in Peter Jackson's *The Hobbit* series), and specialized upsampling hardware is being integrated into standard TV sets. High frame-rate video, however, is not always practical or beneficial. The video must be recorded at shorter exposure times and because there is always some dead time to store each frame (or to open the shutter), only about 60% of the time interval between successive frames is being captured. Again, temporal aliasing may be the result. Some viewers even report perceiving a distracting speedup of the video (www.pcmag.com/article2/0,2817,2379206,00.asp).

Even at 60-Hz capture frame rate, fast-moving objects cannot be recorded free of aliasing. If eye movement is known or can be reliably predicted—e.g., via passive, saliency-based gaze prediction—the minimal sampling rate (or maximal motion) between two frames can be estimated. Specialized displays that support multiflash protocols are able to reduce artifact visibility by showing each video frame multiple times [23]. For conventional displays, however, only appropriate frame prefiltering is able to yield convincing results [2]. To this end, the original video sequence must first be temporally upsampled. For frame interpolation, robust optical-flow-based algorithms may be used [24]. Alternatively, perceptually smooth results may also be obtained by directly making use of predicted gaze [2], [25]. In both cases and for best results, exposure time of the input sequence should be low to minimize motion blur.

When we know the eyes' gaze path p during smooth-pursuit eye movement, the correct retinal image R can be computed for any point in time and any desired integration time T_R by integrating the upsampled video sequence along the path

$$R_i(x) := \int_{t_i}^{t_i+T_R} S(x + p(t), t)\, dt, \qquad (1)$$
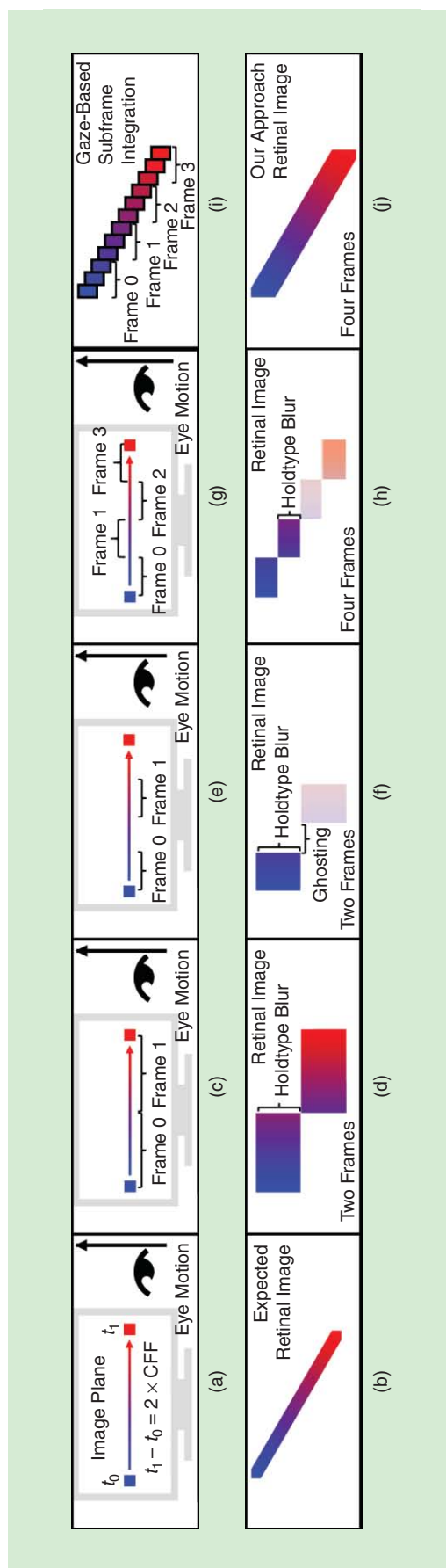
FIGURE 2. Temporal aliasing. (a) An object moves horizontally while eye gaze moves upward. (b) In the physical world, the perceived motion of the object is a continuous, diagonal movement. (c) Long-exposed frames taken with a static camera record the object as horizontal streaks. (d) When these streaks are displayed on a screen, the viewer perceives them as rectangles. (e) Shorter exposure times at fixed-frame rate lead to temporal gaps. (f) On screen, the object is then perceived as a number of discrete, disconnected rectangles, causing perceptually annoying judder. (g) Higher frame rates alleviate the problem but (h) do not eliminate it. By taking smooth-pursuit eye movement into account, (i) the video frames can be optimized along the known gaze path prior to display to achieve (j) a continuous, authentic motion perception [2].

(a)　　　　　　　　　(b)　　　　　　　　　(c)
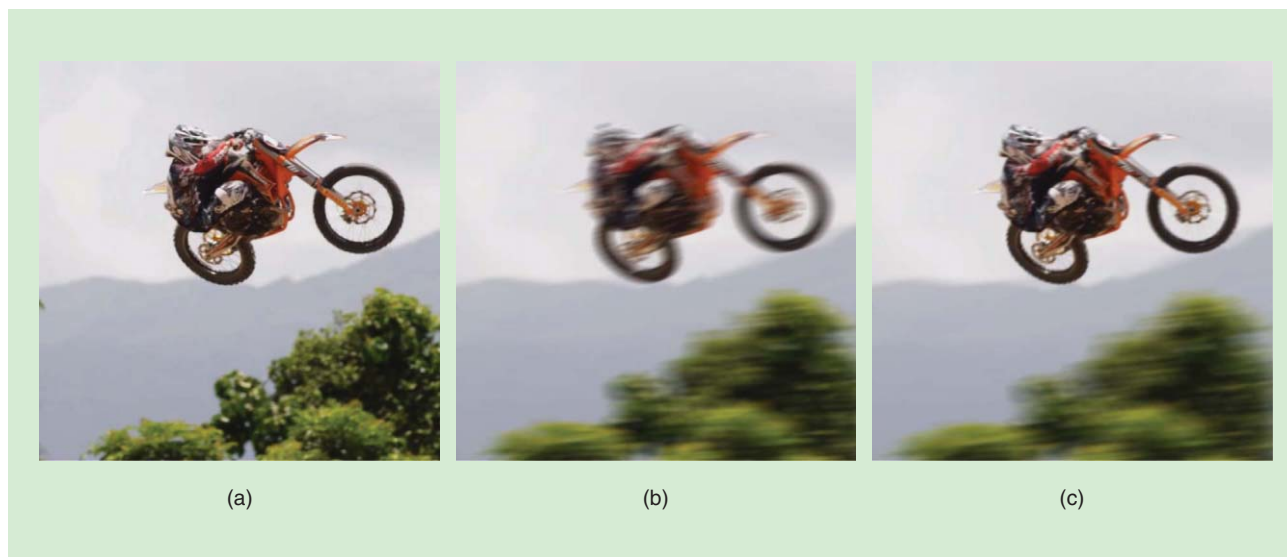
FIGURE 3. Judder cancelation: the original 60-Hz input video is first upsampled to 3,000 Hz, then filtered and downsampled again to 60 Hz to simulate different exposure times. (a) A frame from the original sequence. (b) Conventional long-exposure simulation by blending multiple interpolated frames; note that the entire image is blurred due to camera motion. (c) Gaze-adaptive filtering keeps the foreground sharp while the background is correctly blurred, consistent with smooth-pursuit eye movement of the foreground. (Original footage copyright © 2013 RED Digital Cinema. All rights reserved.).

where t is time, x is a screen coordinate, and $S(x + p(t), t)$ is irradiance received by the retina. This corresponds to translating each frame in the opposite gaze path direction, compensating eye motion, and setting the spatial components along the path to zero: by integrating along the temporal axis and translating the frames back to their original position, the frames are filtered according to the gaze path. This filter operation may be applied in real time for adjusting the integration time if the temporally upsampled video frames can be accessed quickly enough. The result is a consistently filtered, temporally downsampled video sequence that can be displayed on any conventional display or projector (Figure 3).

Besides correct prefiltering of low as well as high frame-rate video, the approach can also be applied to simulate virtually any shutter configuration. Rolling shutter, focal plane shutter, or artistic shutter effects, e.g., shutter times longer than one frame, can be created this way from the same footage during postprocessing. Conversely, still images can be infused with localized motion blur to convey information about scene motion.

Boosting perceived resolution

Big-size screens featuring 60+ in diagonals and 4k pixels per line are entering the consumer market. HMDs that are the driving force behind virtual reality applications, on the other hand, must be small in size but at the same time cover a very wide field of view. Even with state-of-the-art 350 PPI screens, at a natural field of view current HMD displays are still an order of magnitude away from eye acuity (Table 1), exacerbated even more by significant pixel magnification in the central region of lens-based HMDs.

Recently, apparent-display-resolution enhancement techniques have been proposed to provide super-resolution on high refresh-rate displays [26]. By exploiting how the HVS observes and processes moving content, these approaches are able to boost perceived resolution beyond the actual physical resolution of the display. With active-matrix organic light-emitting diode technology, refresh rates in excess of 1,000 Hz are achievable, allowing for a six-fold increase in apparent display resolution. In essence, apparent-display-resolution enhancement allows trading screen refresh rate for perceived resolution as long as the user's gaze continuously and predictably tracks moving foreground via smooth-pursuit eye movement.

Didyk et al. have shown that retinal integration of high-frame-rate, low-resolution images results in an increase in perceived resolution if the frames are displayed above the critical flicker frequency [26]. To perceive an object at higher than screen resolution, the eyes must smoothly pursue the object while it is moving across the screen. Smooth-pursuit eye movement can already be evoked by simply moving a static image at constant velocity. In videos, the effect is achieved by exploiting the movement of foreground objects [27]. Apparent-display-resolution enhancement is governed by two constraints: high display-refresh rates are necessary to achieve high contrast and strong resolution enhancement, and the motion must be predominantly in a diagonal direction to achieve resolution enhancement both horizontally and vertically. The more foreground object motion deviates from the diagonal direction on the pixel grid, the less pronounced the resolution-enhancement effect is in the direction orthogonal to the motion.

Apparent-display-resolution enhancement enables experiencing a standard low-frame-rate, high-resolution video at full resolution on a much lower resolution, yet high refresh-rate display [25]. Based on the assumption that our gaze follows the most salient regions of the sequence, the salient foreground regions are determined in a preprocess. The video frames are then continuously, unnoticeably shifted in such a way that, in
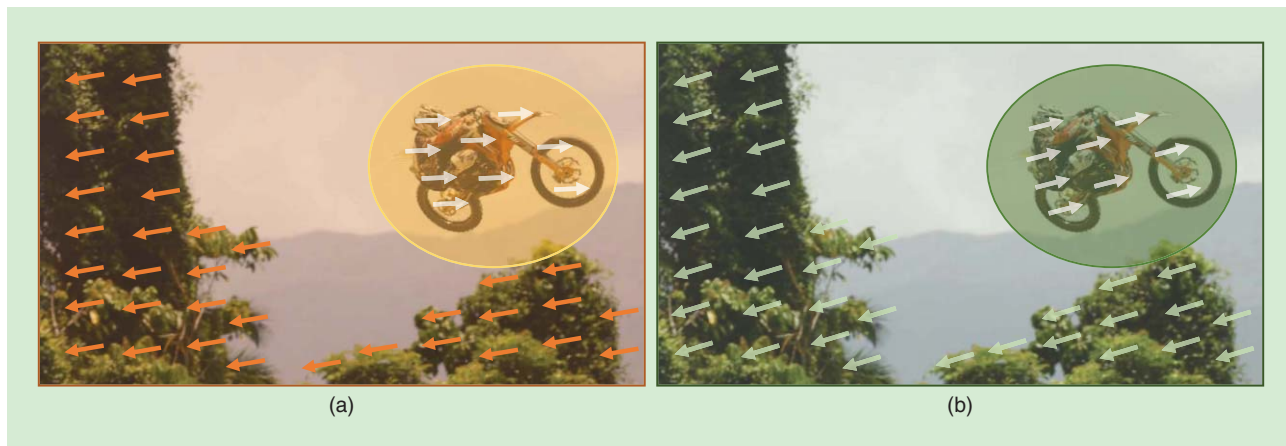
FIGURE 4. Salient-region motion optimization. (a) In the original sequence, the salient region (the motorbike driver) shows only horizontal movement; in this case, apparent-display-resolution enhancement can be applied in a single direction only. (b) After motion optimization, the salient region moves diagonally across the screen, which is ideal for apparent-display-resolution enhancement. (Original footage copyright © 2013 RED Digital Cinema. All rights reserved.)

combination with the postulated smooth-pursuit eye movement, the tracked foreground moves in a diagonal direction, facilitating the exploitation of the resolution enhancement effect (Figure 4):

$$\mathbf{W}(\mathbf{I}_L) - \mathbf{T}(\mathbf{I}_H) = 0, \qquad (2)$$

where $\mathbf{I}_H$ is the vector of all high-resolution frames, $\mathbf{I}_L$ is the vector of low-resolution subframes, $\mathbf{W}$ is a matrix weighting each subframe, and $\mathbf{T}$ is the vector of applied transformations per subframe.

To find the optimal transformations, the optimization is based on energy minimization:

$$E = E_{vel} + E_{smooth} + E_{imp}. \qquad (3)$$

The energy E incorporates saliency by the term E_{imp} (to focus the optimization on regions of interest for a viewer), smoothness

by the term E_{smooth} (to support smooth-pursuit eye motion and to prevent flickering), and resemblance to the optimal direction and velocity by the term E_{vel} (to provide the best possible resolution enhancement effect). Trajectory optimization is performed in matter of seconds for a complete frame sequence.

For low frame-rate videos to facilitate smooth-pursuit eye movement and to avoid simply duplicating frames, frame rate is upsampled by interpolation along the salient region's motion path, boosting the apparent resolution enhancement effect most effectively (Figure 5). By continuously, unnoticeably shifting the video frames around, the salient region can be perceived at much higher resolution than the display physically provides. This approach allows for apparent resolution enhancement even for scenes that originally do not contain any movement or for which optical-flow computation is difficult or impossible. HMDs especially make high demands on spatial resolution and may, therefore, significantly benefit from this technique. The approach takes about 30 seconds per
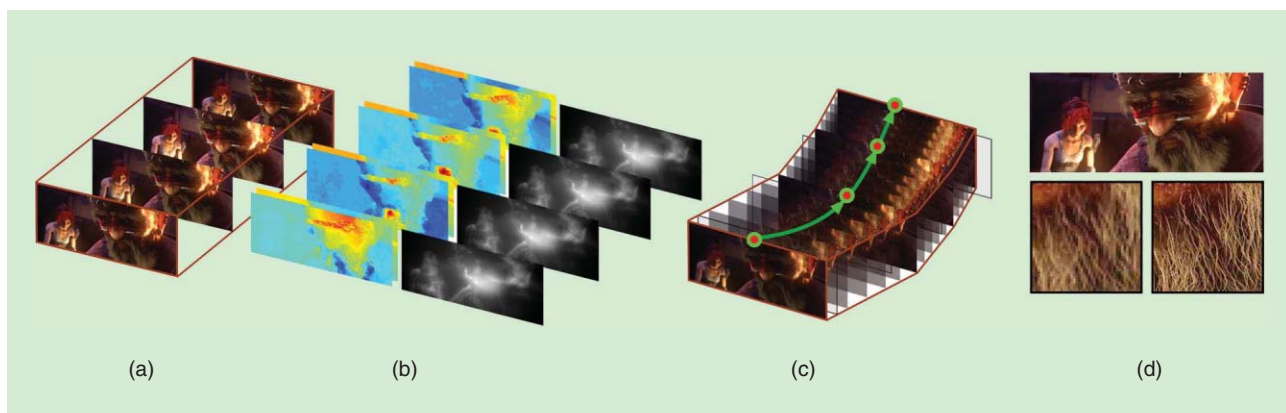
FIGURE 5. Applying apparent-display-resolution enhancement to general footage [25]. (a) Given a standard frame-rate video as input, (b) optical-flow and saliency maps are computed to (c) temporally upsample and shift the video along a smooth, optimized trajectory. (d) When displayed on a high refresh-rate display, the perceived resolution of the salient region is much higher than the physical resolution of the display.
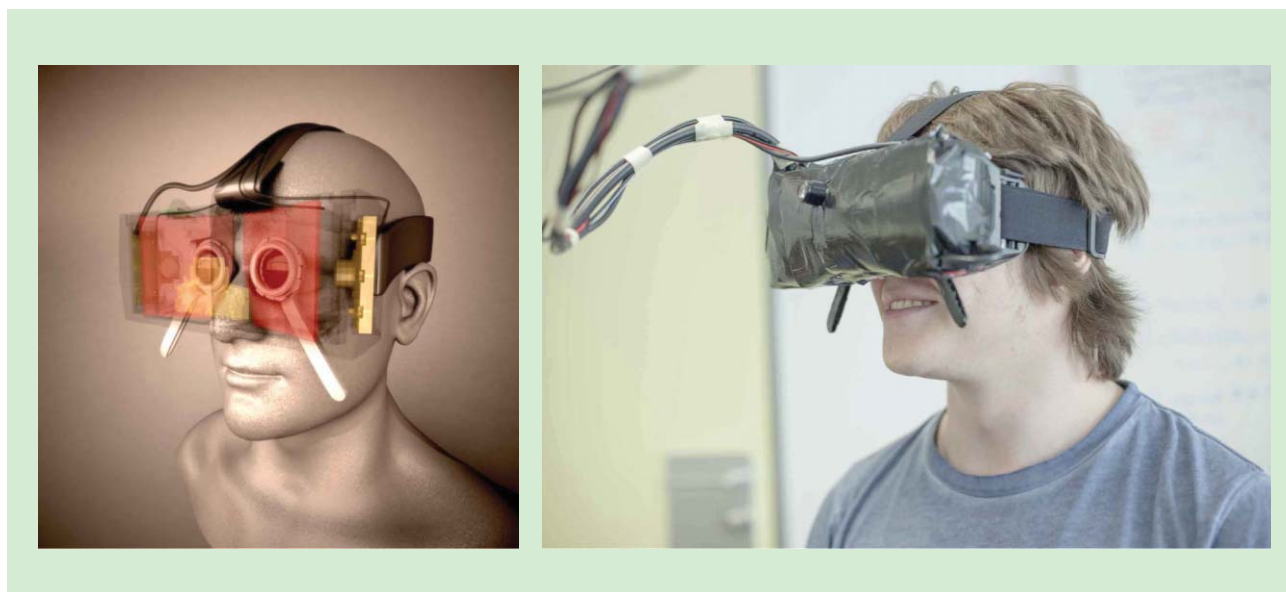
FIGURE 6. An eye-tracking HMD: (a) a mirror-based eye tracker concept and (b) an HMD prototype [28].

output frame (at 4k input) and is, therefore, not yet real-time capable. However, an improved implementation running on the graphics processing unit may enable real-time apparent-display-resolution enhancement based on active gaze input.

Gaze-contingent head-mounted displays

Being able to detect and to adapt to gaze direction facilitates many new ways to enhance digital displays. Especially for HMDs, knowing in real time where the user is looking enables providing a much-improved viewing experience.

By integrating an eye tracker into an HMD (Figure 6), the displayed frames can be directly adapted to the user's current gaze direction [28]. One intriguing application made possible by real-time eye tracking is to simulate the accommodation reflex [29]. Accommodation allows us to focus on objects at arbitrary distances. Objects in front of or behind the focal plane appear naturally blurred. Using a gaze-contingent HMD, this effect can be convincingly simulated because in HMD applications scene-depth information is typically available to facilitate stereoscopic vision [Figure 7(a) and (b)]. To find the correct plane of focus, a ray is cast along the gaze direction, and its intersection with the scene establishes focal distance. Depth-of-field rendering is then applied to create an authentic visual experience of the entire scene. Binocular eye tracking can additionally be exploited to measure eye vergence for very nearby objects.

Another application for eye tracking in HMDs is foveated rendering [22]. The rapid fall off in acuity from our foveal to our peripheral field of vision is exploited to allocate rendering and video processing resources more efficiently.

(a)   (b)   (c)   (d)
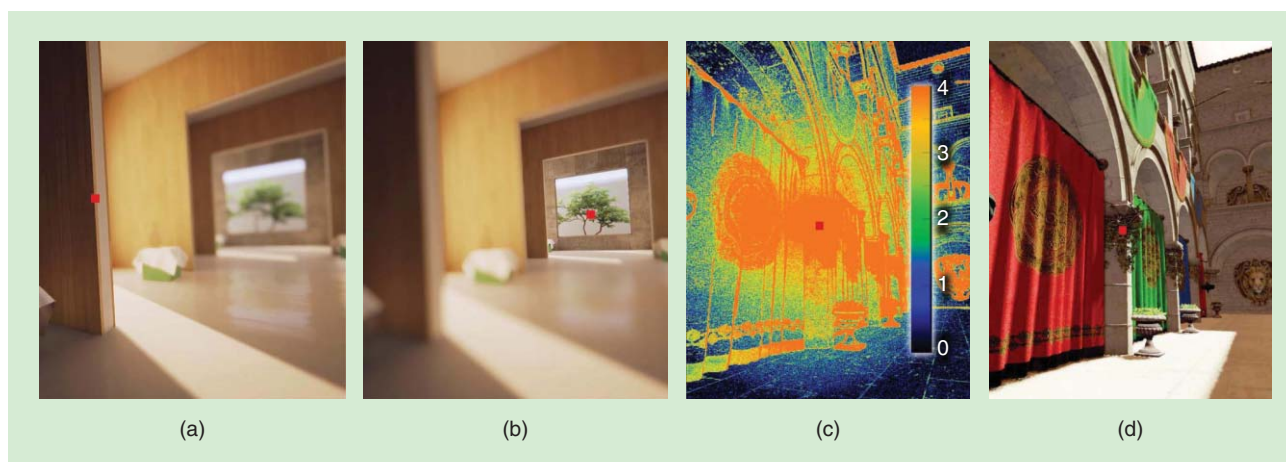
FIGURE 7. Gaze-contingent rendering. By knowing where the user is looking, the effect of our eyes accommodating to different scene depths can be reproduced: (a) and (b) near/far accommodation on an image; the red marker shows gaze position. In addition, (c) gaze-contingent sampling allows one to exploit the limitations of visual perception such as (d) the fall off in acuity from foveal to peripheral vision for faster rendering. The color-coding in (c) represents the number of shaded pixels in a 2×2 neighborhood [22]. [The models shown have been made freely available by Epic Games International (Room) and Crytek GmbH (Sponza Atrium).]
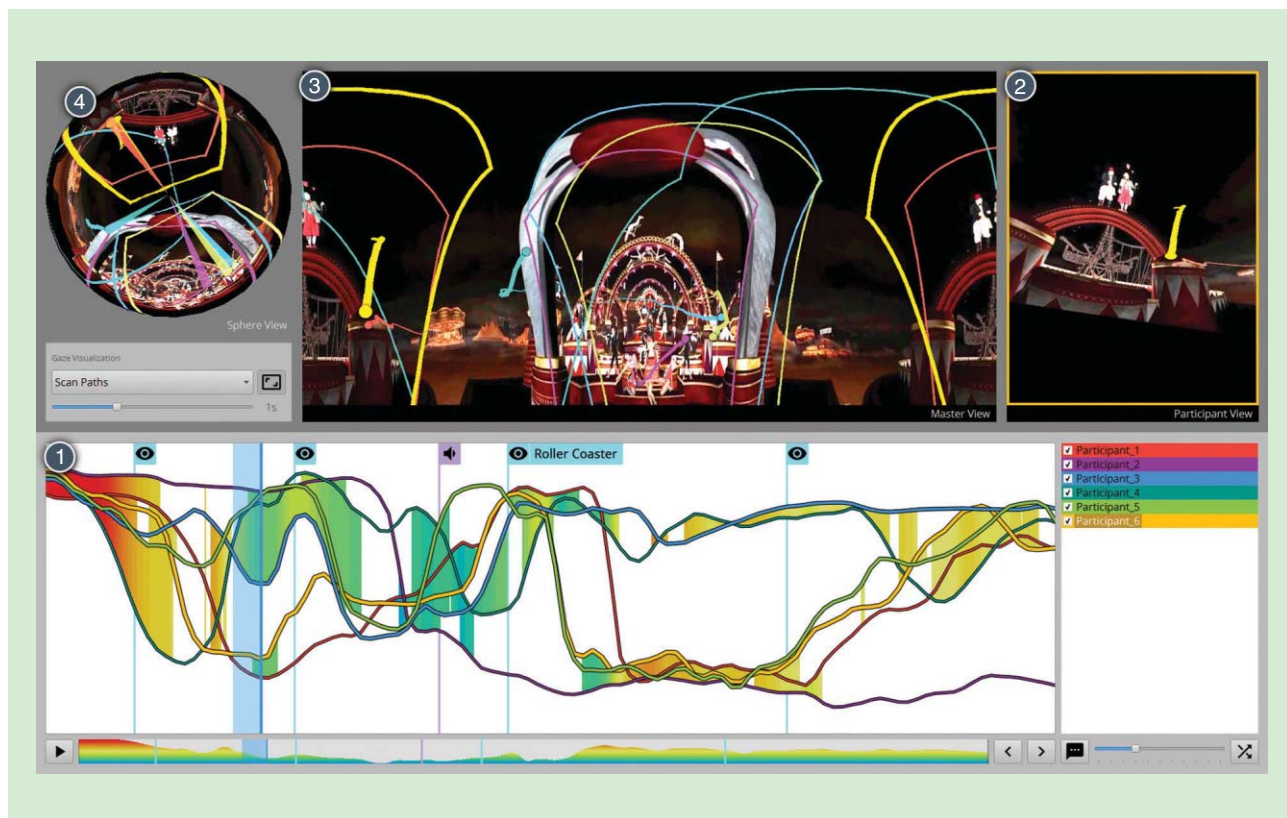
FIGURE 8. A user interface to analyze the viewing behavior of immersive live-action videos [30]. At top, color-coded frames indicate the current field of view of multiple users. Below, the users' scan paths allow analysis of gaze direction of all users over time. In immersive environments, the gaze direction of different users diverges much more than on a conventional TV screen.

The scene is rendered and displayed at full quality only within a small area around the gaze direction [Figure 7(c) and (d)]. Rendered image quality decreases continuously with increasing angular distance from the foveal field of vision. In a similar fashion, a number of ophthalmic diseases or the effects of drugs affecting the visual field can be simulated [19].

The notion of fovated rendering is also useful for broadcasting and for the rendering and display of immersive 360° videos. Currently, video codecs are optimized for encoding blocks of pixels at the same resolution in every part of the video frame. In the light of the retina's vastly varying perception characteristics from foveal to peripheral vision, however, future gaze-contingent video codecs will be able to adapt coding rate to local-view eccentricity. With gaze-contingent coding, only perceptually relevant information needs to be transmitted and rendered, saving bandwidth and memory. In terms of bandwidth and computation complexity, gaze-contingency may be especially valuable for light-field video playback and light-field displays rendering multiple viewpoints for each displayed frame.

Other applications for gaze-contingent HMDs can be found in the context of collaborative virtual reality. With HMD-integrated eye trackers, gaze direction can be truthfully transferred to avatars, enhancing immersion by enabling collaborating users to establish eye contact in virtual reality.

Even eye blinks can be mapped to the avatar, and potentially some facial expressions may also be picked up by the eye tracker and transferred.

Finally, being able to track gaze inside HMDs is a necessary prerequisite for evaluating perception in immersive environments [30]. For example, 360° videos in HMDs are viewed fundamentally differently from conventional movies on a TV screen. Based on eye-tracking-integrated HMD goggles, new visualization and analysis tools can be developed to investigate viewing behavior when one is immersed all around in live-action footage (Figure 8).

Outlook

Knowing where we are looking allows gaze-contingent display algorithms to provide for a much-enhanced viewing experience. Whether by avoiding aliasing artifacts, by exploiting perceptual characteristics, or by allocating computational resources more efficiently, gaze-contingent computational methods are able to boost perceptual fidelity for all kinds of conventional displays. With affordable eye-tracking solutions becoming consumer electronics items, their widespread deployment is only a question of time and social acceptance.

Current limitations include, of course, that gaze-contingent methods concentrate largely on the single-viewer case. To accommodate several people looking at the same screen, the presented methods need to be suitably adapted and extended.

Other limitations arise from imperfect saliency prediction, e.g., for video content that features no or multiple salient regions. More research is also needed into how we perceive with our peripheral vision and how to suitably render and display image information in the periphery of our visual field. We have only just begun exploring the possibilities of gaze-contingent computational displays, and many more exciting methods and applications are certain to be discovered in the coming years.

Acknowledgments

Authors

Michael Stengel (stengel@cg.cs.tu-bs.de) received his diploma degree in computational visualistics from the University of Magdeburg, Germany, in 2011. From 2010 to 2011, he worked at the Virtual Reality Lab at Volkswagen AG. He is currently pursuing his Ph.D. degree in computer graphics at Technische Universität, Braunschweig, Germany. His research interests include virtual reality, eye-tracking, visualization, and perceptual rendering. He is a Student Member of the IEEE.

Marcus Magnor (magnor@cg.cs.tu-bs.de) is full professor of computer science at Technische Universität, Braunschweig, Germany, where he is director of the Computer Graphics Lab. He also holds an appointment as an adjunct professor in the Department of Physics and Astronomy at the University of New Mexico, Albuquerque. His research interests focus on the natural phenomenon of images, from image formation, acquisition, and analysis to image synthesis, display, perception, and cognition. His areas of research include computer graphics, computer vision, visual perception, image processing, computational photography, astrophysics, imaging, optics, visual analytics, and visualization. He is Senior Member of the IEEE.

References

[1] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: A review," *J. Vision*, vol. 11, no. 5, pp. 13, 2011.

[2] M. Stengel, P. Bauszat, M. Eisemann, E. Eisemann, and M. Magnor, "Temporal video filtering and exposure control for perceptual motion blur," *IEEE Trans. Visual. Comp. Graph.*, vol. 21, no. 5, pp. 663–671, 2015.

[3] L. C. Loschky and G. S. Wolverton, "How late can you update gaze-contingent multiresolutional displays without detection?," *ACM Trans. Multim. Comput. Commun. Appl.*, vol. 3, no. 4, pp. 7, 2007.

[4] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford, UK: Oxford University Press, 2011.

[5] D. Cazzato, M. Leo, and C. Distante, "An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation," *Sensors*, vol. 14, no. 5, pp. 8363–8379, 2014.

[6] J. M. Henderson and A. Hollingworth, "Eye movements during scene viewing: An overview," in *Eye Guidance in Reading and Scene Perception*, G. Underwood, ed. Oxford, UK: Elsevier, vol. 11, pp. 269–293, 1998.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[8] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 1, pp. 185–207, 2013.

[9] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012, pp. 733–740.

[10] Q. Zhao and C. Koch, "Learning saliency-based visual attention: A review," *Signal Process.*, vol. 93, no. 6, pp. 1401–1407, 2013.

[11] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, Tech. Rep. MIT-CSAIL-TR-2012-001, 2012.

[12] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, "Subtle gaze direction," *ACM Trans. Graph.*, vol. 28, no. 4, pp. 100, 2009.

[13] E. M. Reingold, L. C. Loschky, G. W. McConkie, and D. M. Stampe, "Gaze-contingent multiresolutional displays: An integrative review," *Hum. Factors*, vol. 45, no. 2, pp. 307–328, 2003.

[14] A. T. Duchowski and A. Çöltekin, "Foveated gaze-contingent displays for peripheral lod management, 3D visualization, and stereo imaging," *ACM Trans. Multim. Comput. Commun. Appl.*, vol. 3, no. 4, pp. 6, 2007.

[15] M. Levoy and R. Whitaker, "Gaze-directed volume rendering," *ACM SIGGRAPH Comp. Graph.*, vol. 24, no. 2, pp. 217–223, 1990.

[16] T. Ohshima, H. Yamamoto, and H. Tamura, "Gaze-directed adaptive rendering for interacting with virtual space," in *Proc. IEEE Virtual Reality Annu. Int. Symp.* 1996, pp. 103–110.

[17] D. Luebke and B. Hallen, "Perceptually driven simplification for interactive rendering," in *Proc. 12th Eurographics Workshop on Rendering Techniques*, 2001, pp. 223–234.

[18] H. Murphy and A. T. Duchowski, "Gaze-contingent level of detail rendering," in *Proc. EuroGraphics*, 2001, pp. 1–10.

[19] J. S. Perry and W. S. Geisler, "Gaze-contingent real-time simulation of arbitrary visual fields," in *Proc. Electronic Imaging: International Society for Optics and Photonics*, 2002, pp. 57–69.

[20] D. Parkhurst and E. Niebur, "A feasibility test for perceptually adaptive level of detail rendering on desktop systems," in *Proc. 1st ACM Symp. Applied Perception in Graphics and Visualization*, 2004, pp. 49–56.

[21] D. M. Levi, S. A. Klein, and A. Aitsebaomo, "Vernier acuity, crowding and cortical magnification," *Vision Res.*, vol. 25, no. 7, pp. 963–977, 1985.

[22] M. Stengel, S. Grogorick, M. Eisemann, and M. Magnor, "Adaptive image-space sampling for gaze-contingent real-time rendering," in *Proc. Eurographics Conf. Rendering Techniques*, 2016, vol. 35, no. 4, pp. 1–11.

[23] D. M. Hoffman, V. I. Karasev, and M. S. Banks, "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," *J. Soc. Inform. Display*, vol. 19, no. 3, pp. 271, 2011.

[24] C. Lipski, C. Linz, T. Neumann, and M. Wacker, and M. Magnor, "High resolution image correspondences for video post-production," in *Proc. IEEE Conf. Visual Media Production*, 2010, pp. 33–39.

[25] M. Stengel, M. Eisemann, S. Wenger, B. Hell, and M. Magnor, "Optimizing apparent display resolution enhancement for arbitrary videos," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3604–3613, 2013.

[26] P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H. P. Seidel, "Apparent display resolution enhancement for moving images," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 113, 2010.

[27] K. Templin, P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H. P. Seidel, "Apparent resolution enhancement for animations," in *Proc. ACM 27th Spring Conf. Computer Graphics*, 2011, pp. 57–64.

[28] M. Stengel, S. Grogorick, M. Eisemann, E. Eisemann, and M. Magnor, "An affordable solution for binocular eye tracking and calibration in head-mounted displays," in *Proc. ACM Multimedia*, 2015, pp. 15–24.

[29] M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath, "Depth perception with gaze-contingent depth of field," in *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems*, 2014, pp. 217–226.

[30] T. Löwe, M. Stengel, E. C. Förster, S. Grogorick, and M. Magnor, "Visualization and analysis of head movement and gaze data for immersive video in head-mounted displays," in *Proc. Workshop on Eye Tracking and Visualization*, 2015, vol. 1.

SP

Min Wu, Adi Hajj-Ahmad, Matthias Kirchner, Yanpin Ren,
Changshui Zhang, and Patrizio Campisi

Location Signatures That You Don't See

Highlights from the IEEE Signal Processing Cup 2016 Student Competition

When thinking about using technologies to find one's location, many of us may consider satellite-based global navigation systems, such as the U.S. global positioning system (GPS). Smartphones today use GPS type signals and cellular network signals to determine the phone user's current location and offer a broad range of location-related services. But when there was no such real-time connection and signal to analyze, for example, having only a piece of video or voice recording as an evidence of a news story or from child exploitation investigations (which have become an important global issue being addressed by the United Nation's UNICEF organization), are we out of luck?

Fortunately, some very weak traces about the location may be "imprinted" to become part of the sound or visual signals during the sensing process [1]. These traces cannot easily be seen or heard by a human, but it is possible to extract them using signal processing techniques. The contribution of these location signatures–perhaps less foreseeable to a layman—come from the variations in electric power supplies [2]. Thanks to information forensics research in the last decade, the frequency variations, known as *electric network frequency* (*ENF*) signals, are being extracted and exploited to answer a range of forensic questions. Interested readers may refer to

publications such as [3] and [4] and the references cited therein.

Exploring location information of these ENF signals was the topic area of this year's IEEE Signal Processing Cup (SP Cup) competition. The SP Cup is an undergraduate competition organized by the IEEE Signal Processing Society (SPS) in which undergraduate students work in teams to tackle a real-life signal processing problem. Launched in 2014, the SP Cup competition has been held annually, and 2016 is the third edition. To join the SP Cup competition, undergraduate students are required to form a team. Each team is composed of one faculty member to advise the team members, up to one graduate student to assist the supervisor in mentoring the team, and three to ten undergraduate students. Three top teams are selected from the initial round of competition and provided travel grants to participate in the final competition at the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). The final results are shown in "Winners of the SP Cup 2016."

This article provides an overview of the technical tasks and highlights representative approaches and participants' reflections.

Location signatures in power and media

As suggested by its name, ENF is the frequency of a power distribution grid. The

> "Everyone has gained knowledge or skills in different aspects. The SP Cup is a catalyst."

nominal value of ENF is 60 Hz in North America and 50 Hz in most other parts of the world. ENF typically does not stay constant at its nominal value but rather fluctuates around it due to load changes across the grid and the control mechanisms that adjust the amount of power generation to stabilize the frequency changes. We refer to the changing values of the ENF over time as an *ENF signal*. An example of an ENF signal can be seen in Figure 1.

What makes the ENF particularly relevant to multimedia forensics is that audio or video recordings captured in areas where there is electrical activity, such as our offices and living rooms, may capture the ENF variations. In audio, this can be from the ambient power hum, and in video, this is due to the near-invisible flickering of electrical lighting. As the exact variations of ENF appear to be random and difficult to predict, the embedded ENF traces can serve as an intrinsic fingerprint that can be used for a number of forensic applications. In recent years, it has been shown that the embedded ENF traces can be used as a timestamp of media recordings as well as an indicator as to whether or not the media recording was tampered with [3], [5].

For the 2016 SP Cup, the competition topic was shaped around using ENF to identify the power grid in which a media recording was made. Upon examining ENF signals collected from different

FIGURE 1. A spectrogram showing ENF signals in concurrent recordings of power main, audio and visual recording. The cross-correlation study can show similarity between media and power line reference at different time lags, where the peak suggests their temporal alignment. (a) Concurrent power-line. (b) From the audio track. (c) From the video track.
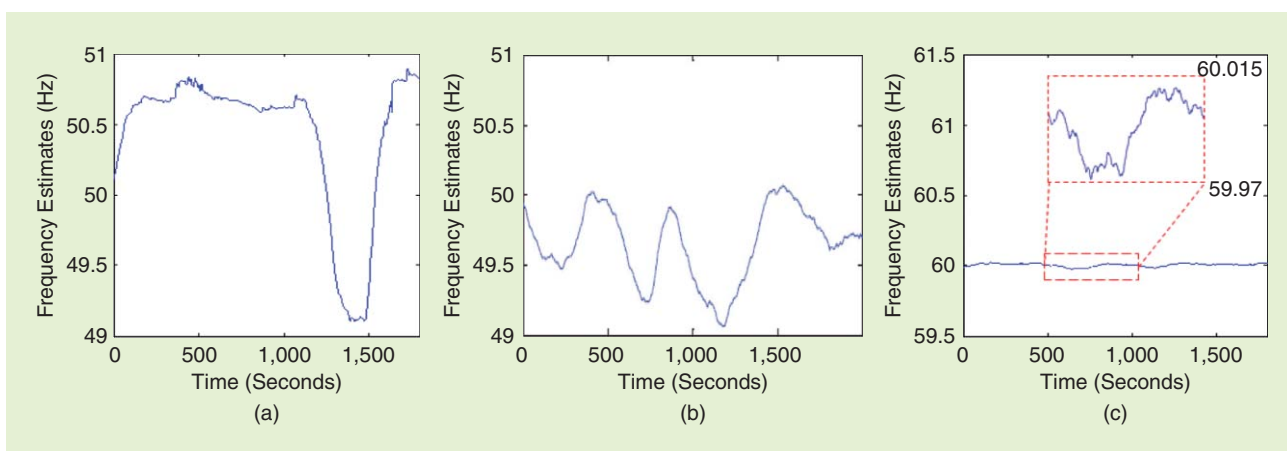
FIGURE 2. Instantaneous ENF variations over time exhibit different characteristics in grids around the world. (a) Lebanon. (b) India. (c) Eastern United States. (Figure used courtesy of [6].)

grids, one can see they differ in the nature and manner of their variations. An example is shown in Figure 2, where we can compare data from Lebanon versus India versus the United States. A machine-learning system can be constructed to learn the characteristics of the ENF signals from different grids to discriminate the grid of origin of a certain recording by examining the ENF traces embedded in it.

Tasks in SP Cup 2016

The open competition stage
The SP Cup started with an open competition stage from September 2015 to January 2016, consisting of two parts. For the first part, participants were first given recordings made in nine grids labeled A–I. Some of the recordings were audio recordings from different locations and environments, and others were power

recordings. Power recordings were made using a custom-made circuit that records the power signal directly from an electric outlet fed into a digital recorder. The main part of a typical circuit for this purpose includes a step-down transformer and a voltage divider. The power recordings generally exhibit very strong ENF traces as sinusoid signals around the base frequency and have much stronger and cleaner traces than the ones seen in regular audio recordings.

The participants were asked to extract the ENF traces from the recordings and build a machine-learning system that could learn the unique characteristics of the ENF signals from different grids. To help the participants test their algorithms and systems and refine their work, they were also provided with 50 practice samples of power and audio recording, each of which was ten minutes long. A simple

oracle website was set up for participants to input their predictions on this practice data set and then receive feedback on the accuracy of their predictions. During the final month of the open competition, participants were provided with the final testing data set, composed of 100 power and audio recordings of ten minutes long each. Both the practice samples and final test samples included some of the recordings from grids outside the nine grids of A–I given for training. Participants were asked to report their predictions on this testing data set as part of their submissions of the open competition. The accuracy of their results was one of the judging criteria in the choosing of the top three teams.

For the first time since the inception of the SP Cup, the competition included hardware and sensing components, which are important aspects of signal processing.

Winners of the SP Cup 2016

Grand Prize: Team Resonance_101
- Bangladesh University of Engineering and Technology, Bangladesh
- Undergraduate students: Sayeed Shafayet Chowdhury, Md Billal Hossain, Jayanta Dey, Ashraful Islam, Rakib Hyder, Md. Samzid bin Hafiz, Ratul Khan, Munif Ishad Mujib, Tauhiduzzaman Khan Himel, and Uday Saha
- Supervisor: Mohammad Ariful Haque.

Second Prize: Team Hammer Down
- Purdue University, United States
- Undergraduate students: Alexandria Moore, Xiangyu Qu, Samuel Sowell, Fangjia Zhu, Siqing Wei, Gregory Dykes, and Christopher Chow
- Supervisor: Stanley Chan
- Graduate Mentor: Benjamin Vondersaar.

Third Prize: Team Vidyut
- Multiuniversity team: Indian Institute of Technology–Madras, National Institute of Technology–Karnatak, Surathkal, and Indian Institute of Science–Bangalore
- Undergraduate students: Priyadarshini Savan Roshan, Pradyumna Byappanahalli Suresha, Aditya Gaonkar P, and Supriya Nagesh

- Supervisor: Prasanta Kumar Ghosh
- Graduate Mentor: Nisha Meenakshi.

In addition to the three overall winning teams, the SP Cup 2016 judging committee made the following special recognitions and honorable mentions.

Special Prize for Sensing Circuit System
- Team "UpatrasECE," University of Patras, Greece
- Team "The ENForcers," University of Novi Sad, Serbia

Special Recognition of Originality
- Team "Hammer Down," Purdue University, United States

Special Recognition of Young Team (student participants in their freshman or sophomore year of college)
- Team "IIT Hyderabad," Indian Institute of Tech Hyderabad

Honorable Mention for Overall Excellence
- Team "UNStoppable," University of Novi Sad, Serbia
- Team "UpatrasECE," University of Patras, Greece

Honorable Mention for High Classification Accuracy
- Team "USTC2016," University of Science and Technology of China.

Participants were asked to design and implement their own circuit to measure a power signal and extract and analyze the ENF traces. After designing the circuit, they were instructed to use it to collect recordings from the power grid in which they live and then analyze these extracted ENF signals and compare them to what they extracted from the recordings of grids A–I provided in the competition. Each team was instructed to provide approximately ten hours of collected recordings together with their analysis as part of their submission for the open competition.

Final competition

After the judging committee evaluated the submissions from the open competition, three finalist teams were chosen to advance to the final competition. Prior to attending the final event at ICASSP, the finalist teams were provided with additional training data from three new grids, which were selected from the recordings submitted in the open competition by other teams. The participants were asked to incorporate this newly acquired data

into their learning systems and were allowed to further refine their systems.

The final SP Cup event was held at the ICASSP in Shanghai, China, on 20 March 2016. At the beginning of the event, the participants of the finalist teams were provided with an additional 50 ten-minute testing samples to combine with the test samples from the open competition round, for a total of 150 testing samples. After submitting the results within an hour, each team presented its work on the ENF extraction, the learning system, the sensing circuits, and the corresponding signal analysis. The final judging committee convened and selected the first-second- and third-prize winners as well as special recognitions.

Approaches to identify grid of origin

As outlined previously, one of the main goals of the SP Cup competition was to identify the grid of origin of power and audio signals containing ENF traces.

Many of the teams extracted the dominant instantaneous frequency around

nominal ENF frequencies over short time windows and used the features proposed in [6] and their variations or extensions to develop a feature set for pattern analysis and classification. The basic types of features employed, shown in Figure 3, include 1) statistical features related to ENF values (mean, variance, and range), 2) features extracted from wavelet analysis of ENF signals, and 3) features based on autoregressive modeling of ENF signals. Two examples of other features proposed by participating teams are histogram-based features around ENF base and harmonic frequencies proposed by Team Hammer Down of Purdue University, United States, and features based on the extrema in the ENF signal and the rising edges that can be present developed by Team UpatrasECE of the University of Upatras, Greece.

A common strategy for pattern analysis and classification that was employed by the participating teams, particularly the three finalist teams, was to differentiate between four different types of signals: 1) power signals with 50-Hz nominal ENF, 2) power signals with 60-Hz
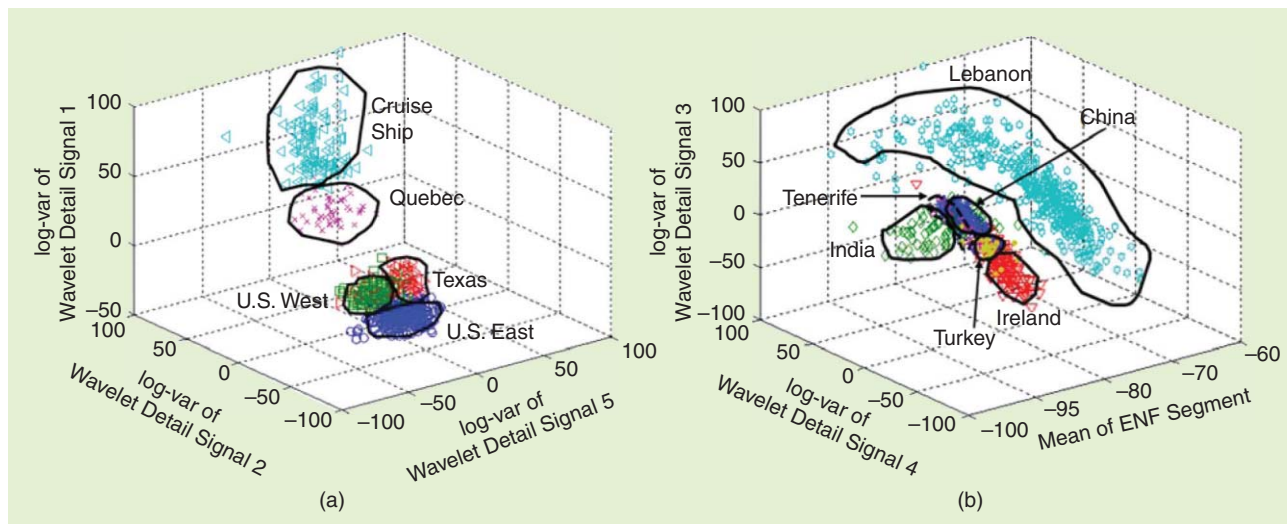
FIGURE 3. An example of features derived from ENF signals for differentiating the originating power grids.

nominal ENF, 3) audio signals with 50-Hz nominal ENF, and 4) audio signals with 60-Hz nominal ENF. While each team had their own design for the learning and classification system, their solutions all consisted of four subsystems, one dedicated to each data type. Some of these subsystems had certain variations in the way the ENF signals were extracted or utilized different discriminating features, depending on which of the four types to which the signal belonged.

The first-place winning team, Team Resonance_1011 from Bangladesh, employed two-stage support vector machine (SVM) classifiers for each data type. The first one is a one-class SVM classifier that decides if the signal belongs to a grid seen in training, and the second SVM classifier narrows down the list of possible grids on the basis of different discriminating features extracted from the embedded ENF signals. The last stage of

the classification is where the testing signal is passed to a "pole-matching" classifier to reach a final decision on the grid of origin based on the minimum distance between the estimated poles of the training and testing grids.

Team Hammer Down from the United States, which took second place, carefully studied the characteristics of the signals for this problem and proposed a different method from what is available in the recent ENF literature. The team's novel method utilized a multiharmonic histogram to analyze ENF signals for identifying power grids. Team members computed ENF signals at multiple harmonic locations and extracted histograms of 1) ENF magnitudes, 2) the signal power around the ENF, 3) the noise power around the ENF, and 4) the signal-to-noise ratio of the ENF. At the classification stage, they proposed a histogram matching method with a multilayer decision rule for identification.

Team Vidyut from India, which took third place, used a multistage SVM system where, for each input test recording, five predictions from five respective classifiers were received. To compute the confidence of a prediction, team members proposed an entropy-based measure of confidence where lower entropy means higher confidence. For audio signals, they used an additional subband classifier that captures the relative strength among the different ENF harmonic components captured by a recording. The team used custom-designed detectors to resolve conflicts in decision between their SVM and subband classifiers.

Sensing circuits

For the hardware and signal acquisition tasks in the SP Cup that aimed at bringing out the synergy of sensing, processing, and learning, nearly all participating teams built their own circuit for collecting reference power recordings from the team's respective locations. Useful references include an article in *IEEE Signal Processing Magazine* on observing the grid [7] as well as ENF-related literature [3], [5], [8]. Among the submissions received, some teams adopted the reference design, while other teams made improvements, incorporated innovative features, or employed different implementations.

Given that this is the first time the SP Cup competitions included sensing hardware design and implementation in addition to algorithmic tasks, the judging
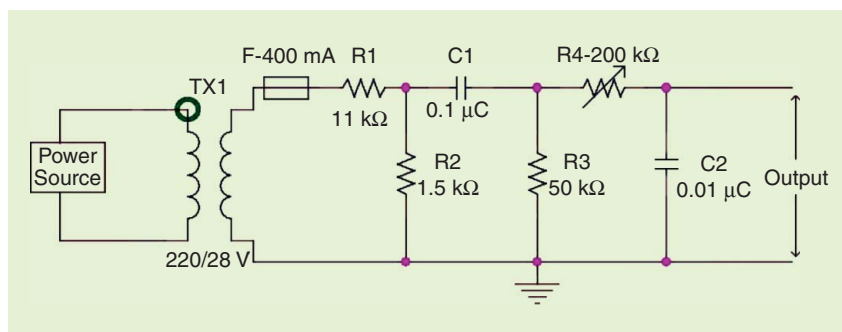
FIGURE 4. A circuit diagram from Team UpatrasECE from the University of Upatras, Greece.

committee reviewed this element in all entries and awarded special recognition to two teams for excellence in this task: Team UpatrasECE from Greece and Team The ENForcers from Serbia.

As shown in Figure 4, Team "UpatrasECE" thoughtfully included a glass fuse for the safety of the circuit operation. After the transformer and voltage divider that creates a reference voltage of 3 V, they included a high-pass filter to block the dc component of the measured signal and thus focused on the variation patterns and an antialiasing filter to limit the bandwidth to 500 Hz and pass on to a recording device through a 3.5-mm jack. Team "UpatrasECE" was also given an honorable mention for the overall excellence of their project.

The circuit of Team The ENForcers shown in Figure 5, included a transformer, a low-pass filter, a voltage divider, an amplifier for voltage adjustment, and an analog-to-digital (A/D) converter provided by an Arduino Uno board. An INA122P instrumentation amplifier was used in conjunction with other components for voltage adjustments so as to make the voltage compatible with the Arduino A/D converter. The sampled digital signal can be sent to a PC using Arduino's built-in serial connection for storage and analysis.

FIGURE 5. The sensing circuits by Team The ENForcers.

SP Cup 2016 statistics

SP Cup 2016 engaged participants from nearly 30 countries, covering every habitable continent. Three hundred thirty-four students from 23 countries and forming 52 teams registered for the competition, as shown in Figure 6. Among them, more than 200 students on 33 teams turned in the required submissions by the open competition deadline in January 2016. Based on team photos and other information submitted, about one quarter of the participating students were female, with many teams populated by women at a 50% or higher rate. Similar to SP Cup 2015, we once again saw strong participation from students in Asia-Pacific regions. At least two teams from Asia were in the final

competition in each of the three SP Cup competitions.

To facilitate interaction and questions-and-answer sessions with participants, the organizing team used Piazza, a popular course interaction platform, as an online bulletin board to post resources and data sets as well as to address questions and engage participants. About 200 participants "enrolled" in SP Cup 2016 on Piazza and made approximately 250 contributions; in total, they have accumulated about 2,300 days of access and over 4,000 views of posts. Interested readers may access the archived Piazza site at https://piazza.com/ieee_sps/other/sp1601/home.

Since its inception, the SP Cup has received generous support from MathWorks, Inc., the maker of the popular MATLAB and Simulink platforms. MathWorks also provided funding support to the SP Cup and contributed their expertise. Each student team that registered for the SP Cup was provided complimentary software access to MATLAB and related toolboxes. The IEEE SPS welcomes continued engagement and support from industry in future SP Cup competitions. Interested supporters may contact Dr. Patrizio Campisi, director for student services, at patrizio.campisi@uniroma3.it.
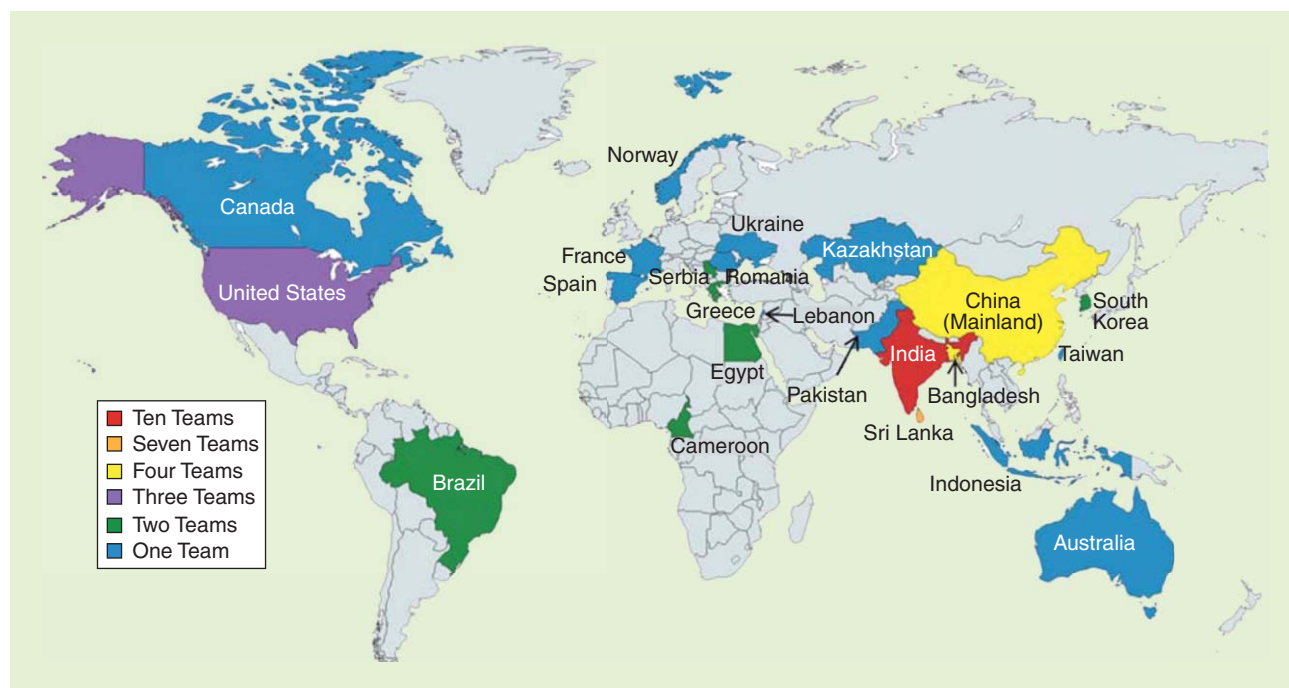
FIGURE 6. A map illustrating the number of registered teams from each participating country around the world.

FIGURE 7. Team Resonance_1011 received the grand prize.

Participants' thoughts

Whether moderating the discussions online or attending the final competition at ICASSP, we could feel the enthusiasm and dedications from the participants. Amid the understandably intense competition, one could also sense a bond, and a community formed between students of different teams through the SP Cup endeavor. Here we highlight a few words from the finalist teams as each team reflected on their journey.

Team Resonance_1011

Bangladesh University of Engineering and Technology (BUET) has had a team as a finalist in every SP Cup so far—and this year, Team Resonance_1011 (see Figure 7) received the grand prize! What an impressive performance and persistent engagement! Three students from this year's BUET team had also participated in the previous SP Cup competition and spoke enthusiastically of the rewarding experiences gained from this unique journey of learning and professional growth.

- "The SP Cup was a great learning curve indeed both times I participated. It taught us to work as a team and follow a stringent time schedule"— Sayeed Shafayet Chowdhury
- "Since this time the task was on a broader scale and we had a team comprising of ten students from different classes, we split up our work and tried to have online contacts regularly, besides having meetings with our supervisor, thus boosting our teamwork capabilities"—Billal Hossain
- "This year we had hardware tasks too... We realized it was possible to take a much simpler approach [than using complex components and boards]. … We were able to devise an extremely low cost (lower than US$5) sensing circuit that provides a perfectly adequate output."—Ratul Khan and Munif Ishad Mujib
- "The SP Cup has profound impacts on our students. Through the competition, they get to know how fundamental signal processing algorithms are used to solve a real and complicated signal processing problem. They also need to explore many advanced signal processing areas outside their text books. They become familiar with signal processing journals and conference proceedings. We have also been able to attract some good and talented students to signal processing research through their participation in the SP Cup.…Each SP Cup gave me the opportunity to work on a new signal processing application. It increases my depth and breadth of knowledge that is really helpful for teaching signal processing courses."—Mohammad Ariful Haque, faculty mentor.
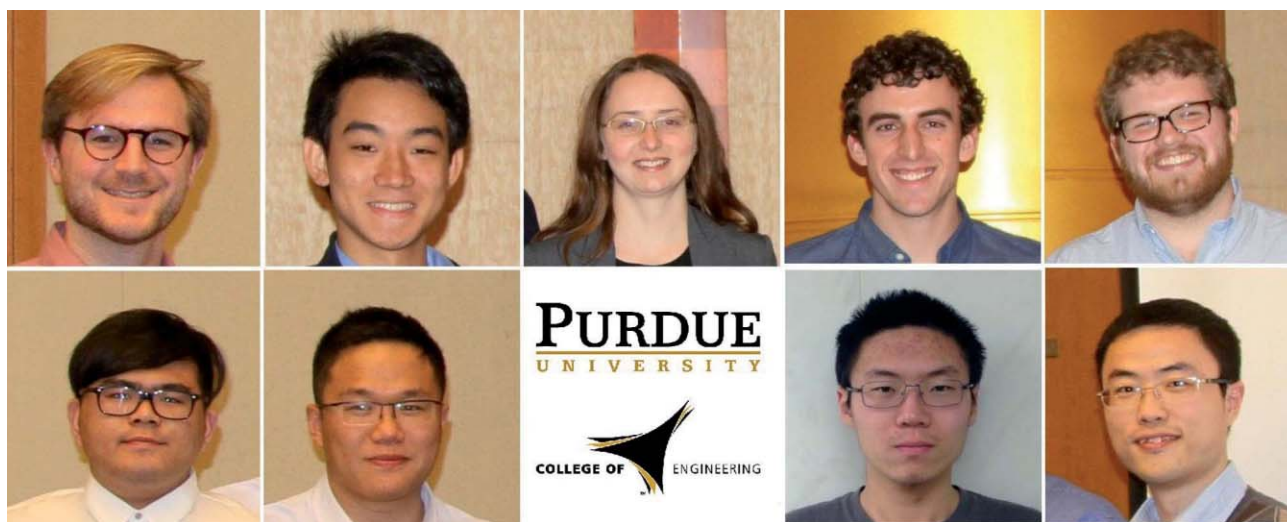
FIGURE 8. Team Hammer Down won second prize and a special recognition for original algorithm.

Team Hammer Down

Team Hammer Down from Purdue University, United States (see Figure 8), was composed of student members who were junior students (in the third year of their college career). They noted that their team was likely one of the most diversified teams in the competition, as they have students from different ethnic and cultural backgrounds with different areas of interest.

- "As a team with all members being junior students,…everyone has gained knowledge or skills in different aspects. Some of us have never done serious research.…Some of us are interested in signal processing but lack a strong motivation to go deeper… The SP Cup is a catalyst"—Team Hammer Down
- "The competition itself provided a new and challenging way to apply what I had learned in my classes. In addition to challenging myself I also learned a lot while at ICASSP. Talking to professionals in my field helped to broaden my view of electrical engineering as a whole. All in all, the SP Cup was a great experience and I recommend it to any undergraduates interested in signal processing"—Sam C. Sowell
- "I have learned a lot during the SP Cup as a junior student. It was a great chance for me to apply and practice what I have learned in the classes in a different way; instead of practicing on idealized problem sets that serve to help students understand the concepts, I was able to face some realistic signal processing problems and challenges"—Xiangyu Qu
- "Mentoring undergraduate students for the SP Cup is a challenging but rewarding experience. The difficult part is how to plan and divide the tasks, for most of the undergraduates do not have adequate background and experience. However, I am impressed by their creativity and the effort they spent on the project. They are a wonderful team. The competition provides a good opportunity for students to explore different aspects of signal processing. I personally most appreciate the integration of signal pro-

FIGURE 9. Team Vidyut from India won third prize.

cessing and circuit, for it bridges the gap between the two seemingly less correlated ends of our curriculum nowadays."—Prof. Stanley Chan, faculty mentor.

Team Vidyut

Team Vidyut (Figure 9) was formed by the collaborated efforts of undergraduate students from the National Institute of Technology–Karnatak, Surathkal (NIT-K) and Indian Institute of Technology—Madras (IIT-M), along with graduate and faculty mentors from the Indian Institute of Science–Bangalore (IISc) where the undergraduate students first met and interned. Returning to their universities in the last month of the competition did not hamper the undergraduate students' teamwork, as they continued to make progress by frequent e-mails and regular phone calls. The team has a perfect gender balance with two female and two male undergraduate students and leveraged the complementary strengths of the members.

- "'Vidyut' means electricity in many Indian languages…Although coordinating amongst ourselves had become a challenge by itself, the sense of team spirit grew in us, as we worked for the competition."—Team Vidyut
- "As an undergraduate, it was very exciting to work on a real-world problem. … Coding was made easier, thanks to MATLAB"—Priyadarshini Savan Roshan, NIT-K
- "In addition to the technical skills gained, this challenge helped me coor-

dinate and communicate better among a team"—Supriya Nagesh, NIT-K
- "Attending the finals in Shanghai was an amazing experience where we had a chance to observe the different views taken by the other teams to approach the same problem. On the whole, participating in the SP Cup proved to be a great learning experience with memorable moments to cherish"—G. Nisha Meenakshi, IISc, graduate mentor.

More reflections from SP Cup finalists and participants along with project reports of selected teams can be found on the IEEE SigPort repository at http://www.sigport.org/events/sp-cup-project-reports. Figure 10 shows a few selected photos provided by the participating teams that showcased their teamwork during the competition.

In closing, we would like to convey our heartfelt congratulations to the winners and teams that received honorable mentions on their excellent performances! Thanks all participants for making IEEE SP Cup 2016 successful! As you read through this issue of *IEEE Signal Processing Magazine*, you will see that the 2017 IEEE SP Cup competition has been announced (see page 5). Please encourage students you know to take part in the competition, mentor a student team if you can, come to support in ways feasible, and stay tuned for another edition of the IEEE SP Cup Global Student Competition!

Acknowledgments

FIGURE 10. Behind-the-scenes teamwork photos of the SP Cup 2016 teams that received special recognitions and honorable mentions (from the top, counter-clockwise) Team UpatrasECE, Greece; Team The Enforcers, Serbia; Team IIT Hyderabad, India; and Team UNStoppable, Serbia.

Authors

Min Wu (minwu@umd.edu) is a professor and distinguished scholar-teacher at the University of Maryland, College Park. She is currently the editor-in-chief of *IEEE Signal Processing Magazine* and an IEEE Distinguished Lecturer.

Adi Hajj-Ahmad (adiha@umd. edu) is a Ph.D. candidate at the University of Maryland, College Park. He received a paper award and two invention awards for his research on media forensics.

Matthias Kirchner (kirchner@ binghamton.edu) is an assistant professor at Binghamton University, New York. He is an elected member of the IEEE Technical Committee on Information Forensics and Security.

Yanpin Ren (renyp@tsinghua.edu.cn) is a senior engineer at Tsinghua University, Beijing, China, where she leads the curriculum development in circuit designs and embedded systems.

Changshui Zhang (zcs@tsinghua. edu.cn) is a professor at Tsinghua University, Beijing, China, with expertise in machine learning. He is an editor of IEEE SigPort and serves on the Student Service Committee of the IEEE Signal Processing Society.

Patrizio Campisi (patrizio.campisi@ uniroma3.it) is a professor at Roma Tre University, Rome, Italy. He chairs the Student Service Committee of the IEEE Signal Processing Society and served as the general chair of the 2015 IEEE Workshop on Information Forensics and Security.

References

[1] C. Carroll. (2015, June 9). Fingerprints of power: Researcher tracks video locations using electrical grid signal. [Online]. *Terp Magazine*, University of Maryland. Available: http://terp.umd.edu/fingerprints-of-power/#.VweLz_krLD4

[2] M. H. Bollen and I. Gu, *Signal Processing of Power Quality Disturbances*, Hoboken, NJ: Wiley, 2006.

[3] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital audio and video recordings," *J. Audio Eng. Soc.*, vol. 57, no. 9, pp. 643–661, Sept. 2009.

[4] M. Stamm, M. Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade, invited paper for the inaugural issue," *IEEE Access*, vol. 1, pp. 167–200, May 2013.

[5] R. Garg, A. L. Varna, A. Hajj-Ahmad, and M. Wu, "'Seeing' ENF: Power signature based timestamp for digital multimedia via optical sensing and signal processing," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 9, pp. 1417–1432, Sept. 2013.

[6] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF-based region-of-recording identification for media signals," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 6, pp. 1125–1136, June 2015.

[7] P. Top, M. R. Bell, E. Coyle, and O. Wasynczuk, "Observing the power grid," *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 24–32, Sept. 2012.

[8] N. Fechner and M. Kirchner, "The Humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings," in *Proc. 8th Int. Conf. IT Security Incident Management & IT Forensics*, Munster, Germany, 2014, pp. 3–13.

SP

IEEE International Conference on Acoustics, Speech and Signal Processing

March 5-9, 2017   New Orleans, USA
Internet of Signals

Call for Papers

As music and rhythm are the heartbeats of life, signal and information processing is the heartbeat of technology development for our daily life. Having both of them capture the hearts and souls of all the attendees is the goal of the 42th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017) which will be held in Hilton Conference Centre, at the Jazz music capital (New Orleans, USA) on March 5-9, 2017. ICASSP is the world's largest and most comprehensive technical conference focused on signal processing and its applications. The conference not only introduces new developments in the field but also provides an engaging forum to exchange ideas, and does so both for researchers and developers. The theme of the ICASSP 2017 is "The Internet of Signals" which is the real technology and world behind the Internet of Things. The conference will feature world-class international speakers, tutorials, exhibits, lectures and poster sessions from around the world. Topics include but are not limited to:

- Audio and acoustic signal processing
- Bio-imaging and biomedical signal processing
- Design & implementation of signal processing systems
- Image, video & multidimensional signal processing
- Industry technology tracks
- Information forensics and security
- Machine learning for signal processing
- Multimedia signal processing
- Remote Sensing and signal processing
- Signal Processing for Smart Systems

- Sensor array & multichannel signal processing
- Signal processing education
- Signal processing for communication & networking
- Signal processing theory & methods
- Signal processing for Big Data
- Internet of Things and RFID
- Speech processing
- Spoken language processing
- Signal Processing for Brain Machine Interface
- Signal Processing for Cyber Security
- Computational Imaging

Submission of Papers:

Prospective authors are invited to submit full-length papers, with up to four pages for technical content including figures and possible references, and with one additional optional 5th page containing only references. A selection of best papers will be made by the ICASSP 2017 committee upon recommendations from Technical Committees.

Special Sessions:

Special session proposals should be submitted by July 11th, 2016. Proposals for special sessions must include a topical title, rationale, session outline, contact information for the session chair, a list of authors, and a tentative title and abstract. Additional information can be found at the ICASSP 2017 website (www.ieee-icassp2017.org).

Tutorials:

Brief proposals for tutorials should be submitted by September 15th, 2016. Proposal for tutorials must include a title, an outline of the tutorial and its motivation, a two-page CV of the presenter(s), and a short description of the material to be covered. Tutorials will be held on March 5th, 2017.

M.Sc./Ph.D. Forum:

The forum will provide an opportunity for M.Sc./Ph.D. students in the fields of electronics, circuits, and embedded systems to present their thesis/dissertation research (completed in the last year), including work in progress within the topics of interest for ICASSP 2017. To encourage interaction among the M.Sc./Ph.D. students and researchers from academia, industry, and government, the forum will be organized as a poster session. A two-page paper on the thesis work will appear in the conference proceedings.

Signal Processing Letters:

Authors of SPL will be given the opportunity to present their work, subject to space availability and approval by the Program Chairs. SPL papers published between January 1st, 2016 and December 31st, 2016 are eligible. Because they are already peer-reviewed and published, SPL papers presented will neither be reviewed nor included in the ICASSP proceedings. Approved papers must have one author/presenter register for the conference.

Demos:

ICASSP 2017 offers a perfect stage to showcase innovative ideas in all technical areas of interest at ICASSP. All demo sessions will be highly interactive and visible. Please refer to the ICASSP 2017 website for additional information regarding demo submission.

Deadlines:

Special session proposals:	*July 11th, 2016*
Tutorial proposals:	*September 15th, 2016*
Notification of special session acceptance:	*August 15th 2016*
Notification of tutorial acceptance:	*October 15th 2016*
M.Sc./Ph.D. papers due:	*October 15th, 2016*
Demo proposals due:	*October 15th, 2016*
Submission of regular papers:	*September 12th, 2016*
Registration of SPL papers:	*November 21st, 2016*
Notification of M.Sc./Ph.D. papers:	*November 30th, 2016*
Notification of acceptance of demos:	*November 30th, 2016*
Notification of paper acceptance:	*December 12th, 2016*
Author registration:	*January 9th, 2017*

General Chair
Magdy Bayoumi, *U. of Louisiana at Lafayette*

Program Chairs
Tülay Adali, *UMBC, USA*
Eli Saber, *Rochester Inst. of Tech*

Special Sessions Chairs
James Fowler, *Mississippi State U.*
Akihiho K. Sugiyana, *NEC, Japan*

Tutorial Chairs
A. K. Katsaggelos, *Northwestern U.*
M. Mansour, *American U. Beirut*
R. de Queiroz, *C. U. D. Riberio, Brazil*

Plenaries Chairs
Edward Delp, *Purdue U.*
Ahmed Tewfik, *U. of Texas at Austin*

Student Papers Contest
Keigo Hirakawa, *U. of Dayton*
S. Theodoridis, *U. of Athens, Greece*

M.Sc/Ph.D Forum Chairs
Ahmed Abdelgawad, *C. Michigan U.*
Patrice Abry, *E. N. S. Lyon, France*
Mayssaa Al Najjar, *UL at Lafayette*

Women in Engineering
Anna Scaglione, *Arizona State U.*

Finance Chair
Ahmed Abdelgawad, *C. Michigan U.*

Publicity Chairs
T. Elarabi, *California S. U. at Fresno*
Fatma Taher, *KUSTAR, UAE*

Publication Chair
Warren Gross, *McGill U., Canada*

International Committee
Jean-luc Dugelay, *Eurecom, France*
Jesef V. Lillie, *IEEE Cust. Elec. Soc.*
S. Rahardja, *NW. PolyTec. U. China*

Exhibit Chair
Nam Ling, *U. of Santa Clara*

Industrial Chair
Tao Mei, *Microsoft, China*

Demos
Yen-Kuang Chen, *Intel*
Nam Ling, *U. of Santa Clara*

Advisors:
Lian-Gee Chen, *National Taiwan U.*
V. Clarkson, *U. of Queensland, AUS*
Scot Douglas, *SMU*
John Mathews, *U. of Utah*
Mark Smith, *Purdue U.*

Management
Billene Cannon, *Conference Management Services, Inc.*

Website
Zaghloul Saad Zaghloul, *CACS*

IEEE · UNIVERSITY of LOUISIANA LAFAYETTE · IEEE Signal Processing Society

LECTURE NOTES

Yonina C. Eldar, Nathaniel Hammen, and Dustin G. Mixon

Recent Advances in Phase Retrieval

In many applications in science and engineering, one is given the modulus squared of the Fourier transform of an unknown signal and then tasked with solving the corresponding inverse problem, known as phase retrieval. Solutions to this problem have led to some noteworthy accomplishments, such as identifying the double helix structure of DNA from diffraction patterns, as well as characterizing aberrations in the Hubble Space Telescope from point spread functions. Recently, phase retrieval has found interesting connections with algebraic geometry, low-rank matrix recovery, and compressed sensing. These connections, together with various new imaging techniques developed in optics, have spurred a surge of research into the theory, algorithms, and applications of phase retrieval. In this lecture note, we outline these recent connections and highlight some of the main results in contemporary phase retrieval.

Relevance

Phase retrieval finds applications in areas such as optics, X-ray crystallography, astronomical imaging, speech processing, and computational biology. Every application of this inverse problem encounters several fundamental questions: How do we reconstruct the desired signal from the measurements? To what extent is the reconstruction unique and/

or stable? Can we develop new measurement devices that allow for robust signal recovery? These questions form the basis for the notes presented below.

Prerequisites

We assume the reader has a basic understanding of linear algebra, optimization, and probability. Some familiarity with compressed sensing is helpful but not necessary.

Problem statement

In standard applications of phase retrieval, we receive measurements of the modulus squared of the Fourier transform of an unknown signal x_0, $y = |Fx_0|^2$, where the magnitude is taken component-wise. (For simplicity, we model x_0 as a vector in $\mathbb{C}^n$ so that the Fourier operator F may be represented by a matrix.) Observe that, without additional information, this inverse problem is terribly ill posed. For example, if y is the vector of all ones, then x_0 can be any standard basis element multiplied by an arbitrary phase factor. In general, the set of solutions can have as many as n real degrees of freedom. Examples of these degrees of freedom include the so-called trivial ambiguities: $|Fz|^2 = |Fx|^2$ if $z = e^{i\phi}x$ for some $\phi \in [0, 2\pi)$, if z is a translation of x, or if z is the conjugate reversal of x.

> Phase retrieval finds applications in areas such as optics, X-ray crystallography, astronomical imaging, speech processing, and computational biology.

Example 1

Suppose that $x = (2, i, 0, 0)$. Then $|Fx|^2 = (5, 9, 5, 1)$. Let z be any of $(2i, -1, 0, 0), (0, 2, i, 0), or (0, 0, -i, 2)$. These are trivial ambiguities for which $|Fz|^2 = (5, 9, 5, 1)$. However, trivial ambiguities are not the only solutions. For example, $z = (\sqrt{5} + 1/2, i, \sqrt{5} - 1/2, -i)$ is not a trivial ambiguity, but still $|Fz|^2 = (5, 9, 5, 1)$.

To obtain a well-posed problem, we must acquire additional information. Historically, this has been accomplished by imposing structure on the signal. For example, one might assume that x_0 is real and has compact support. In fact, this uniquely determines almost every x_0 up to trivial ambiguities when taking a two-dimensional (2-D) Fourier transform [or three-dimensional (3-D) or higher] [5]. However, unfortunately, in 1-D there is no uniqueness even if the signal support is bounded. Furthermore, even in settings in which uniqueness is guaranteed, there is no known general algorithm to find the unknown signal from its Fourier magnitude. A more recent class of signal structures used in phase retrieval relies on sparsity [4]. A vector is defined to be *k-sparse* if it has at most k nonzero entries. Sparsity priors have been used extensively in many fields of engineering and statistics and

are known to closely approximate various classes of images and signals.

Another way to obtain a well-posed problem is to collect additional intensity measurements. For example, in diffractive imaging, one may implement multiple structured illuminations of the form $\left|FD_jx_0\right|^2$, where each D_j is a known diagonal matrix. In other applications, we can take redundant measurements via the short-time Fourier transform. This approach has been used in speech and audio processing, in measurements of ultra-short laser pulses via frequency resolved optical gating, and in ptychographical diffractive imaging, among others.

To account for the apparent multitude of plausible intensity measurements (such as structured illuminations or the short-time Fourier transform), we consider a general phase retrieval setting in which we receive $y = \left|Ax_0\right|^2$ for some known matrix $A \in \mathbb{C}^{m \times n}$. We then seek to solve the following program:

$$\text{find } x \text{ subject to } \left|Ax\right|^2 = y, x \in S, \quad (1)$$

where $S \subseteq \mathbb{C}^n$ corresponds to the imposed structure. In this lecture note, we focus on cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse vectors. For both settings, we discuss transformations A that allow for (1) to uniquely determine x_0 and consider algorithms that were recently designed to solve (1) for various choices of A. The results we present throughout are surveyed in [6], [8], and [10], unless indicated otherwise.

Uniqueness

For a fixed A, we are interested in whether (1) has a "unique" solution for every $x_0 \in S$ (or for most $x_0 \in S$). We focus on the cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse vectors. Note that in both cases, $x \in S$ if and only if $e^{i\phi} x \in S$ for every $\phi \in [0, 2\pi)$, which means (1) never has a unique solution in the literal sense. To account for this technicality, we say (1) has a unique solution (up to a global phase factor) if every solution lies in the set $[x_0] := \{e^{i\phi} x_0 : \phi \in [0, 2\pi)\}$.

Notice that the set $[x_0]$ is determined by the outer product $x_0 x_0^*$ (and vice

versa). Let a_i^* denote the ith row of A (here, $a_i \in \mathbb{C}^n$ is a column vector, and a_i^* denotes its conjugate transpose). Then

$$y_i = \left|a_i^* x_0\right|^2 = a_i^* x_0 \cdot \overline{a_i^* x_0} = a_i^* x_0 x_0^* a_i. \quad (2)$$

Consider the case where $S = \mathbb{C}^n$. In this setting, (2) implies that (1) has a unique solution for every $x_0 \in \mathbb{C}^n$ precisely when the mapping $x_0 x_0^* \mapsto \{a_i^* x_0 x_0^* a_i\}_{i=1}^m$ is one to one. Recent research has investigated the number of measurements m that are necessary or sufficient for this map to be one to one. For example, it has been shown that the inequality $m \geq 4n - O(\log n)$ is a necessary condition. Conversely, for almost every $A \in \mathbb{C}^{m \times n}$ with $m \geq 4n - 4$, (1) has a unique solution for every $x_0 \in \mathbb{C}^n$. Whether such A exist when $m < 4n - 4$ remains an open problem for general n. We know this is impossible when n has the form $n = 2^k + 1$ and yet possible when $n = 4$ [11]. Alternatively, for almost every $A \in \mathbb{C}^{m \times n}$ with $m \geq 2n$, (1) has a unique solution for almost every $x_0 \in \mathbb{C}^n$, and no such A exists when $m < 2n - 1$.

In many real-world applications, A exhibits some sort of Fourier structure. For example, in the classical setting in which x_0 is compactly supported, A may be viewed as an oversampled Fourier matrix. As mentioned before, 1-D uniqueness from Fourier measurements cannot be guaranteed in general. To achieve uniqueness beyond trivial ambiguities, consider the model in which A is a $kn \times n$ matrix composed of k different $n \times n$ blocks of the form FD_j, where F is the $n \times n$ discrete Fourier transform matrix and each D_j is some diagonal matrix. This model is called the structured illumination model, and each D_j is referred to as a mask. While four such masks are required to determine every possible x_0 (by the aforementioned discussion), we currently only know how to do so with $O(\log n)$ masks. On the other hand, we do know how to determine almost every possible x_0 with only two masks, which matches the above theory; in particular, the two masks may be taken to be the identity

matrix and diag$(0, 1, \ldots, 1)$, as illustrated in the following example.

Example 2

Suppose that x_0 is a vector of length $n = 4$ and we measure it using the two masks $D_0 = I$ and $D_1 = D = \text{diag}(0, 1, 1, 1)$. The resulting structured illumination matrix has the form

$$A = \begin{bmatrix} FI \\ FD \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^9 \\ 0 & 1 & 1 & 1 \\ 0 & \omega^1 & \omega^2 & \omega^3 \\ 0 & \omega^2 & \omega^4 & \omega^6 \\ 0 & \omega^3 & \omega^6 & \omega^9 \end{bmatrix},$$

where $\omega = e^{-i\frac{2\pi}{n}} = -i$. The two-mask structured illumination model measures the discrete Fourier transform of the signal and of the signal minus the first component (or any other desired element). A larger dimensional example of this idea is given in Figure 1.

Another structured example of A is the short-time Fourier transform, which can be thought of as a special case of the structured illumination model: The diagonal entries of each D_j come from a different translation of a common window function of width w, and we only consider every lth translation of this window. This particular measurement model finds applications in cross-correlation frequency-resolved optical gating (XFROG), in which one measures ultrafast laser pulses by optically producing a spectrogram; another application is ptychography, a diffractive imaging method where different overlapping patches of the unknown object are measured. For the short-time Fourier transform model, we know that x_0 is not uniquely determined if it has w consecutive zeros, but (1) does uniquely determine most nonvanishing signals when $l \ll w \ll n$.

Example 3

Consider the short-time Fourier transform of a signal of length $n = 6$. We choose the measurement window as a rectangular function of length $w = 3$

(a) Vector x (b) Fx (c) $|Fx|^2$ (d) Dx (e) FDx (f) $|FDx|^2$
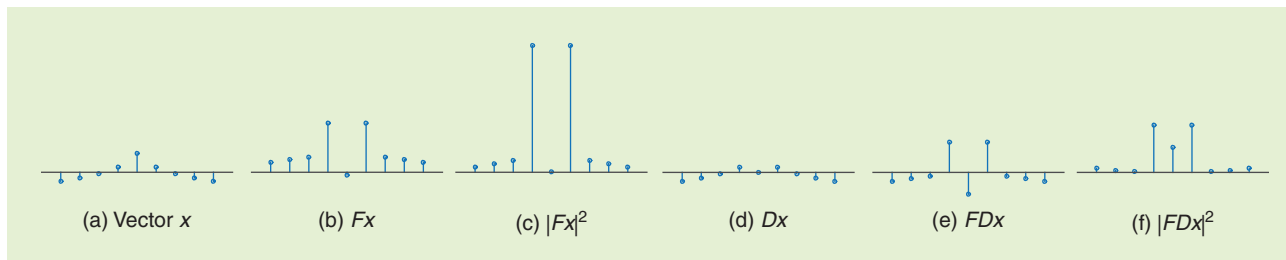
FIGURE 1. An illustration of two deterministic masks. (a) A vector x is chosen with indices from –4 to 4. We take the Fourier transform of x to get (b) and then square it to get (c). Applying the mask with a zero in the zero-indexed component, we obtain (d) a new vector Dx. Note that the Fourier transform (e) and squared Fourier transform (f) of Dx are quite different than those of x. With the data $|Fx|^2$ and $|FDx|^2$ it is possible to recover almost every vector x. Clearly, if x already has a zero in the zero-indexed component; then D will apply no change to x, and it will not be possible to recover x.

and select a step size $l = 2$. The measurements are then the discrete Fourier transforms of vectors that are equal to the signal of interest in the given windows, and 0 elsewhere. With this choice of parameters, the short-time Fourier transform is equivalent to structured illumination using every other translation of diag$(1, 1, 1, 0, 0, 0)$*. The resulting measurement matrix has the form*

$$A = \begin{bmatrix} FD_1 \\ FD_2 \\ FD_3 \end{bmatrix} = \begin{bmatrix} F\text{diag}(1,1,1,0,0,0) \\ F\text{diag}(0,0,1,1,1,0) \\ F\text{diag}(1,0,0,0,1,1) \end{bmatrix}.$$

In the case where S is the set of k-sparse signals, a simple argument gives that for almost every $A \in \mathbb{C}^{m \times n}$, Ax_0 uniquely determines every k-sparse x_0 provided $m \geq 2k$ [4]. When the phases are discarded, the number of measurements increases by a factor of four: It can be shown that for almost every $A \in \mathbb{C}^{m \times n}$ with $m \geq 8k - 4$, (1) uniquely determines every k-sparse x_0. When k is much smaller than n, one may wonder whether the classical phase retrieval problem $|Fx_0|^2 \mapsto [x_0]$ is plausible (we will certainly have $m = n \geq 8k - 4$ intensity measurements, as is typically sufficient). Unfortunately, trivial ambiguities are still present in this case, but we can nonetheless uniquely determine most k-sparse signals with $k = O(n^{1/2 - \epsilon})$ up to trivial ambiguities.

Algorithms

In this section, we describe several methods for solving the phase retrieval problem (1) in the special cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse signals.

The most popular class of phase-retrieval algorithms is based on alternating projections, pioneered by the work of Gerchberg and Saxton and extended by Fienup. These methods consist of iteratively imposing the constraints in time/space and in the Fourier domain, namely, consistency with the measurements. Adapting to our generalized phase-retrieval setup, the basic steps consist of choosing an initial guess, and then alternating between projecting onto the sets $\{x : |Ax|^2 = y\}$ and S:

$$z_n := y^{1/2} \circ \frac{Ax_n}{|Ax_n|}, \quad x_{n+1} := P_S(A^{\dagger} z_n),$$

where $y^{1/2}$ denotes the entrywise square root of y, $\circ$ denotes entrywise product, P_S denotes the nearest-point projection onto S, and $A^{\dagger}$ denotes the pseudoinverse of A. Unfortunately, convergence to the true solution is not guaranteed since the sets are not convex. In what follows, we review a few successful alternatives to this approach.

PhaseLift

As before, we let a_i^* denote the ith row of A, and take $\mathcal{A}$ to be the function that maps Hermitian matrices X to vectors $y \in \mathbb{R}^m$ such that $y[i] = a_i^* X a_i$. It is easy to verify that $\mathcal{A}$ is linear, and considering (2), we also have $\mathcal{A}(xx^*) = |Ax|^2$ for every $x \in \mathbb{C}^n$. As such, the following program is equivalent to (1) when $S = \mathbb{C}^n$:

$$\text{find } X \text{ subject to } \mathcal{A}(X) = y,$$
$$X \succeq 0, \text{rank}(X) = 1. \quad (3)$$

Discarding the rank constraint produces a convex relaxation of the phase retrieval problem:

$$\text{find } X \text{ subject to } \mathcal{A}(X) = y, X \succeq 0. \quad (4)$$

If this relaxation is tight (i.e., every solution satisfies rank$(X) = 1$), then the relaxation solves the phase retrieval problem. If the relaxation is not tight, then we might instead minimize the trace of X subject to $\mathcal{A}(X) = y$ and $X \succeq 0$ so as to encourage X to have low rank. Both options are members of a family of convex relaxations of (3) called PhaseLift. An alternative relaxation, referred to as PhaseCut, is obtained by separating the measurements into an amplitude and phase component, and optimizing only the phase [12].

Amazingly, PhaseLift (4) is typically tight and robust to noise whenever A is "sufficiently random" and the number of measurements m is appropriately large. To see this, we first note that the set of complex Hermitian matrices is an n^2-dimensional vector space over the real numbers, by using a basis consisting of $n(n+1)/2$ real Hermitian matrices $E_{jk} + E_{kj}$ and $n(n-1)/2$ imaginary Hermitian matrices $i(E_{jk} - E_{kj})$. Thus, every Hermitian X is uniquely determined by $\mathcal{A}(X)$, if $\{a_i a_i^*\}_{i=1}^m$ is a spanning set; this typically occurs when $m \geq n^2$ and the a_is are drawn at random. To get away with fewer measurements, we leverage the fact that the solution we seek satisfies $X \succeq 0$. Then, it can be shown that (4) is typically tight when $m = \Omega(n \text{ polylog } n)$, provided $A \in \mathbb{C}^{m \times n}$ is drawn from an appropriate random distribution. For example, the entries of A may be drawn independently

from a complex Gaussian distribution, or A may be composed of $n \times n$ blocks of the form FD_j, where F is the $n \times n$ discrete Fourier transform matrix and D_j is a diagonal matrix with random diagonal entries from an acceptable distribution (the latter case models the structured illumination application of phase retrieval).

Wirtinger flow

While PhaseLift allows solving the phase retrieval problem in polynomial time (say, with an interior-point method), such methods scale poorly with the problem size due to the lifting operation, leading one to seek alternative solvers. To this end, we consider a different program:

$$\text{minimize} \quad \sum_{i=1}^{m} (|a_i^* x|^2 - y_i)^2. \quad (5)$$

Observe that (5) is equivalent to the phase retrieval problem (1) when $S = \mathbb{C}^n$. Unfortunately, since (5) is not convex, we expect to encounter local minima when attempting to solve it. In addition, this particular objective function has a continuum of global optimizers: The true solution x_0 induces a circle of global optimizers $[x_0] = \{e^{i\phi} x_0 : \phi \in [0, 2\pi)\}$. Perhaps surprisingly, (5) admits a fast initialization of gradient descent that allows for convergence to this circle provided A is sufficiently random. This gradient descent iteration is called *Wirtinger flow* because the gradient is conveniently expressed in terms of Wirtinger derivatives [3].

The convergence of Wirtinger flow is established by first showing that initializations sufficiently close to $[x_0]$ yield convergence by verifying a local convexity-type property. Next, a good initialization is found. Suppose the rows $\{a_i^*\}_{i=1}^{m}$ of A are complex Gaussian. Then a simple moment calculation reveals that

$$\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} y_i a_i a_i^*\right] = I + 2x_0 x_0^*,$$

meaning the true solution x_0 is a leading eigenvector of the expected matrix. Furthermore, $1/m \sum_{i=1}^{m} y_i a_i a_i^*$ is typically

spectrally close to its expectation, and so its leading eigenvector (suitably scaled) is close to $[x_0]$. With this initialization, gradient descent converges linearly to $[x_0]$ when $m = \Omega(n \log n)$ and A has complex Gaussian entries; a variant of the gradient descent iteration exhibits similar performance when $m = \Omega(n \log^4 n)$ and A is composed of $n \times n$ blocks of the form FD_j (again, following the structured illumination model). Figure 2 shows a comparison between the computation time required for PhaseLift and for Wirtinger flow.

Sparse phase retrieval

Now suppose that the unknown vector x_0 is known to be k-sparse. If we were given Ax_0 instead of $|Ax_0|^2$, then we could leverage the now-rich theory of compressed sensing to reconstruct x_0 provided A is an $m \times n$ random matrix with $m = \Omega(k \text{ polylog } n)$; see [1] for a short introduction to this theory. We aspire to reconstruct $[x_0]$ from $|Ax_0|^2$ with similar requirements on A. We discuss two algorithms along these lines but note that their performance is strictly worse than the desired $m = \Omega(k \text{ polylog } n)$. Indeed, achieving this performance with complex Gaussian or sufficiently random matrices A remains an open problem.

In compressed sensing, one of the most popular reconstruction algorithms given $Ax_0 = y$ minimizes $\|x\|_1$ subject to $Ax = y$ [4]. In phase retrieval, we receive linear measurements of $x_0 x_0^*$, and since $x_0 x_0^*$ is sparse, it makes sense to minimize $\|X\|_1$ subject to $\mathcal{A}(X) = y$. We also want to encourage X to be rank-1, leading to the following variation of PhaseLift:

$$\text{minimize} \|X\|_1 + \lambda \text{Tr}[X]$$
$$\text{subject to } \mathcal{A}(X) = y, X \succeq 0. \quad (6)$$

If the entries of A are complex Gaussian, then for an appropriate choice of λ, (6) typically recovers $X = x_0 x_0^*$ provided $m = \Omega(k^2 \log n)$. Drawing from compressed sensing-based intuition, the k^2 here comes from the fact that $x_0 x_0^*$ is k^2-sparse, and so the barrier to improving this sample complexity is perhaps an artifact of the lifting approach.

Following the motivation of Wirtinger flow, we seek a faster alternative to this semidefinite program. Let us reformulate (1) in the case where S is the set of k-sparse vectors:

$$\text{minimize} \sum_{i=1}^{m} (|a_i^* x|^2 - y_i)^2$$
$$\text{subject to } \|x\|_0 \leq k. \quad (7)$$

Here, $\|x\|_0$ denotes the number of nonzero entries in x; note the similarity to (5). When $\{a_i\}_{i=1}^{m}$ are complex Gaussian, the solution to (7) typically
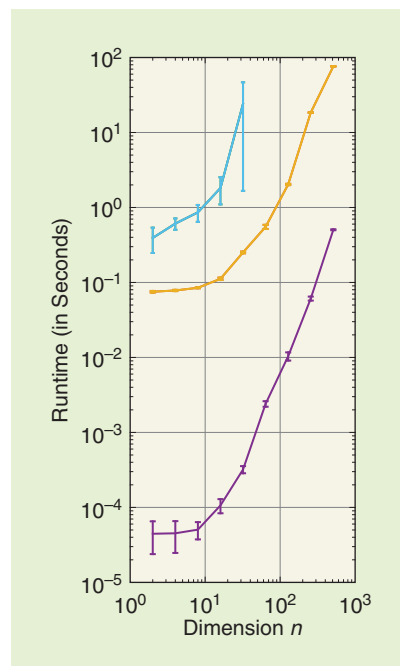
FIGURE 2. The runtime comparison between PhaseLift (blue), Wirtinger flow (orange), and least squares (purple) assuming the known phase. For each dimension $n \in \{2^1, \ldots, 2^9\}$, we perform 20 iterations of the following experiment: Draw a $4.5\, n \times n$ matrix A with independent entries with complex Gaussian distribution $\mathcal{N}(0, 1/2) + i\mathcal{N}(0, 1/2)$ and a signal $x \in \mathbb{C}^n$, also with independent and identically distributed complex Gaussian entries. Compute $z = Ax$ and $y = |z|^2$. Then reconstruct x from z by MATLAB's built-in implementation of least squares, and reconstruct x from y up to global phase using Wirtinger flow and PhaseLift. (For PhaseLift, we solve the semidefinite program using TFOCS v1.3 release 2 [2]; the runtime was prohibitively long for $n > 2^5$.) After conducting all 20 iterations, we plot the average runtime along with error bars that illustrate one standard deviation. As expected, the least-squares solver (which enjoys a phase "oracle") is faster than the phase retrieval solvers. For larger dimensions, Wirtinger flow appears to be about 100 times slower than least-squares, whereas PhaseLift is even slower.

reconstructs the true solution x_0 provided $m = \Omega(k \log(n/k) \log k)$, and furthermore, the reconstruction is robust to noise. To solve this program, we may apply an algorithm called Greedy Sparse Phase Retrieval (GESPAR), which iteratively improves a guess $K \subseteq \{1, \ldots, n\}$ of the support of x_0. For each guess, the objective function reduces to an instance of (5). Therefore, GESPAR optimizes locally to produce an estimate x such that $\text{supp}(x) \subseteq K$, and then updates K by swapping the member of K that contributes least to x with the index outside of K that contributes most to the (negative) gradient of the objective function. This iteration terminates when the swap fails to produce an improvement. See [9] for an implementation of GESPAR. In practice, GESPAR is prone to local minima, and so one must attempt multiple trials with different initializations before succeeding. Still, GESPAR is much faster than the semidefinite programming alternative, and it empirically performs well when $m = \Omega(k^3)$; for comparison, the Fienup-type alternative is about twice as fast, but only performs well when m is much larger. Recently, GESPAR has been used to solve phase retrieval problems in coherent diffraction imaging and in ankylography.

Conclusions

There have been several interesting developments in phase retrieval over the past decade from both the optics and signal processing communities. We discussed various settings in which signals are uniquely determined by intensity measurements, as well as new algorithms for reconstructing signals from such measurements. Understanding conditions under which recovery from Fourier phase retrieval measurements is possible, as well as developing practical measurement systems from which the signal can be recovered efficiently are

> **Understanding conditions under which recovery from Fourier phase retrieval measurements is possible, as well as developing practical measurement systems from which the signal can be recovered efficiently are some of the important directions for future research.**

some of the important directions for future research. We believe that this field will continue to grow and have a significant impact on optical imaging.

Authors

Yonina C. Eldar (yonina@ee.technion. ac.il) is a professor in the Department of Electrical Engineering at the Technion–Israel Institute of Technology, Haifa, where she holds the Edwards Chair in Engineering. Her research interests are in the areas of signal processing and optimization and their applications to communications, radar, biomedical imaging, and optics. She is an IEEE Fellow and a member of the Young Israel Academy of Science and the Israel Committee for Higher Education.

Nathaniel Hammen (nhammen@ gmail.com) is a postdoctoral researcher in the Department of Mathematics and Statistics at the Air Force Institute of Technology, Wright-Patterson, Air Force Base, Ohio. His research interests include applied harmonic analysis and signal processing.

Dustin G. Mixon (dustin.mixon@ afit.edu) is an assistant professor of mathematics in the Department of Mathematics and Statistics at the Air Force Institute of Technology, Wright-Patterson, Air Force Base, Ohio. His research interests include applied harmonic analysis, mathematical signal processing, and mathematical data science.

References

[1] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[2] S. Becker, E. J. Candès, and M. Grant. (Sept. 2011). Templates for convex cone problems with applications to sparse signal recovery. *Math. Program. Comput.* [Online]. *3(3)*, 165–218. Available: http://cvxr.com/tfocs/

[3] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1985–2007, Feb. 2015.

[4] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, UK: Cambridge Univ. Press, 2015.

[5] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 30, no. 2, pp. 140–154, Apr. 1982.

[6] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," submitted for publication.

[7] R. Juri, A. Chebira, Y. M. Lu, and M. Vetterli. (2013). Phase retrieval for sparse signals: Uniqueness conditions. [Online]. Available: arXiv: 1308.3058

[8] D. G. Mixon. (2014, Mar. 7). Phase transitions in phase retrieval. [Online]. Available: arXiv: 1403.1458

[9] Y. Shechtman, A. Beck, and Y. C. Eldar, GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Processing*, vol. 62, no. 4, pp. 928–938, Jan. 2014.

[10] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," *IEEE Signal Processing Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[11] C. Vinzant, "A small frame and a certificate of its injectivity," in *Proc. Int. Conf. SampTA*, Washington, DC, 2015, pp. 197–200.

[12] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut, and complex semidefinite programming," *Math. Program.*, vol. 149, nos. 1–2, pp. 47–81, Feb. 2015.

SP

new

IEEE TRANSACTIONS ON

SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

Now accepting paper submissions

The new *IEEE Transactions on Signal and Information Processing over Networks* publishes high-quality papers that extend the classical notions of processing of signals defined over vector spaces (e.g. time and space) to processing of signals and information (data) defined over networks, potentially dynamically varying. In signal processing over networks, the topology of the network may define structural relationships in the data, or may constrain processing of the data. Topics of interest include, but are not limited to the following:

Adaptation, Detection, Estimation, and Learning
o   Distributed detection and estimation
o   Distributed adaptation over networks
o   Distributed learning over networks
o   Distributed target tracking
o   Bayesian learning; Bayesian signal processing
o   Sequential learning over networks
o   Decision making over networks
o   Distributed dictionary learning
o   Distributed game theoretic strategies
o   Distributed information processing
o   Graphical and kernel methods
o   Consensus over network systems
o   Optimization over network systems

Communications, Networking, and Sensing
o   Distributed monitoring and sensing
o   Signal processing for distributed communications and networking
o   Signal processing for cooperative networking
o   Signal processing for network security
o   Optimal network signal processing and resource allocation

Modeling and Analysis
o   Performance and bounds of methods
o   Robustness and vulnerability
o   Network modeling and identification

Modeling and Analysis (cont.)
o   Simulations of networked information processing systems
o   Social learning
o   Bio-inspired network signal processing
o   Epidemics and diffusion in populations

Imaging and Media Applications
o   Image and video processing over networks
o   Media cloud computing and communication
o   Multimedia streaming and transport
o   Social media computing and networking
o   Signal processing for cyber-physical systems
o   Wireless/mobile multimedia

Data Analysis
o   Processing, analysis, and visualization of big data
o   Signal and information processing for crowd computing
o   Signal and information processing for the Internet of Things
o   Emergence of behavior

Emerging topics and applications
o   Emerging topics
o   Applications in life sciences, ecology, energy, social networks, economic networks, finance, social sciences, smart grids, wireless health, robotics, transportation, and other areas of science and engineering

Editor-in-Chief: Petar M. Djurić, Stony Brook University (USA)

To submit a paper, go to: https://mc.manuscriptcentral.com/tsipn-ieee

IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society
http://www.signalprocessingsociety.org/

IEEE COMPUTER SOCIETY

STANDARDS IN A NUTSHELL

Ronggang Wang, Tiejun Huang, Sang-hyo Park, Jae-Gon Kim,
Euee S. Jang, Cliff Reader, and Wen Gao

The MPEG Internet Video-Coding Standard

To address the diversified needs of the Internet, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) started the project of Internet video coding (IVC) in July 2011. It is anticipated that any patent declaration associated with the baseline profile of this standard will indicate that the patent owner is prepared to grant a free-of-charge license to an unrestricted number of applicants worldwide. IVC has been developed in MPEG from scratch by combining well-known existing technology elements and new contributions with free-of-charge licenses. Recently, IVC's compression performance has been determined to be approximately equal to that of the advanced video coding high profile (AVC HP) for typical operational settings, both for streaming and low-delay applications. In June 2015, the IVC project was approved as ISO/IEC 14496-33 (MPEG-4 IVC). It is believed that this standard can be highly beneficial for video services in the Internet domain. This article describes the main coding tools adopted in IVC; evaluates its performance compared with web video coding (WVC), video coding for browsers (VCB), and AVC HP; and provides the subjective comparison results between IVC and AVC HP.

Background

Video-coding standards lie at the heart of every aspect of video in our lives,

including broadcast television, streaming video on the Internet, digital cinema, movies on optical disks, home movies, and video conferencing. The most famous image-coding standards, JPEG and JPEG 2000, are royalty free (Type-1). To address the diversified needs of the Internet, ISO/IEC JTC1/SC29/WG11 MPEG issued the call for proposals (CfP) for the type-1 standard of IVC [1] in July 2011.

Three codecs respond to the CfP, including WVC [2], VCB [3] and IVC [4]. WVC is proposed jointly by Apple, Cisco, Fraunhofer Heinrich Hertz Institute, Magnum Semiconductor, Polycom, and Research in Motion Ltd. The coding tools in constrained AVC baseline plus hierarchical P frames are adopted in WVC. VCB is proposed by Google, and the coding tools are the same as those in VP8. IVC is proposed by several universities (including Peking University, Tsinghua University, Zhejiang University, Hanyang University, Korea Aerospace University, and the University of Electronic Science and Technology of China), and its coding tools are developed from scratch. These three codecs try to meet the intellectual property rights policy requirement of IVC with different strategies [5]. WVC and VCB expect the patent holders would like to grant a free-of-charge license for the internet application scenarios. IVC aims to create a new platform by utilizing coding tools for which patents have expired and new contributions with free-of-charge licenses.

In June 2015, the compression performance of IVC was determined to be approximately equal to that of AVC HP for typical operational settings, both for streaming and low-delay applications [6], and the IVC project was formally approved as ISO/IEC 14496-33 (MPEG-4 IVC).

Coding tools in IVC

Similar to previous standards, IVC is based on the traditional hybrid transform and motion compensation framework, as shown in Figure 1. Only progressive scan sequences are supported by IVC, and the input format of an IVC encoder is YUV420. The basic coding unit is macroblock. A macroblock consists of a 16×16 luminance block and two corresponding 8×8 chroma blocks. The input macroblock can either be coded with intramode or intermode, which is decided by the mode decision. If intramode is selected, the blocks in a macroblock are first predicted with intraprediction, and then the residues are processed with the modules of transform, quantization, and entropy coding, sequentially. At last, the blocks are reconstructed and processed with deblocking to obtain the decoded blocks. The decoded blocks are put into a forward frame buffer or backward frame buffer for being referred to by motion compensation. Otherwise, if intermode is selected for the current macroblock, motion compensation is invoked to get the interpredictor, and the motion vectors used in the motion compensation
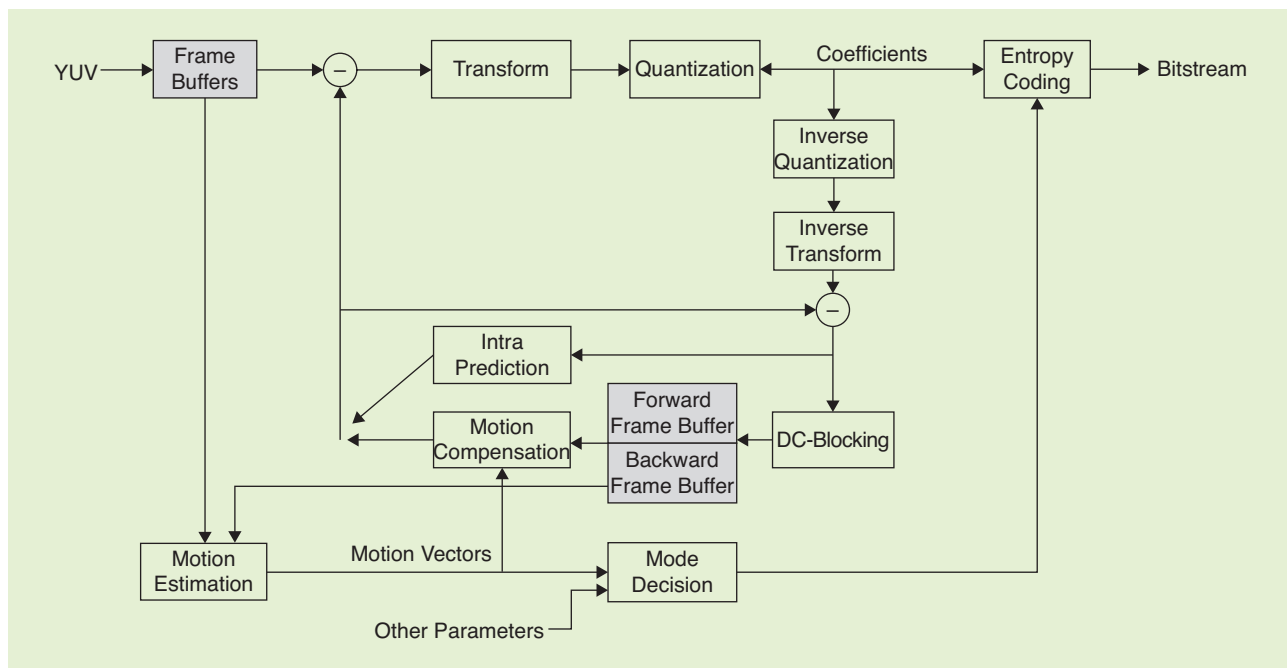
FIGURE 1. A block diagram of an IVC encoder.

are derived with motion estimation. The main coding tools of the IVC are described in the following paragraphs.

Intraprediction

Spatial domain intraprediction is used in intramacroblock coding. The decoded boundary samples of adjacent blocks are used as reference data for the spatial prediction. For the luma component in a macroblock, it can be coded by either one 16×16 macroblock partition or four 8×8 macroblock partitions. Each 8×8 macroblock partition can be further coded with four 4×4 macroblock partitions. The four prediction modes (i.e., Intra_Vertical, Intra_Horizontal, Intra_DC, Intra_Down_left, and Intra_Down_right) shown in Figure 2 can be used for a 16×16 macroblock partition, an 8×8 macroblock partition or a 4×4 macroblock partition. For the chroma components in a macroblock, they are only coded by an 8×8 macroblock partition, and four prediction modes (i.e., Intra_Chroma_DC, Intra_Chroma_Horizontal, Intra_Chroma_Vertical, and Intra_Chroma_Plane) can be used for each 8×8 macroblock partition. The intraprediction mode for each macroblock partition is directly coded into the bitstream without prediction

from the intraprediction modes of neighboring macroblock partitions.

Interprediction

An intermacroblock can be coded by one 16×16 macroblock partition, two 16×8 macroblock partitions, two 8×16 macroblock partitions, or four 8×8 macroblock partitions. Five interprediction modes (i.e., skip, forward prediction, backward prediction, multiple hypothesis, and symmetrical prediction) are defined for intermacroblock partitions. For each intermacroblock partition, one to four modes are available to be selected depending on the current picture coding type and partition size, as shown in Table 1.

1) Skip: Skip mode is one of the prediction modes that skips the encoding of all syntax elements except its mode-type information. If skip mode is selected for the current macroblock, both the motion vector difference and prediction residuals are forced to zeros.

2) Forward prediction: Forward prediction mode uses only one block in one of its forward reference pictures to predict the current macroblock partition. The motion vector difference and the prediction residuals of
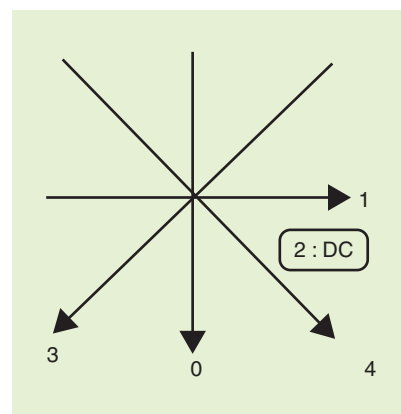
FIGURE 2. The intraprediction modes for the luma component.

the current macroblock partition are transmitted in the bitstream.

3) Backward prediction: Backward prediction mode uses only one block in its backward reference picture to predict the current macroblock partition. The motion vector difference and the interprediction residuals of the current macroblock partition are transmitted in the bitstream.

4) Multiple hypothesis: In this mode, as shown in Figure 3, the interpredictor of the macroblock partition (c) is derived by averaging two forward predictors (H1 and H2) [10]. The motion vector (MV1) of the first

Table 1. The interprediction modes for each type of macroblock partition.

MBPart \ Mode	Skip	Forward Prediction	Backward Prediction	Multiple Hypothesis	Symmetrical Prediction
P_16x16	√	√		√	
P_16x8		√		√	
P_8x16		√		√	
P_8x8		√		√	
B_Skip	√				
B_16x16	√	√	√		√
B_16x8		√	√		√
B_8x16		√	√		√
B_8x8	√	√	√		√

FIGURE 3. An illustration of multiple hypothesis.
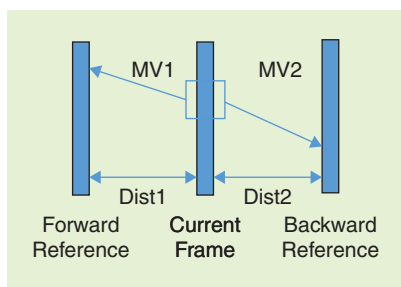
FIGURE 4. An illustration of symmetrical prediction.
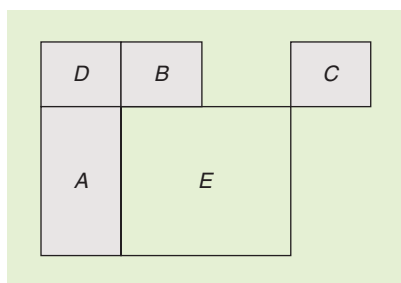
FIGURE 5. The neighboring macroblock partitions for MVP.

forward predictor is the predicted motion vector derived by the process of motion vector prediction, and the motion vector (MV2) of the second forward predictor is derived

by the motion estimation constrained with the first predictor. Only the motion vector difference of the second predictor and the interprediction residuals are transmitted in the bitstream.

5) Symmetrical prediction: Symmetrical mode averages one forward predictor and a backward predictor to get the final interpredictor for the macroblock partition, as shown in Figure 4. The motion vector (MV1) of the forward predictor is derived by forward motion estimation. The backward motion vector (MV2) is derived by scaling the forward motion vector. The scaling factor is decided by the distance (Dist1) between the forward reference frame and the current frame and the distance (Dist2) between the backward reference frame and the current frame. Only the motion vector difference of the forward predictor and the interprediction residuals are transmitted in the bitstream.

Motion vector prediction

A motion vector is predicted, and only the motion vector difference (between the motion vector and its prediction) is coded into bitstream. To predict the motion vector of the current macroblock partition (given by E), its four neighboring macroblock partitions of left, above, left-above, and right-above (their motion vectors are given by A, B, D, and C, respectively) are used, as shown in Figure 5. If there is no motion vector

for one macroblock partition (intramacroblock partition) or the macroblock partition has not been reconstructed, the motion vector of this macroblock partition is set as a zero vector. When the partitions A, B, C, and D are all unavailable, the predictor of E is set as a zero vector. When only one of A, B, C, or D is available, the predictor of E is set as that available motion vector. Otherwise, if C is unavailable, then it is replaced with D, and A, B, and C are used to predict E by the following process.

First, the signs of each horizontal component of A, B, and C are checked. If the sign of one motion vector (denoted by X) is different from the other two, then this motion vector X is excluded from the motion vector prediction process, and the predictor of the horizontal component is the averaging of the horizontal component of the other two motion vectors. Otherwise, the Euler distance of the horizontal component of each pair of neighboring motion vectors is calculated, the motion vector pair with the smallest Euler distance is selected, and the predictor of the horizontal component is the average of the horizontal component of the selected motion vector pair. The vertical component of E is predicted in the same way as the horizontal component.

Subpel interpolation

Quarter-pel motion compensation is adopted for the luma component. A two-dimensional separable Lanczos filter is used to generate the subpel position values [11], as shown in Figure 6. Three one-dimensional filters, F1, F2, and F3, are used to generate the subpel values as follows:

$$a_{x,y} = \sum_z F1(z) * A_{x+z,y}, b_{x,y}$$
$$= \sum_z F2(z) * A_{x+z,y}, c_{x,y}$$
$$= \sum_z F3(z) * A_{x+z,y} \qquad (1)$$

$$d_{x,y} = \sum_z F1(z) * A_{x,y+z}, h_{x,y}$$
$$= \sum_z F2(z) * A_{x,y+z}, n_{x,y}$$
$$= \sum_z F3(z) * A_{x,y+z} \qquad (2)$$

$$e_{x,y} = \sum_z F1(z) * a_{x,y+z}, \ i_{x,y}$$
$$= \sum_z F2(z) * a_{x,y+z}, \ p_{x,y}$$
$$= \sum_z F3(z) * a_{x,y+z} \qquad (3)$$

$$f_{x,y} = \sum_z F1(z) * b_{x,y+z}, \ j_{x,y}$$
$$= \sum_z F2(z) * b_{x,y+z}, \ q_{x,y}$$
$$= \sum_z F3(z) * b_{x,y+z} \qquad (4)$$

$$g_{x,y} = \sum_z F1(z) * c_{x,y+z}, \ k_{x,y}$$
$$= \sum_z F2(z) * c_{x,y+z}, \ r_{x,y}$$
$$= \sum_z F3(z) * c_{x,y+z}, \qquad (5)$$

where A is the integer pixel, $a–r$ are the subpels, x and y are the horizontal and vertical coordinates of subpels, F1–F3 are the one-dimensional interpolation filters given in Table 2, and z is the index of filter coefficients. Either 4-tap, 6-tap or 10-tap filters can be used as the interpolation filter depending on the spatial resolution of the given sequence. The 4-tap filters are used on a sequence that is larger than 1080p, the 6-tap filters are used on a sequence that is between 1080p and 720p, and the 10-tap filters are used on a sequence that is smaller than 720p.

Eighth-pel motion compensation is adopted for chroma components. A two-dimensional separable filter similar to that of the luma component is used to generate the subpel position values. The 4-tap filters specified in Table 3 are used for calculating the subpels of chroma.

Reference frames

Figure 7 illustrates the relationship between the current frame with its reference frames in forward direction and/or backward direction. The interprediction process can refer to multiple reference frames in the forward direction that are used in the forward prediction and the multiple-hypothesis modes. In the current IVC Test Model (ITM), the number of forward reference frames can be customized up to eight. Let the temporal position of the current frame be t, and the current frame refers to the reference pictures at the following locations, as shown in Figure 7: $t-1, t-2$, and $t-4*n$ (for $n = 1, 2, 3, …$). On the other hand, the macroblock coded with the backward prediction mode, the skip mode, or the symmetrical mode, can refer to only one backward reference frame (i.e., the t + 1 frame in Figure 7).

In low-delay coding cases, nonreference P-frame [13] coding uses three different levels of quantization parameter (QP) values for each group of pictures. A coding structure with nonreference P-frame coding is shown in Figure 8. As a typical example of QP setting, the lowest value of QP is assigned to the P frame of P0 and P4, and then a larger QP value is assigned to the P frame of P2, and the largest QP is assigned to the nonreference frames of P1 and P3. As a result, a three-level hierarchical coding structure in terms of QP values is used in nonreference P-frame coding. It can be

FIGURE 6. Integer samples (shaded blocks) and fractional sample positions (unshaded blocks) for luma interpolation.

Table 2. Interpolation filter coefficients for luma.			
Position (filter)	4-Tap	6-Tap	10-Tap
1/4(F1)	{−6, 56, 15, −1}/64	{2, −9, 57, 17, −4, 1}/64	{1, −2, 4, −10, 57, 19, −7, 3, −1, 0}/64
2/4(F2)	{−4, 36, 36, −4}/64	{2, −9, 39, 39, −9, 2}/64	{1, −2, 5, −12, 40, 40, −12, 5, −2, 1}/64
3/4(F3)	{−1, 15, 56, −6}/64	{1, −4, 17, 57, −9, 2}/64	{0, −1, 3, −7, 19, 57, −10, 4, −2, 1}/64

Table 3. Interpolation filter coefficients for chroma components.

Position	Filter Coefficients
1/8	{−4, 62, 6, 0}/64
2/8	{-6, 56, 15, −1}/64
3/8	{−5, 47, 25, −3}/64
4/8	{−4, 36, 36, −4}/64
5/8	{−3, 25, 47, −5}/64
6/8	{−1, 15, 56, −6}/64
7/8	{0, 6, 62, −4}/64

adaptively determined whether non-reference P-frame coding is used (e.g., P0, P1, P2, and P3) or not (e.g., P4, P5, P6, and P7) for every four frames based on the temporal correlation that is measured by the amount of motion and bitrate in an adaptive manner.

Adaptive block size transform

Adaptive block size transforms are applied on the prediction residuals to reduce the spatial redundancy. The transform block size can be either 16×16, 8×8, or 4×4. Integer transforms of 16×16, 8×8, and 4×4 are derived by scaling and rounding the DCT cores of 16×16, 8×8, and 4×4 respectively [8], [9]. For the intermacroblock partition, if the partition size is 16×16, a 16×16 transform is applied; otherwise, an 8×8 transform is applied on each 8×8 block within this macroblock partition. For the intramacroblock partition, the transform size is coupled with the

partition size. If the partition size is 16×16, then a 16×16 transform is applied; else if the partition size is 8×8, an 8×8 transform is applied; otherwise, a 4×4 transform is applied. The inverse transform processes are specified as

$$R_{N \times N} = (T^T{}_{N \times N} * C_{N \times N} * T_{N \times N} + (1 << \text{left\_shift})) >> \text{right\_shift}, \quad (6)$$

where $R_{N \times N}$ is the $N \times N$ residual matrix, $T_{N \times N}$ is the $N \times N$ transform matrix, and $C_{N \times N}$ is the transformed $N \times N$ matrix. For the 16×16, 8×8 and 4×4 inverse transforms, the parameters of {N, left_shift, right_shift} are set as {16, 14, 15}, {8, 4, 5}, and {4, 16, 17} respectively.

Logarithmic domain arithmetic coding

Coding a data symbol involves the following steps: 1) binarization, 2) context model selection, and 3) arithmetic encoding. For a given nonbinary valued syntax element, it is uniquely mapped to a binary sequence, a so-called bin string. Each of the given binary decisions, which are referred to as a *bin in the sequence*, enters the context modeling stage where a context model is selected. Then, the bin value, along with its associated context model, is passed to the regular coding engine or bypass coding where the final stage of arithmetic encoding together with a subsequent context updating takes

place. Binary arithmetic coding is based on the principle of recursive interval subdivision that involves the following elementary multiplication operation. Suppose that an estimate of the probability P_{MPS} $(0.5, 1)$ of the most probable symbol (MPS) is given, and the given interval is represented by its lower bound, L, and its width (range), R. Based on these settings, the given interval is subdivided into two subintervals: one interval of width, R_{MPS} and R_{LPS}.

$$R_{MPS} = R * P_{MPS} \quad (7)$$
$$R_{LPS} = R - R_{MPS}. \quad (8)$$

An arithmetic coding method in a the logarithmic domain is adopted as the entropy coding engine [14]. In the binary arithmetic coder, the multiplication in the R_{MPS} calculation is substituted with addition in the logarithm domain. When an MPS happens, the renewal of range is given as

$$\text{Log}R_{MPS} = \text{Log}R + \text{Log}P_{MPS}. \quad (9)$$

Assume the value of $\log R$ is represented by its integer part s1 and fractional part t1, and the value of $\text{Log}(R_{MPS})$ is represented by its integer part s2 and fractional part t2, respectively. When an LPS happens, the range is updated as follows:

$$R_{LPS} = R - R_{MPS}$$
$$= 2^{-s1+t1} - 2^{-s2+t2} \approx 2^{-s1}*(1+t1)$$
$$- 2^{-s2}*(1+t2), \text{ since}$$
$$2^x \approx 1 + x(0 < x < 1). \quad (10)$$

So, R_{MPS} and R_{LPS} are all calculated by just addtions and shifts operations.
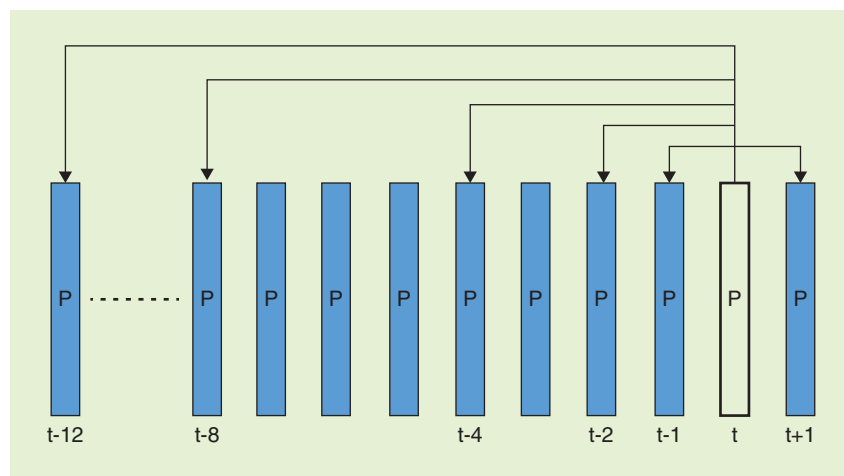
FIGURE 7. The relationship between the current frame with its reference frames.
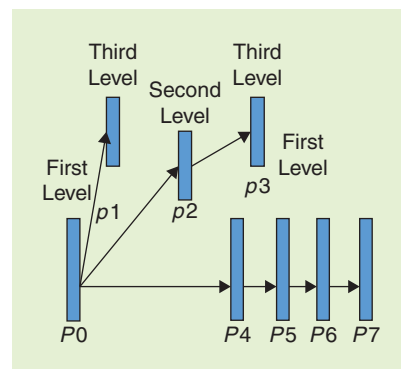
FIGURE 8. An adaptive nonreference P coding structure.

After the value of R_{LPS} is obtained, the renewed lower bound is updated. Then the renormalization process is carried out to guarantee that the most significant bit of the updated range value is always one. After one bin is encoded by the arithmetic coder, the estimated probability of the chosen context should also be updated. Actually, the probability of each context model is initialized to be 0.5 for both MPS and LPS at the start of coding. With the coding of some bins, the adaptive probability estimation of MPS on the logarithm domain is performed. The probability estimation is fulfilled using only additions/subtractions and shifts, as in the following formulas:

$$\begin{cases} \text{Log}P_{\text{MPS}} = \text{Log}_{\text{MPS}} + \text{Log}\,f, \\ \qquad \text{if}\,(\text{LPS happens}) \\ \text{Log}P_{\text{MPS}} = +(P_{\text{MPS}} >> cw), \\ \qquad \text{if}\,(\text{MPS happens}) \end{cases} \quad (11)$$

where f is equal to $(1 - 2^{-cw})$. Here, cw is the size of the sliding widow to control the speed of probability adaptation, and it is constant.

In summary, the arithmetic coder in IVC replaces the traditional multiplications for range updating and probability estimation updating with additions by combining the original domain and the logarithmic domain. Refer to [14] for detailed explanations.

Loop filtering

An expired patent deblocking filter [15] is utilized to process all 8×8 block edges of a picture to reduce the blocking artifact, except the edges at the boundary of the picture, as shown in Figure 9. This filtering process is performed on a macroblock basis after the completion of the picture-reconstruction process prior to the deblocking filter process for the entire decoded picture, with all macroblocks
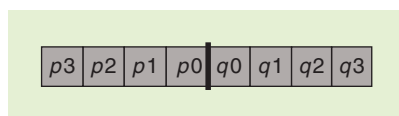
FIGURE 9. The horizontal or vertical edge sample of an 8×8 block.

in a picture processed in order of increasing macroblock addresses. The deblocking filter process is invoked for the luma and chroma components separately. For each macroblock, the vertical edges are filtered first, from left to right, and then the horizontal edges are filtered from top to bottom. Sample values above and to the left of the current macroblock that may have already been modified by the deblocking filter process operation on previous macroblocks shall be used as input to the deblocking filter process on the current macroblock and may be further modified during the filtering of the current macroblock. Sample values modified during the filtering of the vertical edges are used as input for the filtering of the horizontal edges for the same macroblock. If the level differences between the two border pixels in the same block and between the two border pixels in different adjacent blocks meet certain conditions, the edge is filtered. Here, the edge is defined as edges between all 8×8 blocks inside the macroblock and the upper and left edges of the current macroblock. There are three kinds of filtering methods: strong loop filtering, normal loop filtering, and weak loop filtering.

The conditions of loop filtering are:
- $\text{abs}(p0 - p1) < \beta\,\&\&\,\,\text{abs}(q0 - q1) < \beta$
- $\text{abs}(q0 - p0) > \text{abs}(p0 - p1)\&\&\,\text{abs}(q0 - p0) > \text{abs}(q0 - q1)$
- $\text{abs}(q0 - p0) < \alpha$
- $\text{abs}(p2 - p0) < \beta\,\&\&\,\text{abs}(q2 - q0) < \beta$
- $\text{abs}(p3 - p0) < \beta\,\&\&\,\text{abs}(q3 - q0) < \beta$
- $\text{abs}(p0 - p1) < \min(3, \beta)\,\&\&\,\text{abs}(q0 - q1) < \min(3, \beta)$.

As long as the first three conditions, 1–3, are all satisfied, the loop filtering is active. For the luma component, if condition 4) is also satisfied, the normal loop filter is applied. Strong loop filtering is used when all of the six conditions are satisfied for macroblock luminance edges. For the chroma components, only when conditions 4) and 6) are also satisfied, the normal loop filtering is used. The filtering processes are defined as follows:

- Weak loop filtering
$$P0 = (q0 - p0)/4 + p0;$$
$$Q0 = (p0 - q0)/4 + q0.$$

- Normal loop filtering
$$P1 = (3 * (p2 - q0) \\ + 4 * (p0 - q0) \\ + 8 * (q0 - p1))/16 + p1;$$
$$P0 = ((p2 - q0) \\ + 4 * (p1 - q0) + (q1 - p0) \\ + 9 * (q0 - p0))/16 + p0;$$
$$Q0 = ((q2 - p0) \\ + 4 * (q1 - p0) + (p1 - q0) \\ + 9 * (p0 - q0))/16 + q0;$$
$$Q1 = (3 * (q2 - p0) \\ + 4 * (q0 - p0) \\ + 8 * (p0 - q1))16 + q1.$$

- Strong loop filtering
$$P2 = (4 * (p0 - q0) \\ + 5 * (q0 - p2))/8 + p2;$$
$$P1 = (16 * (q0 - p1) \\ + 6 * (p2 - q0) \\ + 7 * (p0 - q0))/16 + p1;$$
$$P0 = (9 * (p2 - q0) \\ + 6 * (q2 - p0) \\ + 17 * (q0 - p0))/32 + p0;$$
$$Q0 = (9 * (q2 - p0) \\ + 6 * (p2 - q0) \\ + 17 * (p0 - q0))32 + q0;$$
$$Q1 = (16* (p0 - q1) \\ + 6* (q2 - p0) \\ + 7* (q0 - p0))/16 + q1;$$
$$Q2 = (4* (q0 - p0) \\ + 5* (p0 - q2)/8 + q2,$$

where $P0$ and $Q0$ are sample values obtained after the filtering process of $p0$ and $q0$, $P1$ and $Q1$ denote the sample values obtained after the filtering process of $p1$ and $q1$, and $P2$ and $Q2$ denote the sample values obtained after the filtering process of $p2$ and $q2$.

Performance evaluation

To evaluate the performance of IVC, we conduct a performance test on the newest ITM14 and compare it against

WVC, VCB, and AVC HP. The following approach had been agreed within the MPEG video group to enable comparison at approximately the same bit rate points [16]:

■ Produce bitstreams for each of the codec designs, which are within ±3% of the target bit rates for the sequences given in Table 4.

■ Allow QP (or quantizer step size) variation within a sequence within a periodic pattern of frame types (where frame types are differentiated by syntax or by a reference picture handling mechanism) within a sequence.

■ No per-sequence adaptation of the pattern of frame types could be used.

■ No sequence-specific tuning of coding parameters (such as enabling/disabling of special tools, certain modes, limitation of motion search range, etc.) was allowed to be used.

■ No rate control was allowed to be used.

■ No preprocessing was allowed to be used.

■ No postprocessing of the decoder output was allowed to be used.

Encoded bitstreams were provided for the following two constraint cases:

■ Constraint set 1 [CS1, also known as *random access* (*RA*)]: the structural delay of the processing units not larger than an 8-picture group of pictures and random access intervals of 1.1 seconds or fewer.

■ Constraint set 2 [CS2, also known as *low delay*, (*LD*)]: no structural delay of the processing units, with essentially no picture reordering between the decoder processing and the output.

The tests included AVC HP anchors produced by a JM 18.6 reference software encoder. Encoding of those anchors was performed under same configuration constraints as for the other encoders. Detailed encoding settings of WVC, VCB, IVC, and AVC HP can be found in [16].

Table 5 shows the performance of the three tested encoders according to the established Bjøntegaard delta bit rate (BD-BR) criterion [17], using AVC HP as the anchor. Positive percentages indicate a bit rate increase relative to the reference of the comparison. In the RA constraint cases (CS1), IVC clearly outperforms WVC and VCB in terms of BD-BR rate in overall average by 25.2% and 23.7%, respectively, and underperforms AVC HP by 10.4%. In the LD constraint cases (CS2), IVC clearly outperforms WVC

Table 4. Test sequences and rate points.

Class A [1920x1080p]	Rate 1	Rate 2	Rate 3	Rate 4
S03 Kimono, S04 Park Scene	1.6 Mbit/s	2.5 Mbit/s	4.0 Mbit/s	6.0 Mbit/s
S05 Cactus, S06 BasketballDrive	3.0 Mbit/s	4.5 Mbit/s	7.0 Mbit/s	10.0 Mbit/s
Class B [836x480p (WVGA)]	**Rate 1**	**Rate 2**	**Rate 3**	**Rate 4**
S08 BasketballDrill, S09 BQMall, S10 PartyScene, S11 RaceHorses	512 kbit/s	768 kbit/s	1.2 Mbit/s	2.0 Mbit/s
Class D [1280x720p]	**Rate 1**	**Rate 2**	**Rate 3**	**Rate 4**
S16 Johnny, S17 KristenAndSara, S18 FourPeople	384 kbit/s	512 kbit/s	850 kbit/s	1.5 Mbit/s

Table 5. Performances of IVC, VCB, and WVC relative to AVC HP.

Class	Sequences	RA			LD		
		WVC	VCB	IVC	WVC	VCB	IVC
Class A	Kimono	47.9%	24.5%	9.3%	37.0%	2.8%	−0.4%
	ParkScene	25.4%	38.0%	18.6%	17.0%	8.1%	4.5%
	Cactus	45.9%	32.2%	10.5%	25.4%	9.5%	3.2%
	BasketballDrive	41.5%	32.1%	15.3%	28.1%	8.6%	5.6%
Class B	BasketballDrill	28.5%	15.5%	6.6%	17.9%	17.6%	3.8%
	BQMall	30.2%	36.9%	5.5%	18.2%	7.3%	3.8%
	PartyScene	25.0%	32.5%	-5.7%	13.5%	5.1%	-7.3%
	RaceHorses	22.2%	20.4%	20.1%	16.1%	4.2%	7.7%
Class D	FourPeople	46.2%	67.8%	17.6%	27.5%	40.9%	12.0%
	Johnny	40.8%	41.2%	8.5%	22.9%	23.1%	11.1%
	KristenAndSara	37.6%	34.3%	7.9%	21.8%	15.8%	5.4%
	Average	35.6%	34.1%	10.4%	22.3%	13.0%	4.5%
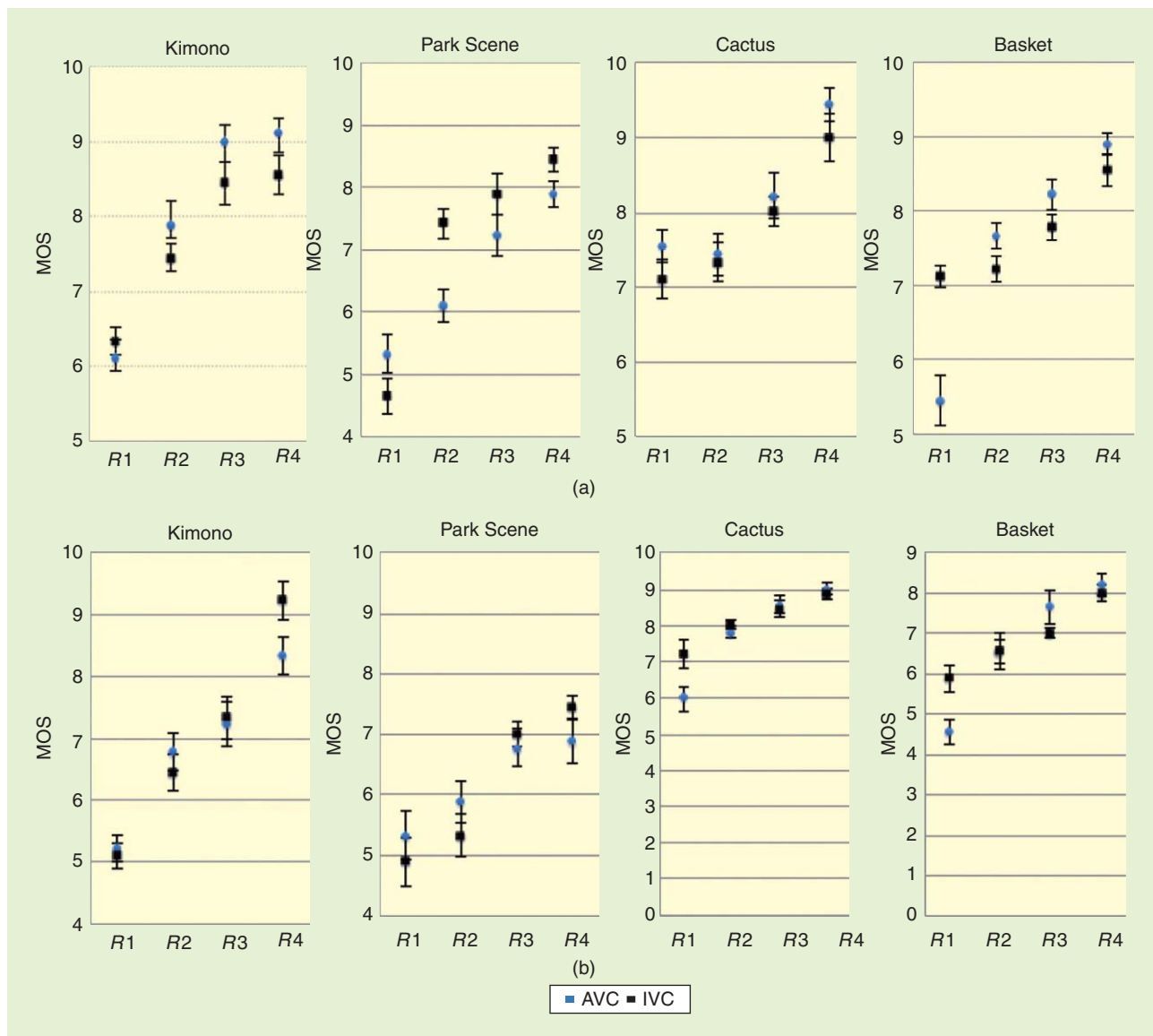
FIGURE 10. Subjective test results for 1080p sequences.

and VCB by 17.8% and 8.5%, respectively. IVC underperforms AVC HP by 4.5%. Although, in some sequences, e.g., Racehorses, IVC underperforms VCB. However, LD cases are mainly used in video-conference scenarios, and, for these video sequences of Class D, IVC is still clearly better than VCB.

In addition to objective evaluation, the MPEG video group has organized a viewing test to compare the subjective performance between IVC and AVC HP, and detailed test mythology and results can be found in [6]. From the results, it is concluded that IVC and AVC HP provide very similar results for the tested cases (in most cases with confidence intervals that are overlapping, in some cases IVC is visually better than AVC HP, and in some cases AVC HP is better than IVC). In general, IVC seems to have slightly better performance than the AVC HP anchors used in the LD cases. Figure 10 gives some examples of the test results on 1080p sequences.

As a general conclusion about the IVC performance evaluation, the results show that IVC is better than WVC and VCB and is comparable with AVC HP under both RA and LD constraints.

Conclusions

This article gives an overview of the coding tools adopted in the MPEG IVC standard, which is a Type-1 standard aiming at being used in various internet applications. The coding tools in IVC are developed from scratch and consist of well-known expired-patent techniques and new tools with free-of-charge licenses. During each coding tool adoption process, comprehensive prior art searches are conducted by the proponents. All prior art of IVC coding tools are recorded in an output document called the *collection of information related to adopted IVC technologies*, and it is updated after each new normative tool was adopted. Both the objective and subjective performance tests have been conducted within the MPEG video

group, and it has been determined that the performance of IVC is comparable with the AVC high profile. The next steps are to push this standard into the market and investigate new royalty-free technologies for the next version of IVC. It is anticipated that as existing patents for video coding tools expire, these tools may be added to IVC, further improving its performance.

Resources

MPEG resources

The MPEG homepage (http://wg11.sc29. org/) provides information on its past and present meeting documents. All of the input contributions and output documents of IVC can be found on the website.

Open documents

The website, http://mpeg.chiariglione. org/standards/mpeg-4/internet-video-coding, has links to all IVC opened publications. The IVC working documents are available, including the CD text, test models, performance reports, and prior art techniques related to IVC.

Acknowledgments

Authors

Ronggang Wang (rgwang@pkusz.edu.cn) is an associate professor at Peking University Shenzhen Graduate School. He is a cochair of the IVC AhG group.

Tiejun Huang (tjhuang@pku.edu.cn) is a professor at Peking University. He is an active proponent for the IVC project.

Sang-hyo Park (foriamweak@gmail.com) is a Ph.D. student at Hanyang University.

Jae-Gon Kim (jgkim@kau.ac.kr) is an associate professor at Korean Aerospace University.

Euee S. Jang (esjang@hanyang.ac.kr) is a professor at Hanyang University. He is a cochair of the IVC AhG group.

Cliff Reader (cliff@reader.com) is an adjunct professor at Peking University.

Wen Gao (wgao@pku.edu.cn) is a professor at Peking University.

References

[1] ISO/IEC, *Call for Proposals (CfP) for Internet Video Coding Technologies*, ISO/IEC JTC1/SC29/WG11 N12204, July 2011.

[2] K. Kolarov, D. Singer, D. Benham, G. Jouret, T. Wiegand, L. Winger, S. Botzko, J. Sampedro, and G. Martin-Cocher, *Joint Response to Call for Proposals (CfP) for Internet Video Coding Technologies*, ISO/IEC JTC1/SC29/WG11/M22492, Nov. 2011.

[3] H. Alvestrand, A. Grange, J. Luther, L. Bivolarski, and M. Raad, *Google Inc.'s Response to the CfP on Internet Video Technologies*, ISO/IEC JTC1/SC29/WG11 MPEG2013/M 29693, July 2013.

[4] R. Wang, X. Zhang, H. Lv, Z. Wang, X. Zhu, J. Chen, S. Ma, T. Huang, Y. He, and L. Yu, Ciff Reader, Wen Gao, *RFM2.0 for Internet Video Coding*, ISO/IEC JTC1/SC29/WG11 MPEG2012/M26716, Oct. 2012.

[5] K. Choi and E. S. Jang, "Royalty-free video coding standards in MPEG," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 145–148, Jan. 2014.

[6] B. Vittorio, *Report of Expert Viewing Visual Test of Internet Video Coding*, ISO/IEC JTC1/SC29/WG11 MPEG2015/N15428, June 2015.

[7] G. Bjontegaard, "Improvements to the Telenor Proposal for H.26L: More Block sizes for Prediction and RD Constrained Quantization of Transform Co-efficients," ITU: Telecommunications Standardization Sector, STUDY GROUP 16, Video Coding Experts Group (Question 15), Q15-H-10, 1999.

[8] W.-K. Cham, "Development of integer cosine transforms by the principle of dyadic symmetry," Proc. *Inst. Elect. Eng.*, Pt. 1, vol. 136, no. 4, pp. 276–282, 1989.

[9] C.-T. Chen, "Adaptive transform coding via quadtree-based variable block size DCT," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Glasgow, 1989, vol. 3, pp. 1854–1857.

[10] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* 1993, pp. 437–440.

[11] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, Aug. 1979.

[12] S.-H. Lee, S. Park, and E. S. Jang, *Improved Set of Reference Frames for Internet Video Coding (IVC)*, ISO/IEC JTC1/SC29/WG11 MPEG2015/m35748, Feb. 2015.

[13] X. Zhang, Y. Tian, R. Wang, T. Tian, et al., *Adaptive Non-reference P Optimization for Internet Video Coding*, ISO/IEC JTC1/SC29/WG11 MPEG2012/M27964, Jan. 2013.

[14] Q. Yu, W. Yu, P. Yang, J. Zheng, X. Zheng, and Y. He, "An efficient adaptive binary arithmetic coder based on logarithmic domain," *IEEE Trans. Image Processing*, vol. 24, no. 11, pp. 4225–4239, Nov. 2015.

[15] M. Honjo, "Method of correcting an image signal decoded in block units," U.S. Patent 5337088, Aug. 1993.

[16] MPEG Video, "Conditions for visual comparison of VCB, IVC and WVC codecs," MPEG 106, Geneva, Switzerland, Output Doc. N13943, Nov. 2013.

[17] G. Bjøntegaard, "*Calculation of average PSNR differences between RD-Curves*," ITU-Telecommunication Standard, Austin, TX, Doc. SG16 Q.6, VCEG-M33, Apr. 2001.

SP

ERRATA

In the article, "Bayesian Machine Learning: EEG/MEG Signal Processing Measurements" [1], the authors regret that the original acknowledgement of funding support from the National Natural Science Foundation of China was incorrect.

The correct acknowledgement as per Wei Wu is as follows: Wei Wu acknowledges support from the 973 Program of China (2015CB351703), the 863 Program of China (2012AA011601), the National Natural Science Foundation of China (61403144), the Guangdong Natural Science Foundation (2014A030312005 and S2013010013445), and the Steven and Alexandra Cohen Foundation.

The authors apologize for any confusion this may have caused.

Reference

[1] W. Wu, S. Nagarajan, and Z. Chen, "Bayesian machine learning: EEG/MEG signal processing measurements," *IEEE Signal Processing Mag.*, vol. 33, no. 1, pp. 14–36, Jan. 2016.

SP

WORLD'S FIRST

VISUAL INNOVATION AWARD

IEEE
2016.IEEEICIP.ORG

Finalists - Final Voting at IEEE ICIP 2016 on Tuesday September 27, 2016

Achin Bhowmik
Inventor
(intel) REALSENSE
TECHNOLOGY
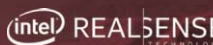
Steve Chen
Co-Founder
You Tube

Bill Dally
Chief Scientist
NVIDIA CUDA

Reed Hastings
Co-Founder
NETFLIX

Brendan Iribe
Co-Founder
oculus

Alex Kipman
Inventor
KINECT

Ren Ng
Founder
LYTRO

Keynote Speakers - Innovation Program at IEEE ICIP 2016
on Tuesday September 27, 2016. 12:30-17:00, Phoenix

Michael Antonov
Co-Founder
Oculus

Hanno Basse
CTO
20th Century Fox

Bo Begole
VP
Huawei

Achin Bhowmik
VP
Intel

Bill Dally
SVP
nVidia

C. –C. Jay Kuo
Dean's Professor
USC

Matthew Mengerink
VP
Google

Tim Milliron
VP
Lytro

Anthony Park
VP
Netflix

Jamie Shotton
Co-Inventor, Kinect
Microsoft

Raj Talluri
SVP
Qualcomm

Susie Wee
CTO
Cisco
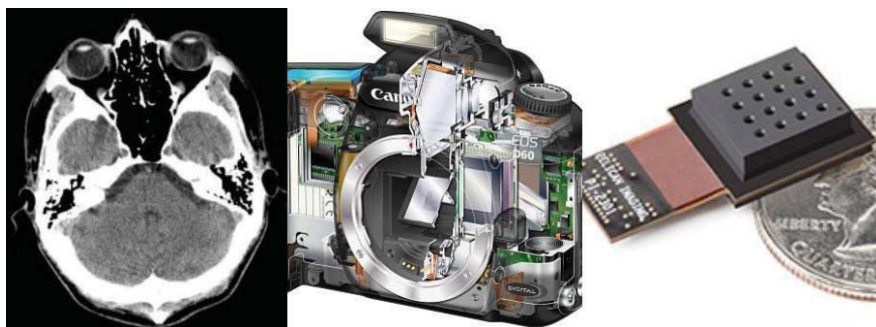
NEW!

IEEE TRANSACTIONS ON
COMPUTATIONAL IMAGING

Editor-in-Chief

W. Clem Karl
Boston University

Technical Committee
Charles Bouman
Eric Miller
Peter Corcoran
Jong Chul Ye
Dave Brady
William Freeman

The IEEE Transactions on Computational Imaging publishes research results where computation plays an integral role in the image formation process. All areas of computational imaging are appropriate, ranging from the principles and theory of computational imaging, to modeling paradigms for computational imaging, to image formation methods, to the latest innovative computational imaging system designs. Topics of interest include, but are not limited to the following:

Computational Imaging Methods and Model
- Coded imaging sensing
- Compressed sensing
- Sparse and low-rank models
- Learning-based models, dictionary methods
- Graphical image models
- Perceptual models

Computational Image Formation
- Sparsity-based reconstruction
- Statistically-based inversion methods
- Multi-image and sensor fusion
- Optimization-based methods; proximal iterative methods, ADMM

Computational Photography
- Non-classical image capture
- Generalized illumination
- Time-of-flight imaging
- High dynamic range imaging
- Plenoptic imaging

Computational Consumer Model
- Mobile imaging, cell phone imaging
- Camera-array systems
- Depth cameras, multi-focus imaging
- Pervasive imaging, camera networks

Computational Acoustic Imaging
- Multi-static ultrasound imaging
- Photo-acoustic imaging
- Acoustic tomography

Computational Microscopy
- Holographic microscopy
- Quantitative phase imaging
- Multi-illumination microscopy
- Lensless microscopy
- Light field microscopy

Imaging Hardware and Software
- Embedded Computing Systems
- Big data computational imaging
- Integrated hardware/digital design

Tomographic Imaging
- X-ray CT
- PET
- SPECT

Magnetic Resonance Imaging
- Diffusion tensor imaging
- Fast acquisition

Radar Imaging
- Synthetic aperture imaging
- Inverse synthetic imaging
- Terahertz imaging

Geophysical Imaging
- Multi-spectral imaging
- Ground penetrating radar
- Seismic tomography

Multi-spectral Imaging
- Multi-spectral imaging
- Hyper-spectral imaging
- Spectroscopic imaging

For more information on the **IEEE Transactions on Computational Imaging** see
http://www.signalprocessingsociety.org/publications/periodicals/tci/

No Page Charges in 2016 and Fast Turn Around!

IEEE Signal Processing Society

EMB CO

GRSS

ADVERTISING & SALES

© GRAPHIC STOCK

The Advertisers Index contained in this issue is compiled as a service to our readers and advertisers: the publisher is not liable for errors or omissions although every effort is made to ensure its accuracy. Be sure to let our advertisers know you found them through *IEEE Signal Processing Magazine*.

CVR 2
IEEE MDL/Marketing
www.ieee.org/go/freemonth

Page 157
IEEE ICASSP 2017
www.ieee-icassp2017.org

CVR 4
Mathworks
www.mathworks.com/wireless
+1 508 647 7040

James A. Vick
Sr. Director, Advertising
Phone: +1 212 419 7767;
Fax: +1 212 419 7589
jv.ieeemedia@ieee.org

Marion Delaney
Advertising Sales Director
Phone: +1 415 863 4717;
Fax: +1 415 863 4717
md.ieeemedia@ieee.org

Mark David
Sr. Manager Advertising & Business Development
Phone: +1 732 465 6473
Fax: +1 732 981 1855
m.david@ieee.org

Mindy Belfer
Advertising Sales Coordinator
Phone: +1 732 562 3937
Fax: +1 732 981 1855
m.belfer@ieee.org

PRODUCT ADVERTISING

MIDATLANTIC
Lisa Rinaldo
Phone: +1 732 772 0160;
Fax: +1 732 772 0164
lr.ieeemedia@ieee.org
NY, NJ, PA, DE, MD, DC, KY, WV

NEW ENGLAND/ SOUTH CENTRAL/ EASTERN CANADA
Jody Estabrook
Phone: +1 774 283 4528;
Fax: +1 774 283 4527
je.ieeemedia@ieee.org
ME, VT, NH, MA, RI, CT, AR, LA, OK, TX
Canada: Quebec, Nova Scotia, Newfoundland, Prince Edward Island, New Brunswick

SOUTHEAST
Cathy Flynn
Phone: +1 770 645 2944;
Fax: +1 770 993 4423
cf.ieeemedia@ieee.org
VA, NC, SC, GA, FL, AL, MS, TN

MIDWEST/ CENTRAL CANADA
Dave Jones
Phone: +1 708 442 5633;
Fax: +1 708 442 7620
dj.ieeemedia@ieee.org
IL, IA, KS, MN, MO, NE, ND, SD, WI, OH
Canada: Manitoba, Saskatchewan, Alberta

MIDWEST/ ONTARIO, CANADA
Will Hamilton
Phone: +1 269 381 2156;
Fax: +1 269 381 2556
wh.ieeemedia@ieee.org
IN, MI. Canada: Ontario

WEST COAST/ MOUNTAIN STATES/ WESTERN CANADA
Marshall Rubin
Phone: +1 818 888 2407;
Fax: +1 818 888 4907
mr.ieeemedia@ieee.org
AZ, CO, HI, NM, NV, UT, AK, ID, MT, WY, OR, WA, CA. Canada: British Columbia

EUROPE/AFRICA/ MIDDLE EAST ASIA/FAR EAST/ PACIFIC RIM
Michael O'Kane
Phone: +44 1875 825 700;
Fax: +44 1875 825 701
mo.ieeemedia@ieee.org
Europe, Africa, Middle East
Asia, Far East, Pacific Rim, Australia, New Zealand

RECRUITMENT ADVERTISING

MIDATLANTIC
Lisa Rinaldo
Phone: +1 732 772 0160;
Fax: +1 732 772 0164
lr.ieeemedia@ieee.org
NY, NJ, CT, PA, DE, MD, DC, KY, WV

NEW ENGLAND/ EASTERN CANADA
Liza Reich
Phone: +1 212 419 7578;
Fax: +1 212 419 7589
e.reich@ieee.org
ME, VT, NH, MA, RI.
Canada: Quebec, Nova Scotia, Prince Edward Island, Newfoundland, New Brunswick

SOUTHEAST
Cathy Flynn
Phone: +1 770 645 2944;
Fax: +1 770 993 4423
cf.ieeemedia@ieee.org
VA, NC, SC, GA, FL, AL, MS, TN

MIDWEST/ SOUTH CENTRAL/ CENTRAL CANADA
Darcy Giovingo
Phone: +224 616 3034;
Fax: +1 847 729 4269
dg.ieeemedia@ieee.org;
AR, IL, IN, IA, KS, LA, MI, MN, MO, NE, ND, SD, OH, OK, TX, WI. Canada: Ontario, Manitoba, Saskatchewan, Alberta

WEST COAST/ SOUTHWEST/ MOUNTAIN STATES/ASIA
Tim Matteson
Phone: +1 310 836 4064;
Fax: +1 310 836 4067
tm.ieeemedia@ieee.org
AZ, CO, HI, NV, NM, UT, CA, AK, ID, MT, WY, OR, WA. Canada: British Columbia

EUROPE/AFRICA/ MIDDLE EAST
Michael O'Kane
Phone: +44 1875 825 700;
Fax: +44 1875 825 701
mo.ieeemedia@ieee.org
Europe, Africa, Middle East

SP

DATES AHEAD

Please send calendar submissions to:
Jessica Barragué
E-mail: j.barrague@ieee.org

2016

SEPTEMBER

IEEE International Workshop on Machine Learning for Signal Processing (MLSP)
13–16 September, Salerno, Italy.

Fourth International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa)
19–22 September, Aachen, Germany.

IEEE International 18th International Workshop on Multimedia Signal Processing (MMSP)
20–23 September, Montréal, Québec, Canada.

Sensor Signal Processing for Defence (SSPD)
22–23 September, Edinburgh, Great Britain.
URL: http://www.sspd.eng.ed.ac.uk

IEEE International Conference on Image Processing (ICIP)
25–28 September, Phoenix, Arizona, USA.
General Chair: Lina Karam
URL: http://www.ieeeicip2016.org

NOVEMBER

50th Annual Asilomar Conference on Signals, Systems, and Computers (ASILOMAR)
6–9 November, Pacific Grove, California, USA.
General Chair: Philip Schniter
URL: http://www.asilomarsscconf.org/

Digital Object Identifier 10.1109/MSP.2016.2587164
Date of publication: 2 September 2016

ICIP 2016 will be held in Phoenix, Arizona, United States, 25–28 September 2016, at the Phoenix Convention Center.

©ISTOCKPHOTO.COM/TONDA

DECEMBER

Picture Coding Symposium (PCS)
4–7 December, Nuremberg, Germany.
General Chair: André Kaup
URL: http://www.pcs2016.com/

Eighth IEEE International Workshop on Information Forensics and Security (WIFS)
5–7 December, Abu Dhabi, UAE.
General Chairs: Ernesto Damiani and Nasir Memon
URL: http://wifs2016.mdabaie.com/

IEEE Global Conference on Signal and Information Processing (GlobalSIP)
7–9 December, Greater Washington, D.C., USA.
General Chairs: Zhi Tian and Brian Sadler
URL: http://2016.ieeeglobalsip.org

IEEE World Forum on Internet of Things (WF-IoT)
12–14 December, Reston, Virginia, USA.
General Chairs: Geoff Mulligan and Latif Ladid
URL: http://wfiot2016.ieee-wf-iot.org/

IEEE Spoken Language Technology Workshop (SLT)
13–16 December, San Juan, Puerto Rico.

Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)
13–16 December, Jeju, South Korea.

2017

MARCH

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
5–9 March, New Orleans, Louisiana, USA.
General Chair: Magdy Bayoumi
http://www.ieee-icassp2017.org/

APRIL

IEEE International Symposium on Biomedical Imaging (ISBI)
18–21 April, Melbourne, Australia.
URL: http://biomedicalimaging.org/2017/

SEPTEMBER

IEEE International Conference on Image Processing (ICIP)
17–20 September, Beijing, China.
General Chairs: Xinggang Lin, Anthony Vetro, and Min Wu
URL: http://2017.ieeeicip.org/

DECEMBER

17th IEEE International Workshop on Computational Advances in Multisensor Adaptive Processing (CAMSAP)
10–13 December, Curacao, Dutch Antilles.
General Chairs: André L.F. de Almeida and Martin Haardt
URL: http://www.cs.huji.ac.il/conferences/CAMSAP17/

SP

SIGNAL PROCESSING

The Science Behind Our Digital Life

Cast Your Vote for the Future of the IEEE and the Signal Processing Society

Voting is now open until **3 October 2016** on important issues and board elections. Take a moment to become an informed voter, and vote to have your voice heard.

IEEE CONSTITUTIONAL AMENDMENT: Should the IEEE remain a "Bottom-Up" volunteer association, or change to a "Top-Down" management? Learn about this amendment from the President's Message (page 4 of this issue).

SPS BOARD OF GOVERNORS ELECTION: Meet the candidates of member-at-large and regional-director-at-large of the Signal Processing Society's governing body (page 6 of this issue). Elected board members represent your voices in important decision making of the Society.

signalprocessingsociety.org

MATLAB SPEAKS
WIRELESS DESIGN

You can simulate, prototype, and verify wireless systems right in MATLAB. Learn how today's MATLAB supports RF, LTE, WLAN and 5G development and SDR hardware.

mathworks.com/wireless

©2016 The MathWorks, Inc

MathWorks®

IEEE SIGNAL PROCESSING SOCIETY

Content**Gazette**

SEPTEMBER 2016

ISSN 2167-5023

IEEE Signal Processing Society

◆IEEE

IEEE International Symposium on Biomedical Imaging

April 18-21, 2017, Melbourne, Australia

CALL FOR PAPERS

Connect with us:

The IEEE International Symposium on Biomedical Imaging (ISBI) is a scientific conference dedicated to mathematical, algorithmic, and computational aspects of biomedical imaging, across all scales of observation. It fosters knowledge transfer among different imaging communities and contributes to an integrative approach to biomedical imaging.

ISBI is a joint initiative from the IEEE Signal Processing Society (SPS) and the IEEE Engineering in Medicine and Biology Society (EMBS). The 2017 meeting will include tutorials, and a scientific program composed of plenary talks, invited special sessions, challenges, as well as oral and poster presentations of peer-reviewed papers.

High-quality papers are requested containing original contributions to the topics of interest including image formation and reconstruction, computational and statistical image processing and analysis, dynamic imaging, visualization, image quality assessment, and physical, biological, and statistical modeling. Accepted 4-page regular papers will be published in the symposium proceedings published by IEEE and included in IEEE Xplore.

To encourage attendance by a broader audience of imaging scientists and offer additional presentation opportunities, ISBI 2017 will continue to propose a second track featuring posters selected from 1-page abstract submissions without subsequent archival publication.

Venue:

ISBI 2017 will be held at the Melbourne Convention and Exhibition Centre, a world class venue in the heart of the vibrant city of Melbourne, Australia within walking distance of hotels and dining options. The conference starts on the Tuesday immediately after the Easter holiday period to provide attendees with the opportunity to vacation before the conference enjoying the many attractions of Australia.

Important Dates:

4-page paper submission
August 1st - October 24th, 2016

Author Notification for 4-page papers
December 19th, 2016

Final version of 4-page papers & registration
January 9th, 2017

Conference Chairs

Gary Egan
Monash University, Australia

Olivier Salvado
CSIRO, Brisbane, Australia

Program Chairs

Simon Warfield
Harvard Medical School, Boston, USA

Arrate Munoz-Barrutia
Universidad Carlos III de Madrid, Spain

Organizing Committee

Roger Ordidge, Parvin Mousavi, Andrew Zalesky, Fabrice Meriaudeau, Tammy Riklin Raviv, Stephen Aylward, Bram Van Ginneken, Adrienne Mendrik, Wiro Niessen, Stuart Crozier, Steve Lee, Mehrnaz Shoushtarian, Eduardo Romero, Amir Armini

Contact

Janice Sandler *j.sandler@ieee.org*
http://biomedicalimaging.org/2017

https://www.facebook.com/ISBI-2017-1586861521610962/
https://twitter.com/IEEEISBI2017

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

REGULAR PAPERS

◆IEEE

CORRECTION

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed

MEDLINE
U.S. National Library of Medicine

new

Now with no page charges!

IEEE TRANSACTIONS ON

SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

Now accepting paper submissions

The new *IEEE Transactions on Signal and Information Processing over Networks* publishes high-quality papers that extend the classical notions of processing of signals defined over vector spaces (e.g. time and space) to processing of signals and information (data) defined over networks, potentially dynamically varying. In signal processing over networks, the topology of the network may define structural relationships in the data, or may constrain processing of the data. Topics of interest include, but are not limited to the following:

Adaptation, Detection, Estimation, and Learning
- Distributed detection and estimation
- Distributed adaptation over networks
- Distributed learning over networks
- Distributed target tracking
- Bayesian learning; Bayesian signal processing
- Sequential learning over networks
- Decision making over networks
- Distributed dictionary learning
- Distributed game theoretic strategies
- Distributed information processing
- Graphical and kernel methods
- Consensus over network systems
- Optimization over network systems

Communications, Networking, and Sensing
- Distributed monitoring and sensing
- Signal processing for distributed communications and networking
- Signal processing for cooperative networking
- Signal processing for network security
- Optimal network signal processing and resource allocation

Modeling and Analysis
- Performance and bounds of methods
- Robustness and vulnerability
- Network modeling and identification

Modeling and Analysis (cont.)
- Simulations of networked information processing systems
- Social learning
- Bio-inspired network signal processing
- Epidemics and diffusion in populations

Imaging and Media Applications
- Image and video processing over networks
- Media cloud computing and communication
- Multimedia streaming and transport
- Social media computing and networking
- Signal processing for cyber-physical systems
- Wireless/mobile multimedia

Data Analysis
- Processing, analysis, and visualization of big data
- Signal and information processing for crowd computing
- Signal and information processing for the Internet of Things
- Emergence of behavior

Emerging topics and applications
- Emerging topics
- Applications in life sciences, ecology, energy, social networks, economic networks, finance, social sciences, smart grids, wireless health, robotics, transportation, and other areas of science and engineering

Editor-in-Chief: Petar M. Djurić, Stony Brook University (USA)

To submit a paper, go to: https://mc.manuscriptcentral.com/tsipn-ieee

IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society ®
http://www.signalprocessingsociety.org/

IEEE COMPUTER SOCIETY
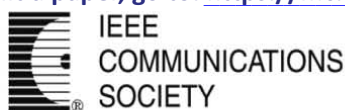
Call for Papers
IEEE Signal Processing Society
IEEE Transactions on Signal and Information Processing over Networks

Special Issue on Distributed Information Processing in Social Networks

Over the past few decades, online social networks such as *Facebook* and *Twitter* have significantly changed the way people communicate and share information with each other. The opinion and behavior of each individual are heavily influenced through interacting with others. These local interactions lead to many interesting collective phenomena such as herding, consensus, and rumor spreading. At the same time, there is always the danger of mob mentality of following crowds, celebrities, or gurus who might provide misleading or even malicious information. Many efforts have been devoted to investigating the collective behavior in the context of various network topologies and the robustness of social networks in the presence of malicious threats. On the other hand, activities in social networks (clicks, searches, transactions, posts, and tweets) generate a massive amount of decentralized data, which is not only big in size but also complex in terms of its structure. Processing these data requires significant advances in accurate mathematical modeling and computationally efficient algorithm design.

Many modern technological systems such as wireless sensor and robot networks are virtually the same as social networks in the sense that the nodes in both networks carry disparate information and communicate with constraints. Thus, investigating social networks will bring insightful principles on the system and algorithmic designs of many engineering networks. An example of such is the implementation of consensus algorithms for coordination and control in robot networks. Additionally, more and more research projects nowadays are data-driven. Social networks are natural sources of massive and diverse big data, which present unique opportunities and challenges to further develop theoretical data processing toolsets and investigate novel applications. This special issue aims to focus on addressing distributed information (signal, data, etc.) processing problems in social networks and also invites submissions from all other related disciplines to present comprehensive and diverse perspectives.

Topics of interest include, but are not limited to:

- Dynamic social networks: time varying network topology, edge weights, etc.
- Social learning, distributed decision-making, estimation, and filtering
- Consensus and coordination in multi-agent networks
- Modeling and inference for information diffusion and rumor spreading
- Multi-layered social networks where social interactions take place at different scales or modalities
- Resource allocation, optimization, and control in multi-agent networks
- Modeling and strategic considerations for malicious behavior in networks
- Social media computing and networking
- Data mining, machine learning, and statistical inference frameworks and algorithms for handling big data from social networks
- Data-driven applications: attribution models for marketing and advertising, trend prediction, recommendation systems, crowdsourcing, etc.
- Other topics associated with social networks: graphical modeling, trust, privacy, engineering applications, etc.

Important Dates:

- Manuscript submission due: September 15, 2016
- First review completed: November 1, 2016
- Revised manuscript due: December 15, 2016
- Second review completed: February 1, 2017
- Final manuscript due: March 15, 2017
- Publication: June 1, 2017

Guest Editors:

- Zhenliang Zhang, Qualcomm Corporate R&D (zhenlian@qti.qualcomm.com)
- Wee Peng Tay, Nanyang Technological University (wptay@ntu.edu.sg)
- Moez Draief, Imperial College London (m.draief@imperial.ac.uk)
- Xiaodong Wang, Columbia University (xw2008@columbia.edu)
- Edwin K. P. Chong, Colorado State University (edwin.chong@colostate.edu)
- Alfred O. Hero III, University of Michigan (hero@eecs.umich.edu)

IEEE/ACM TRANSACTIONS ON

AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

REGULAR PAPERS

◆IEEE

Call for Papers
IEEE Signal Processing Society
IEEE Transactions on Signal and Information Processing over Networks

Special Issue on Distributed Signal Processing for Security and Privacy in Networked Cyber-Physical Systems

GUEST EDITORS:
Arash Mohammadi, Concordia University, Montreal, Canada, arashmoh@encs.concordia.ca
Peng Cheng, Zhejiang University, Hangzhou, China, pcheng@iipc.zju.edu.cn
Vincenzo Piuri, Universita degli Studi di Milano, Milan, Italy, vincenzo.piuri@unimi.it
Konstantinos N. Plataniotis, University of Toronto, Toronto, Canada, kostas@ece.utoronto.ca
Patrizio Campisi, Universitá degli Studi Roma Tre, Italy, patrizio.campisi@uniroma3.it

SCOPE

The focus of this special issue is on distributed information acquisition, estimation, and adaptive learning for security and privacy in the context of networked cyber-physical systems (CPSs) which are engineering systems with integrated computational and communication capabilities that interact with humans through cyber space. The CPSs have recently emerged in several practical applications of engineering importance including aerospace, industrial/manufacturing process control, multimedia networks, transportation systems, power grids, and medical systems. The CPSs typically consist of both wireless and wired sensor/agent networks with different capacity/reliability levels where the emphasis is on real-time operations, and performing distributed, secure, and optimal sensing/processing is the key concern. To satisfy these requirements of the CPSs, it is of paramount importance to design innovative "Signal Processing" tools to provide unprecedented performance and resource utilization efficiency.

A significant challenge for implementation of signal processing solutions in CPSs is the difficulty of acquiring data from geo- graphically distributed observation nodes and storing/processing the aggregated data at the fusion center (FC). As such, there has been a recent surge of interest in development of distributed and collaborative signal processing technologies where adaptation, estimation, and/or control are performed locally and communication is limited to local neighborhoods. Distributed signal processing over networked CPSs, however, raise significant privacy and security concerns as local observations are being shared by neighboring nodes in a collaborative and iterative fashion. On one hand, applications of CPSs are severely safety critical where potential cyber and physical attacks by adversaries on signal processing modules could lead to a variety of severe consequences including customer information leakage, destruction of infrastructures, and endangering human lives. On the other hand, the need for cooperation be- tween neighboring nodes makes it imperative to prevent the disclosure of sensitive local information during distributed information fusion step. At the same time, efficient usage of available resources (communication, computation, bandwidth, and energy) is a pre- requisite for productive operation of the CPSs. To accommodate these critical aspects of CPSs, it is of great practical importance and theoretical significance to develop advanced "Secure and Privacy Preserving Distributed Signal Processing" solutions.

The spirit and wide scope of distributed signal processing in revolutionized CPSs calls for novel and innovative techniques beyond conventional approaches to provide precise guarantees on security and privacy of CPSs. The objective of this special issue is to further advance recent developments of distributed signal processing to practical aspects of CPSs for real-time processing and monitoring of the underlying system in a secure and privacy preserving manner while avoiding degradation of the processing performance and preserving the valuable resources. To provide a systematic base for future advancements of CPSs, this special issue aims to provide a research venue to investigate distributed signal processing techniques with adaptation, cooperation, and learning capabilities which are secure against cyber-attacks and protected against privacy leaks. The emphasis of this special issue is on distributed/network aspects of security and privacy in CPSs. Papers with primary emphasis on forensics and security will be redirected to IEEE Transactions on Information Forensics and Security (TIFS). Topics of interest include, but are not limited to:

- Security and Privacy of distributed signal processing in networked CPSs.
- Distributed and secure detection, estimation, and information fusion.
- Security and privacy of consensus and diffusive strategies in networked systems.
- Secure and privacy preserving distributed adaptation and learning.
- Security and privacy of distributed sensor resource management in networked systems.
- Distributed event-based estimation/control in networked CPSs.
- Detection and identification of potential attacks on distributed signal processing mechanisms.
- Application domains including but not limited to, smart grids, camera networks, multimedia network, and vehicular networks.

SUBMISSION GUIDELINES

Authors are invited to submit original research contributions by following the detailed instructions given in the "Information for Authors" at http://www.signalprocessingsociety.org/publications/periodicals/tsipn/. Manuscripts should be submitted via S c h o l a r O n e (Manuscript Central) at http://mc.manuscriptcentral.com/tsipn-ieee . Questions about the special issue should be directed to the Guest Editors.

IMPORTANT DATES

Paper submission deadline:	December 15, 2016	Final notification:	September 1, 2017
Notification of the first review:	March 1, 2017	Final manuscript due:	October 15, 2017
Revised paper submission:	April 15, 2017	Publication:	Advance posting in IEEE explore as
Notification of the re-review:	June 15, 2017		soon as authors approve galley proofs
Minor revision deadline:	August 1, 2017	Expected inclusion in an issue:	March 2018

IEEE TRANSACTIONS ON

IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society ®

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed

MEDLINE
U.S. National Library of Medicine

◆IEEE

COMMENTS AND CORRECTIONS

NEW!

IEEE TRANSACTIONS ON
COMPUTATIONAL IMAGING

30% FASTER TURN AROUND THAN TIP!

Editor-in-Chief

W. Clem Karl
Boston University

Technical Committee

Charles Bouman
Eric Miller
Peter Corcoran
Jong Chul Ye
Dave Brady
William Freeman

The IEEE Transactions on Computational Imaging publishes research results where computation plays an integral role in the image formation process. All areas of computational imaging are appropriate, ranging from the principles and theory of computational imaging, to modeling paradigms for computational imaging, to image formation methods, to the latest innovative computational imaging system designs. Topics of interest include, but are not limited to the following:

Computational Imaging Methods and Models

- Coded image sensing
- Compressed sensing
- Sparse and low-rank models
- Learning-based models, dictionary methods
- Graphical image models
- Perceptual models

Computational Image Formation

- Sparsity-based reconstruction
- Statistically-based inversion methods
- Multi-image and sensor fusion
- Optimization-based methods; proximal iterative methods, ADMM

Computational Photography

- Non-classical image capture
- Generalized illumination
- Time-of-flight imaging
- High dynamic range imaging
- Plenoptic imaging

Computational Consumer Imaging

- Mobile imaging, cell phone imaging
- Camera-array systems
- Depth cameras, multi-focus imaging
- Pervasive imaging, camera networks

Computational Acoustic Imaging

- Multi-static ultrasound imaging
- Photo-acoustic imaging
- Acoustic tomography

Computational Microscopy

- Holographic microscopy
- Quantitative phase imaging
- Multi-illumination microscopy
- Lensless microscopy
- Light field microscopy

Imaging Hardware and Software

- Embedded computing systems
- Big data computational imaging
- Integrated hardware/digital design

Tomographic Imaging

- X-ray CT
- PET
- SPECT

Magnetic Resonance Imaging

- Diffusion tensor imaging
- Fast acquisition

Radar Imaging

- Synthetic aperture imaging
- Inverse synthetic aperture imaging

Geophysical Imaging

- Multi-spectral imaging
- Ground penetrating radar
- Seismic tomography

Multi-spectral Imaging

- Multi-spectral imaging
- Hyper-spectral imaging
- Spectroscopic imaging

For more information on the **IEEE Transactions on Computational Imaging** see
http://www.signalprocessingsociety.org/publications/periodicals/tci/

NOW WITH NO PAGE CHARGES!

IEEE Signal Processing Society ®   EMB   GRSS   IEEE

General Chairs

André L. F. de Almeida
andre@gtel.ufc.br
Federal University of Ceará,
Brazil

Martin Haardt
martin.haardt@tu-ilmenau.de
Ilmenau University of Technology,
Germany

Technical Program Chairs

Xiaoli Ma
xiaoli@gatech.edu
Georgia Institute of Technology,
USA

Shahram ShahbazPanahi
shahram.shahbazpanahi@uoit.ca
University of Ontario Institute of
Technology,
Canada

Finance Chair

Cihan Tepedelenlioglu
cihan@asu.edu
Arizona State University,
USA

Publicity and Publications Chair

Ami Wiesel
ami.wiesel@huji.ac.il
Hebrew University of Jerusalem,
Israel

Local Arrangements Chair

Geert Leus
g.j.t.leus@tudelft.nl
Delft University of Technology,
The Netherlands

Robert W. Heath Jr.
rheath@utexas.edu
University of Texas at Austin,
USA

Join Twitter Conversation
#CAMSAP2017

◆IEEE

CAMSAP 2017

Call for Papers

The Seventh IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing

Curaçao, Dutch Antilles
December 10-13, 2017
http://www.cs.huji.ac.il/conferences/CAMSAP17

Following the success of the first six editions of the IEEE workshop on Computational Advances in Multi-Sensor Adaptive Processing, we are pleased to announce the seventh workshop in this series. IEEE CAMSAP 2017 will be held in Curaçao, Dutch Antilles, and will feature a number of plenary talks from the world's leading researchers in the area, special focus sessions, and contributed papers. All papers will undergo peer review in order to provide feedback to the authors and ensure a high-quality program.

Topics and applications of interest for the workshop include, but are not limited to, the following.

TOPICS OF INTEREST
- Array processing, waveform diversity, space-time adaptive processing
- Convex optimization and relaxation
- Computational linear & multi-linear algebra
- Computer-intensive methods in signal processing (bootstrap, MCMC, EM, particle filtering, etc.)
- Signal and information processing over networks
- Sparse signal processing

APPLICATIONS
- Big data
- Biomedical signal processing
- Communication systems
- Computational imaging
- Radar
- Sensor networks
- Smart grids
- Sonar

Submission of Papers: Prospective authors are invited to submit original full-length papers, with up to four pages for technical content including figures and references, using the formatting guidelines on the website for reviewing purposes. All accepted papers must be presented at the workshop to appear in the proceedings. Best student paper awards, selected by a CAMSAP committee, will also be presented at the workshop.

Special Session Proposals: In addition to contributed sessions, the workshop will also have a number of special sessions. Prospective organizers of special sessions are invited to submit a proposal form, available on the workshop website, by e-mail to the Special Sessions Chair.

IMPORTANT DEADLINES

Submission of proposals for special sessions	**March, 2017**
Notification of special session acceptance	**March 15, 2015**
Submission of papers	**July, 2017**
Notification of paper acceptance	**September, 2017**
Final paper submission	

IEEE
Signal Processing Society

IEEE TRANSACTIONS ON
INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

REGULAR PAPERS

ANNOUNCEMENTS

IEEE TRANSACTIONS ON
MULTIMEDIA

IEEE JOURNAL OF
SELECTED TOPICS IN SIGNAL PROCESSING

IEEE Signal Processing Society ®

www.ieee.org/sp/index.html

ISSUE ON PERSON-CENTERED SIGNAL PROCESSING FOR ASSISTIVE, REHABILITATIVE, AND WEARABLE HEALTH TECHNOLOGIES

IEEE

IEEE Journal on Selected Topics in Signal Processing

IEEE Transactions on Signal and Information Processing over Networks

Special Issues on Graph Signal Processing

Numerous applications rely on the processing of high-dimensional data that resides on irregular or otherwise unordered structures which are naturally modeled as networks (such as social, economic, energy, transportation, telecommunication, sensor, and neural, to name a few). The need for new tools to process such data has led to the emergence of the field of graph signal processing, which merges algebraic and spectral graph theoretic concepts with computational harmonic analysis to process signals on structures such as graphs. This important new paradigm in signal processing research, coupled with its numerous applications in very different domains, has fueled the rapid development of an inter-disciplinary research community that has been working on theoretical aspects of graph signal processing and applications to diverse problems such as big data analysis, coding and compression of 3D point clouds, biological data processing, and brain network analysis.

The purpose of these special issues is to gather the latest advances in graph signal processing and disseminate new ideas and experiences in this emerging field to a broad audience. We encourage the submission of papers with new results, methods or applications in graph signal processing. In particular, the topics of interest include (but are not limited to):

- Sampling and recovery of graph signals
- Graph filter and filter bank design
- Uncertainty principles and other fundamental limits
- Graph signal transforms
- Graph topology inference
- Prediction and learning in graphs
- Statistical graph signal processing
- Non-linear graph signal processing
- Applications to visual information processing
- Applications to neuroscience and other medical fields
- Applications to economics and social networks
- Applications to various infrastructure networks

Submission procedure:

Prospective authors should follow the instructions given on the IEEE JSTSP webpages and submit their manuscript with the web submission system at https://mc.manuscriptcentral.com/jstsp-ieee. The decisions on whether the accepted papers will be published in IEEE JSTSP or IEEE TSIPN will depend on the respective themes of the papers and will be made by the Guest Editors.

Schedule (all deadlines are firm)

Manuscript due:	Nov 1, 2016
First Review Completed:	Jan 1, 2017
Revised manuscript due:	Mar 1, 2017
Second Review Completed:	May 1, 2017
Final manuscript due:	June 1, 2017
Publication date:	September 2017

Guest Editors:

Pier-Luigi Dragotti, Imperial College, London (p.dragotti@imperial.ac.uk)
Pascal Frossard, EPFL, Lausanne (pascal.frossard@epfl.ch)
Antonio Ortega, USC, Los Angeles (ortega@sipi.usc.edu)
Michael Rabbat, McGill University, Montreal (michael.rabbat@mcgill.ca)
Alejandro Ribeiro, UPenn, Philadelphia (aribeiro@seas.upenn.edu)

CALL FOR PAPERS
IEEE Journal of Selected Topics in Signal Processing

Special Issue on Signal Processing and Machine Learning for Education and Human Learning at Scale

Aims and Scope

The surge in popularity of Massive Open Online Courses (MOOCs) and other online and blended learning platforms has demonstrated the potential of the Internet for scaling education. While advances in technology have enabled content delivery to massive numbers of students, these platforms remain limited in their ability to provide an effective learning experience for each individual.

Recent advances in machine learning and signal processing offer promising avenues to move beyond this "one size fits all" educational approach. The key is that today's learning technology platforms can capture big data about learners as they proceed through courses. Examples of learning data include performance on homeworks and exams, click actions made while watching lecture videos or interacting with simulations, the social learning networks formed among the students, and the content posted on discussion forums. Going even further, prototype platforms are being built that use cameras and other sensors to continuously monitor students' affect and engagement. The large volumes of empirical learning data being collected present novel opportunities to study the process of student learning, to design systems that improve learning at scale by closing the learning feedback loop.

This special issue of IEEE J-STSP will showcase the research from the signal processing community that is providing leadership in advancing effective learning at scale. Particularly of interest to this special issue will be novel methods for defining and extracting signals of a student's behavior and performance from big learning data and using these measures in the design of intelligent algorithms and systems.

Topics of interest in the special issue include (but are not limited to):

- Processing and Representing Learning Behavioral Data
- Generative/Low Dimensional Modeling of Student Learning
- Learner Knowledge Tracing and Performance Prediction
- Social Learning Networks
- Algorithms for Identifying Learner Collaborations
- Automating Course/Content Individualization, Automatic Grading Methods, Automatic Feedback Generation
- Learning Analytics with Actionable Intelligence for Instructors
- Algorithms for More Effective Peer Grading Allocation
- Relationships between Learning Behavior, Performance, and Content
- Machine Vision Algorithms for Processing Student Biometric Data
- Trials for Demonstrating Efficacy for Learners and/or Instructors

Important Dates:

Manuscript submission due: October 1, 2016
First review completed: December 15, 2016
Revised manuscript due: February 1, 2017
Second review completed: May 1, 2017
Final manuscript due: June 15, 2017
Publication: August 2017

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted using the Manuscript Central system at http://mc.manuscriptcentral.com/jstsp-ieee.

Guest Editors:

Mihaela van der Schaar (Lead Guest Editor), University of California Los Angeles, USA, mihaela@ee.ucla.edu
Mung Chiang, Princeton University, USA, chiangm@princeton.edu
Richard Baraniuk, Rice University, USA, richb@rice.edu
Jonathan Chung-Kuan Huang, Google, USA, jonathanhuang@google.com
Shengdong Zhao, National University of Singapore, Singapore, zhaosd@comp.nus.edu.sg

IEEE

SIGNAL PROCESSING LETTERS

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

www.ieee.org/sp/index.html

IEEE TRANSACTIONS ON
COMPUTATIONAL IMAGING

A PUBLICATION OF
IEEE SIGNAL PROCESSING SOCIETY
IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY
IEEE CONSUMER ELECTRONICS SOCIETY

TECHNICALLY CO-SPONSORED BY
IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY

PAPERS

IEEE TRANSACTIONS ON
SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE COMMUNICATIONS SOCIETY
THE IEEE COMPUTER SOCIETY

Modeling and Analysis

General Chairs
Jörn Ostermann, LUH, Germany
Kenneth K.M. Lam, PolyU, Hong Kong
Technical Program Chairs
Thomas Sikora (Coordinator), TU Berlin, Germany
Qian Zhang, HKUST, Hong Kong
Prabha Balakrishnan, U. Texas at Dallas, USA
Winston Hsu, NTU, Taiwan
Yonggang Wen, NTU, Singapore
Ce Zhu, UESTC, China
Finance Chairs
Daniel P.K. Lun, PolyU, Hong Kong
Chris Y.H. Chan, PolyU, Hong Kong
Workshop Chairs
Béatrice Pesquet-Popescu, ParisTech. France
Chong-Wah Ngo, CityU, Hong Kong
Special Sessions Chair
Gene Cheung, National Institute of Informatics, Japan
Zhu Li, University of Missouri – Kansas City, USA
Panel Session Chair
Lap Pui Chau, NTU, Singapore
Tutorial Chair
Atanas Gotchev, TUT, Finland
Grand Challenge Chair
Axel Becker-Lakus, Five Open Books LLC, USA
Student Program Chair
Hong Zhao, Tsinghua University, China
Award Chair
Moncef Gabbouj, Tampere University of Technology, Finland
Sponsorship Chair
Xian-Sheng Hua, Alibaba, China
Registration Chairs
Tan Lee, CUHK, Hong Kong
Publications Chairs
Yui Lam Chan, PolyU, Hong Kong
Publicity Chair
Thomas Fang Zheng, Tsinghua University, China
Ling Guan, Ryerson University, Canada
Web Chair
Thorsten Laude, LUH, Germany
Local Arrangement Chairs
Frank Leung, PolyU, Hong Kong
Secretaries
Bonnie Law, PolyU, Hong Kong
Jiantao Zhou, U. Macau, Macau
Europe Liaison
N.N.
North America Liaison
N.N.
South America Liaison
N.N.
Australia/Liaison
N.N.
Asia Liaison
Xiaopeng Fan, HIT, China
Mark Liao, Academia Sinica, Taiwan
Advisor
Wen Gao, Peking University, China
Wan Chi Siu, PolyU, Hong Kong

IEEE International Conference on Multimedia and Expo (ICME) 2017

CALL FOR PAPERS

IEEE International Conference on Multimedia and Expo (ICME) 2017
July 10 – 14, 2017 ● Hong Kong

IEEE International Conference on Multimedia & Expo (ICME) has been the flagship multimedia conference sponsored by four IEEE societies since 2000. It serves as a forum to promote the exchange of the latest advances in multimedia technologies, systems, and applications from both the research and development perspectives of the circuits and systems, communications, computer, and signal processing communities. An Exposition of multimedia products, animations and industries will be held in conjunction with the conference. The theme of the 2017 conference will be "The New Media Experience", enabling next generation 3D/AR/VR experiences and applications, based on which various sessions and events will be organized.

Authors are invited to submit a full paper (two-column format, 6 pages maximum) according to the guidelines available on the conference website at http://www.icme2017.org/. Reviewing will be double blind. Only electronic submissions will be accepted. Topics of interest include, but are not limited to:

- Speech, audio, image, video, text and new sensor signal processing
- Signal processing for media integration
- Interactive 3D media and immersive VR/AR environments
- Multi-modal multimedia computing systems and human-machine interaction
- Multimedia Cloud and Big Data
- Multimedia IoT
- Mobile Multimedia Systems and Applications
- Multimedia compression
- Multimedia security, privacy, and content protection
- Multimedia databases and digital libraries
- Multimedia applications and services
- Media content analysis, analytics, and search
- Hardware and software for multimedia systems
- Multimedia standards and related issues
- Multimedia quality assessment

ICME 2017 will showcase high quality oral and poster presentations and demonstration sessions. Best paper awards (one diamond and five platinum awards) will be selected and recognized in the conference. Several awards sponsored by industry and institutions will be given out. Accepted papers have to be registered and presented; otherwise they will not be included in the IEEE Xplore Library.

ICME2017 will feature IEEE societies sponsored Workshops. We welcome researchers, developers and practitioners to organize Workshops on various new emerging topics. Industrial exhibitions are held in conjunction with the main conference. Proposals for Tutorials, Special Sessions, and Demos are also invited. Please visit the ICME 2017 website for submission details.

Regular Paper Abstract Submission:	**November 25, 2016**
Regular Paper Submission:	**December 2, 2016**
Notification of Regular Paper Acceptance:	**February 24, 2017**
Special Session/Workshop Proposals Due:	**November 1, 2016**
Notification of Special Session/Workshop Proposal Acceptance:	**November 7, 2016**
Tutorial Proposals Due:	**January 23, 2017**
Workshop Paper Submission:	**February 27, 2017**
Notification of Workshop Paper Acceptance:	**March 27, 2017**
Camera-Ready Paper Due:	**April 10, 2017**

www.icme2017.org

Contact Email:
icme@tnt.uni-hannover.de

SPAWC 2017

The 18th IEEE International Workshop on Signal Processing Advances in Wireless Communications, July 3 – 6, Sapporo, Japan

SPAWC 2017 will be held at Hokkaido University in Sapporo, Japan on July 3-6, 2017. The workshop is devoted to advances in signal processing for wireless communications, networking, and information theory. The technical program features plenary talks as well as invited and contributed papers presented in poster format.

Call for Papers

Prospective authors are invited to submit papers in the following areas:
• Smart antennas, MIMO systems, massive MIMO, and space-time processing
• Single-carrier, multi-carrier, and multi-rate systems
• Multiple-access and broadcast channels, multi-user receivers
• Signal processing for ad-hoc, multi-hop, and sensor networks
• Cooperative communication, coordinated multipoint transmission and reception
• Distributed resource allocation and scheduling
• Convex and non-convex optimization; Game theory for communications
• Interference management, dynamic spectrum management
• Heterogeneous networks, small cells
• Millimeter wave, 60GHz communications
• Full duplex systems
• Physical layer security
• Feedback in wireless networks
• Cognitive radio and networks
• Cooperative sensing, compressed sensing, sparse signal processing
• Machine-to-machine, device-to-device communications
• Ultra-wideband radio, localization, RFID
• Modeling, estimation and equalization of wireless channels
• Acquisition, synchronization, and tracking
• Cross-layer issues, joint source-channel coding, delay-limited communication
• Signal processing for optical, satellite, and underwater communications
• Signal processing for nano- and molecular communications
• Energy efficiency and energy harvesting
• Large networks and big data in wireless communications
• Emerging techniques, technologies, and new waveforms for 5G

Full papers up to five-page limit should be submitted via EDAS.

March 1, 2017	Paper submission deadline
April 30, 2017	Acceptance notification
May 10, 2017	Final paper due

General Chairs
Yasutaka Ogawa
Wei Yu
Fumiyuki Adachi

Technical Program Chairs
Tomoaki Ohtsuki
Lutz Lampe
Wing-Kin (Ken) Ma

Special Session Chair
Tony Q. S. Quek

Technical Program Committee
Waheed Bajwa
Chong-Yung Chi
Philippe Ciblat
Min Dong
FeiFei Gao
Joakim Jaldén
Shi Jin
Yindi Jing
Markku Juntti
Eleftherios Karipidis
Erik G. Larsson
David J. Love
Matthew McKay
Urbashi Mitra
Chandra Murthy
Michael Rabbat
Alejandro Ribeiro
Ahmed Sadek
Mathini Sellathurai
Milica Stojanovic
Weifeng Su
Cihan Tepedelenlioglu
Wolfgang Utschick
Sergiy A. Vorobyov
Pengfei Xia
Rui Zhang
Wei Zhang

Finance Chairs
Yasushi Yamao
Hai Lin

Local Arrangement Chairs
Takeo Ohgane
Toshihiko Nishimura

Publication Chairs
Kazunori Hayashi
Kenichi Higuchi

Publicity Chairs
Julian Webber
Koichi Adachi

IEEE

IEEE Signal Processing MAGAZINE

Volume 33 | Number 5 | September 2016

COMPUTATIONAL PHOTOGRAPHY AND DISPLAYS

Interfacing with the Visual World

Recent Advances in Phase Retrieval

The MPEG Internet Video-Coding Standard

Highlights from the IEEE SP Cup 2016 Student Competition

IEEE Signal Processing Society

◆IEEE

Contents

Volume 33 | Number 5 | September 2016

FEATURES

See the "President's Message" and "Society News" in this issue!

ON THE COVER

IEEE Signal Processing Magazine is uniquely positioned to convey and embrace an evolving scope of signal processing. This issue showcases an example of a cross-disciplinary area—fascinating advances of computational photography and display. A variety of articles showcase the potential for the field to revolutionize imaging and displays and transform the way in which we capture, share, and interact with the visual world around us.

COVER IMAGE: ©ISTOCKPHOTO.COM/SCYTHER5

PG. 149

◆IEEE ORDER FORM FOR REPRINTS

Purchasing IEEE Papers in Print is easy, cost-effective and quick.
Complete this form, send via our secure fax (24 hours a day) to 732-981-8062 or mail it back to us.

PLEASE FILL OUT THE FOLLOWING

Author: _____

Publication Title: _____

Paper Title: _____

_____

RETURN THIS FORM TO:
IEEE Publishing Services
445 Hoes Lane
Piscataway, NJ 08855-1331

**Email the Reprint Department at
reprints@ieee.org for questions regarding
this form**

PLEASE SEND ME

■ 50 ■ 100 ■ 200 ■ 300 ■ 400 ■ 500 or _____ (in multiples of 50) reprints.

■ YES ■ NO Self-covering/title page required. COVER PRICE: $74 per 100, $39 per 50.

■ $58.00 Air Freight must be added for all orders being shipped outside the U.S.

■ $21.50 must be added for all USA shipments to cover the cost of UPS shipping and handling.

PAYMENT

■ Check enclosed. Payable on a bank in the USA.

■ Charge my: ■ Visa ■ Mastercard ■ Amex ■ Diners Club

Account # _____ Exp. date _____

Cardholder's Name (please print): _____

_____

■ Bill me (you must attach a purchase order) Purchase Order Number _____

Send Reprints to: Bill to address, if different:

_____   _____

_____   _____

_____   _____

Because information and papers are gathered from various sources, there may be a delay in receiving your reprint request. This is especially true with postconference publications. Please provide us with contact information if you would like notification of a delay of more than 12 weeks.

Telephone: _____ Fax: _____ Email Address: _____

2012 REPRINT PRICES (without covers)

Number of Text Pages

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40	41-44	45-48
50	$129	$213	$245	$248	$288	$340	$371	$408	$440	$477	$510	$543
100	$245	$425	$479	$495	$573	$680	$742	$817	$885	$953	$1021	$1088

Larger quantities can be ordered. Email reprints@ieee.org with specific details.

Tax Applies on shipments of regular reprints to CA, DC, FL, MI, NJ, NY, OH and Canada (GST Registration no. 12534188).
Prices are based on black & white printing. Please call us for full color price quote, if applicable.

Authorized Signature: _____ Date: _____

2016 IEEE MEMBERSHIP APPLICATION

(students and graduate students must apply online)

IEEE
Advancing Technology for Humanity

Start your membership immediately: Join online www.ieee.org/join

Please complete both sides of this form, typing or **printing in capital letters**. Use only English characters and abbreviate only if more than 40 characters and spaces per line. We regret that incomplete applications cannot be processed.

1 Name & Contact Information

Please PRINT your name as you want it to appear on your membership card and IEEE correspondence. As a key identifier for the IEEE database, circle your last/surname.

☐ Male   ☐ Female   Date of birth (Day/Month/Year) _____ / _____ / _____

Title   First/Given Name   Middle   Last/Family Surname

▼ **Primary Address** ☐ Home ☐ Business (All IEEE mail sent here)

Street Address

City   State/Province

Postal Code   Country

Primary Phone

Primary E-mail

▼ **Secondary Address** ☐ Home ☐ Business

Company Name   Department/Division

Street Address   City   State/Province

Postal Code   Country

Secondary Phone

Secondary E-mail

To better serve our members and supplement member dues, your postal mailing address is made available to carefully selected organizations to provide you with information on technical services, continuing education, and conferences. Your e-mail address is not rented by IEEE. Please check box only if you do not want to receive these postal mailings to the selected address. ☐

2 Attestation

I have graduated from a three- to five-year academic program with a university-level degree.
☐ Yes   ☐ No

This program is in one of the following fields of study:
☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

This academic institution or program is accredited in the country where the institution is located. ☐ Yes   ☐ No   ☐ Do not know

I have _____ years of professional experience in teaching, creating, developing, practicing, or managing within the following field:
☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

3 Please Tell Us About Yourself

Select the numbered option that best describes yourself. This information is used by IEEE magazines to verify their annual circulation. Please enter numbered selections in the boxes provided.

A. Primary line of business ⟶ ☐

1. Computers
2. Computer peripheral equipment
3. Software
4. Office and business machines
5. Test, measurement and instrumentation equipment
6. Communications systems and equipment
7. Navigation and guidance systems and equipment
8. Consumer electronics/appliances
9. Industrial equipment, controls and systems
10. ICs and microprocessors
11. Semiconductors, components, sub-assemblies, materials and supplies
12. Aircraft, missiles, space and ground support equipment
13. Oceanography and support equipment
14. Medical electronic equipment
15. OEM incorporating electronics in their end product (not elsewhere classified)
16. Independent and university research, test and design laboratories and consultants (not connected with a mfg. co.)
17. Government agencies and armed forces
18. Companies using and/or incorporating any electronic products in their manufacturing, processing, research or development activities
19. Telecommunications services, telephone (including cellular)
20. Broadcast services (TV, cable, radio)
21. Transportation services (airline, railroad, etc.)
22. Computer and communications and data processing services
23. Power production, generation, transmission and distribution
24. Other commercial users of electrical, electronic equipment and services (not elsewhere classified)
25. Distributor (reseller, wholesaler, retailer)
26. University, college/other educational institutions, libraries
27. Retired
28. Other_____

B. Principal job function ⟶ ☐

1. General and corporate management
2. Engineering management
3. Project engineering management
4. Research and development management
5. Design engineering management—analog
6. Design engineering management—digital
7. Research and development engineering
8. Design/development engineering—analog
9. Design/development engineering—digital
10. Hardware engineering
11. Software design/development
12. Computer science
13. Science/physics/mathematics
14. Engineering (not elsewhere specified)
15. Marketing/sales/purchasing
16. Consulting
17. Education/teaching
18. Retired
19. Other_____

C. Principal responsibility ⟶ ☐

1. Engineering and scientific management
2. Management other than engineering
3. Engineering design
4. Engineering
5. Software: science/mngmnt/engineering
6. Education/teaching
7. Consulting
8. Retired
9. Other_____

D. Title ⟶ ☐

1. Chairman of the Board/President/CEO
2. Owner/Partner
3. General Manager
4. VP Operations
5. VP Engineering/Dir. Engineering
6. Chief Engineer/Chief Scientist
7. Engineering Management
8. Scientific Management
9. Member of Technical Staff
10. Design Engineering Manager
11. Design Engineer
12. Hardware Engineer
13. Software Engineer
14. Computer Scientist
15. Dean/Professor/Instructor
16. Consultant
17. Retired
18. Other_____

Are you now or were you ever a member of IEEE?
☐ Yes ☐ No   If yes, provide, if known:

Membership Number   Grade   Year Expired

4 Please Sign Your Application

I hereby apply for IEEE membership and agree to be governed by the IEEE Constitution, Bylaws, and Code of Ethics. I understand that IEEE will communicate with me regarding my individual membership and all related benefits. **Application must be signed.**

Signature   Date

Over Please

(continued on next page)

5 — Add IEEE Society Memberships (Optional)

The 39 IEEE Societies support your technical and professional interests. Many society memberships include a personal subscription to the core journal, magazine, or newsletter of that society. **For a complete list of everything included with your IEEE Society membership, visit www.ieee.org/join.** All prices are quoted in US dollars.

Please check ✔ the appropriate box.

Society	Code	BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
IEEE Aerospace and Electronic Systems ⌘ ☑	AES010	25.00 ☐	12.50 ☐
IEEE Antennas and Propagation ⌘ ☑	AP003	15.00 ☐	7.50 ☐
IEEE Broadcast Technology ▪ ⌘ ☑	BT002	15.00 ☐	7.50 ☐
IEEE Circuits and Systems ▪ ⌘ ☑	CAS004	22.00 ☐	11.00 ☐
IEEE Communications ⌘ ☑	COM019	30.00 ☐	15.00 ☐
IEEE Components, Packaging, & Manu. Tech. ▪ ⌘ ☑	CPMT021	15.00 ☐	7.50 ☐
IEEE Computational Intelligence ⌘ ☑	CIS011	29.00 ☐	14.50 ☐
IEEE Computer ▪ ⌘ ☑	C016	56.00 ☐	28.00 ☐
IEEE Consumer Electronics ▪ ⌘ ☑	CE008	20.00 ☐	10.00 ☐
IEEE Control Systems ⌘ ☑	CS023	25.00 ☐	12.50 ☐
IEEE Dielectrics and Electrical Insulation ⌘ ☑	DEI032	26.00 ☐	13.00 ☐
IEEE Education ⌘ ☑	E025	20.00 ☐	10.00 ☐
IEEE Electromagnetic Compatibility ▪ ⌘ ☑	EMC027	31.00 ☐	15.50 ☐
IEEE Electron Devices ▪ ⌘ ☑	ED015	18.00 ☐	9.00 ☐
IEEE Engineering in Medicine and Biology ⌘ ☑	EMB018	40.00 ☐	20.00 ☐
IEEE Geoscience and Remote Sensing ⌘ ☑	GRS029	19.00 ☐	9.50 ☐
IEEE Industrial Electronics ⌘ ☑	IE013	9.00 ☐	4.50 ☐
IEEE Industry Applications ⌘ ☑	IA034	20.00 ☐	10.00 ☐
IEEE Information Theory ▪ ⌘ ☑	IT012	30.00 ☐	15.00 ☐
IEEE Instrumentation and Measurement ⌘ ☑	IM009	29.00 ☐	14.50 ☐
IEEE Intelligent Transportation Systems ▪ ⌘ ☑	ITSS038	35.00 ☐	17.50 ☐
IEEE Magnetics ▪ ☑ ♦ ⌘	MAG033	26.00 ☐	13.00 ☐
IEEE Microwave Theory and Techniques ⌘ ☑	MTT017	17.00 ☐	8.50 ☐
IEEE Nuclear and Plasma Sciences ▪ ⌘ ☑	NPS005	35.00 ☐	17.50 ☐
IEEE Oceanic Engineering ▪ ⌘ ☑	OE022	19.00 ☐	9.50 ☐
IEEE Photonics ▪ ⌘ ☑	PHO036	34.00 ☐	17.00 ☐
IEEE Power Electronics ⌘ ☑	PEL035	25.00 ☐	12.50 ☐
IEEE Power & Energy ▪ ⌘ ☑	PE031	35.00 ☐	17.50 ☐
IEEE Product Safety Engineering ▪	PSE043	35.00 ☐	17.50 ☐
IEEE Professional Communication ▪ ⌘ ☑	PC026	31.00 ☐	15.50 ☐
IEEE Reliability ▪ ⌘ ☑	RL007	35.00 ☐	17.50 ☐
IEEE Robotics and Automation ⌘ ☑	RA024	9.00 ☐	4.50 ☐
IEEE Signal Processing ▪ ⌘ ☑	SP001	22.00 ☐	11.00 ☐
IEEE Social Implications of Technology ▪ ⌘ ☑	SIT030	33.00 ☐	16.50 ☐
IEEE Solid-State Circuits ⌘ ☑	SSC037	22.00 ☐	11.00 ☐
IEEE Systems, Man, & Cybernetics ⌘ ☑	SMC028	12.00 ☐	6.00 ☐
IEEE Technology & Engineering Management ⌘ ☑	TEM014	35.00 ☐	17.50 ☐
IEEE Ultrasonics, Ferroelectrics, & Frequency Control ▪ ⌘ ☑	UFFC020	20.00 ☐	10.00 ☐
IEEE Vehicular Technology ▪ ⌘ ☑	VT006	18.00 ☐	9.00 ☐

Legend—Society membership includes:
⌘ One or more Society publications
▪ Society newsletter
☑ Online access to publication
♦ CD-ROM of selected society publications

Complete both sides of this form, sign, and return to:
IEEE MEMBERSHIP APPLICATION PROCESSING
445 HOES LN, PISCATAWAY, NJ 08854-4141 USA
or fax to +1 732 981 0225
or join online at www.ieee.org/join

Please reprint your full name here

6 — 2016 IEEE Membership Rates
(student rates available online)

IEEE member dues and regional assessments are based on where you live and when you apply. Membership is based on the calendar year from 1 January through 31 December. All prices are quoted in US dollars.

Please check ✔ the appropriate box.

RESIDENCE	BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
United States	$197.00 ☐	$98.50 ☐
Canada (GST)*	$173.35 ☐	$86.68 ☐
Canada (NB, NF and ON HST)*	$185.11 ☐	$92.56 ☐
Canada (Nova Scotia HST)*	$188.05 ☐	$94.03 ☐
Canada (PEI HST)*	$186.58 ☐	$93.29 ☐
Canada (GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

*IEEE Canada Business No. 125634188

Minimum Income or Unemployed Provision
Applicants who certify that their prior year income did not exceed US$14,700 (or equivalent) or were not employed are granted 50% reduction in: full-year dues, regional assessment and fees for one IEEE Membership plus one Society Membership. If applicable, please check appropriate box and adjust payment accordingly. Student members are not eligible.

☐ I certify I earned less than US$14,700 in 2015
☐ I certify that I was unemployed in 2015

7 — More Recommended Options

Proceedings of the IEEE .... print $47.00 ☐ **or** online $41.00 ☐
Proceedings of the IEEE (print/online combination) ....$57.00 ☐
IEEE Standards Association (IEEE-SA) ....$53.00 ☐
IEEE Women in Engineering (WIE) ....$25.00 ☐

8 — Payment Amount

Please total the Membership dues, Society dues, and other amounts from this page:
IEEE Membership dues ⑥ ....$_____
IEEE Society dues (optional) ⑤ ....$_____
IEEE-SA/WIE dues (optional) ⑦ ....$_____
Proceedings of the IEEE (optional) ⑦ ....$_____
Canadian residents pay 5% GST or appropriate HST (BC—12%; NB, NF, ON-13%;NS-15%) on Society payments & publications only.....................TAX $_____
AMOUNT PAID .... **TOTAL** $_____

Payment Method
All prices are quoted in US dollars. You may pay for IEEE membership by credit card (see below), check, or money order payable to IEEE, drawn on a US bank.

☐ Check   ☐ VISA   ☐ MasterCard   ☐ Diners Club   ☐ American Express   ☐ Discover

Credit Card Number

MONTH / YEAR EXPIRATION DATE   CARDHOLDER'S 5-DIGIT ZIPCODE (BILLING STATEMENT ADDRESS) USA ONLY

Name as it appears on card

Signature

Auto Renew my Memberships and Subscriptions (available when paying by credit card).
☐ I agree to the Terms and Conditions located at www.ieee.org/autorenew

9 — Were You Referred to IEEE?

☐ Yes   ☐ No   If yes, provide the following:
Member Recruiter Name _____
IEEE Recruiter's Member Number (Required) _____

CAMPAIGN CODE   PROMO CODE

15-MEM-385-P 6/15

Information for Authors

(Updated/Effective January 2015)

For Transactions and Journals:

Authors are encouraged to submit manuscripts of Regular papers (papers which provide a complete disclosure of a technical premise), or Comment Correspondences (brief items that provide comment on a paper previously published in these TRANSACTIONS). Submissions/resubmissions must be previously unpublished and may not be under consideration elsewhere.

Every manuscript must:

    i. provide a clear statement of the problem and what the contribution of the work is to the relevant research community;

    ii. state why this contribution is significant (what impact it will have);

    iii. provide citation of the published literature most closely related to the manuscript; and

    iv. state what is distinctive and new about the current manuscript relative to these previously published works.

By submission of your manuscript to these TRANSACTIONS, all listed authors have agreed to the authorship list and all the contents and confirm that the work is original and that figures, tables and other reported results accurately reflect the experimental work. In addition, the authors all acknowledge that they accept the rules established for publication of manuscripts, including agreement to pay all overlength page charges, color charges, and any other charges and fees associated with publication of the manuscript. Such charges are not negotiable and cannot be suspended. The corresponding author is responsible for obtaining consent from all co-authors and, if needed, from sponsors before submission.

In order to be considered for review, a paper must be within the scope of the journal and represent a novel contribution. A paper is a candidate for an Immediate Rejection if it is of limited novelty, e.g. a straightforward combination of theories and algorithms that are well established and are repeated on a known scenario. Experimental contributions will be rejected without review if there is insufficient experimental data. These TRANSACTIONS are published in English. Papers that have a large number of typographical and/or grammatical errors will also be rejected without review.

In addition to presenting a novel contribution, acceptable manuscripts must describe and cite related work in the field to put the contribution in context. Do not give theoretical derivations or algorithm descriptions that are easily found in the literature; merely cite the reference.

New and revised manuscripts should be prepared following the "Manuscript Submission" guidelines below, and submitted to the online manuscript system, ScholarOne Manuscripts. Do not send original submissions or revisions directly to the Editor-in-Chief or Associate Editors; they will access your manuscript electronically via the ScholarOne Manuscript system.

Manuscript Submission. Please follow the next steps.

1. *Account in ScholarOne Manuscripts.* If necessary, create an account in the on-line submission system ScholarOne Manuscripts. Please check first if you already have an existing account which is based on your e-mail address and may have been created for you when you reviewed or authored a previous paper.

2. *Electronic Manuscript.* Prepare a PDF file containing your manuscript in double-column, single-spaced format using a font size of 10 points or larger, having a margin of at least 1 inch on all sides. Upload this version of the manuscript as a PDF file "double.pdf" to the ScholarOne-Manuscripts site. Since many reviewers prefer a larger font, you are strongly encouraged to also submit a single-column, double-spaced version (11 point font or larger), which is easy to create with the templates provided **IEEE Author Digital Toolbox (http://www.ieee.org/publications_standards/publications/authors/authors_journals.html)**. Page length restrictions will be determined by the double-column

    Digital Object Identifier

version. Proofread your submission, confirming that all figures and equations are visible in your document before you "SUBMIT" your manuscript. Proofreading is critical; once you submit your manuscript, the manuscript cannot be changed in any way. You may also submit your manuscript as a .PDF or MS Word file. The system has the capability of converting your files to PDF, however it is your responsibility to confirm that the conversion is correct and there are no font or graphics issues prior to completing the submission process.

3. *EDICS (Not applicable to Journal of Selected Topics in Signal Processing).* All submissions must be classified by the author with an EDICS (Editors' Information Classification Scheme) selected from the list of EDICS published online at the at the publication's EDICS webpage (*please see the list below). Upon submission of a new manuscript, please choose the EDICS categories that best suit your manuscript. Failure to do so will likely result in a delay of the peer review process.

4. *Additional Documents for Review.* Please upload pdf versions of all items in the reference list that are not publicly available, such as unpublished (submitted) papers. Graphical abstracts and supplemental materials intended to appear with the final paper (see below) must also be uploaded for review at the time of the initial submission for consideration in the review process. Use short filenames without spaces or special characters. When the upload of each file is completed, you will be asked to provide a description of that file.

5. *Supplemental Materials.* IEEE Xplore can publish multimedia files (audio, images, video), datasets, and software (e.g. Matlab code) along with your paper. Alternatively, you can provide the links to such files in a README file that appears on Xplore along with your paper. For details, please see IEEE Author Digital Toolbox under "Multimedia." To make your work reproducible by others, these TRANSACTIONS encourages you to submit all files that can recreate the figures in your paper.

6. *Submission.* After uploading all files and proofreading them, submit your manuscript by clicking "Submit." A confirmation of the successful submission will open on screen containing the manuscript tracking number and will be followed with an e-mail confirmation to the corresponding and all contributing authors. Once you click "Submit," your manuscript cannot be changed in any way.

7. *Copyright Form and Consent Form.* By policy, IEEE owns the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers; and to facilitate the appropriate reuse of this material by others. To comply with the IEEE copyright policies, authors are required to sign and submit a completed "IEEE Copyright and Consent Form" prior to publication by the IEEE. The IEEE recommends authors to use an effective electronic copyright form (eCF) tool within the ScholarOne Manuscripts system. You will be redirected to the "IEEE Electronic Copyright Form" wizard at the end of your original submission; please simply sign the eCF by typing your name at the proper location and click on the "Submit" button.

Comment Correspondence. Comment Correspondences provide brief comments on material previously published in these TRANSACTIONS. These items may not exceed 2 pages in double-column, single spaced format, using 9 point type, with margins of 1 inch minimum on all sides, and including: title, names and contact information for authors, abstract, text, references, and an appropriate number of illustrations and/or tables. Correspondence items are submitted in the same way as regular manuscripts (see "Manuscript Submission" above for instructions).

Authors may also submit manuscripts of overview articles, but note that these include an additional white paper approval process http://www.signalprocessingsociety.org/publications/overview-articles/. *[This does not apply to the Journal of Selected Topics in Signal Processing. Please contact the Editor-in-Chief.]*

Manuscript Length. For the initial submission of a regular paper, the manuscript may not exceed 13 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. Supplemental materials and graphical abstracts are not included in the page count. For regular papers, the revised manuscript may not exceed 16 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. For Overview Papers, the maximum length is double that for regular submissions at each stage (please reference http://www.signalprocessingsociety.org/publications/overview-articles/ for more information).

Note that any paper in excess of 10 pages will be subject to mandatory overlength page charges. Since changes recommended as a result of peer review may require additions to the manuscript, it is strongly recommended that you practice economy in preparing original submissions. Note: Papers submitted to the TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.

Exceptions to manuscript length requirements may, under extraordinary circumstances, be granted by the Editor-in-Chief. However, such exception does not obviate your requirement to pay any and all overlength or additional charges that attach to the manuscript.

Resubmission of Previously Rejected Manuscripts. Authors of manuscripts rejected from any journal are allowed to resubmit their manuscripts only once. At the time of submission, you will be asked whether your manuscript is a new submission or a resubmission of an earlier rejected manuscript. If it is a resubmission of a manuscript previously rejected by any journal, you are expected to submit supporting documents identifying the previous submission and detailing how your new version addresses all of the reviewers' comments. Papers that do not disclose connection to a previously rejected paper or that do not provide documentation as to changes made may be immediately rejected.

Author Misconduct. Author misconduct includes plagiarism, self-plagiarism, and research misconduct, including falsification or misrepresentation of results. All forms of misconduct are unacceptable and may result in sanctions and/or other corrective actions. Plagiarism includes copying someone else's work without appropriate credit, using someone else's work without clear delineation of citation, and the uncited reuse of an author's previously published work that also involves other authors. Self-plagiarism involves the verbatim copying or reuse of an authors own prior work without appropriate citation, including duplicate submission of a single journal manuscript to two different journals, and submission of two different journal manuscripts which overlap substantially in language or technical contribution. For more information on the definitions, investigation process, and corrective actions related to author misconduct, see the Signal Processing Society Policies and Procedures Manual, Section 6.1. http://www.signalprocessingsociety.org/about-sps/governance/policy-procedure/part-2. Author misconduct may also be actionable by the IEEE under the rules of Member Conduct.

Extensions of the Author's Prior Work. It is acceptable for conference papers to be used as the basis for a more fully developed journal submission. Still, authors are required to cite their related prior work; the papers cannot be identical; and the journal publication must include substantively novel aspects such as new experimental results and analysis or added theoretical work. The journal publication should clearly specify how the journal paper offers novel contributions when citing the prior work. Limited overlap with prior journal publications with a common author is allowed only if it is necessary for the readability of the paper, and the prior work must be cited as the primary source.

Submission Format. Authors are required to prepare manuscripts employing the on-line style files developed by IEEE, which include guidelines for abbreviations, mathematics, and graphics. All manuscripts accepted for publication will require the authors to make final submission employing these style files. The style files are available on the web at the **IEEE Author Digital Toolbox** under "Template for all TRANSACTIONS." (LaTeX and MS Word). Please note the following requirements about the abstract:

- The abstract must be a concise yet comprehensive reflection of what is in your article.
- The abstract must be self-contained, without abbreviations, footnotes, displayed equations, or references.

- The abstract must be between 150-250 words.
- The abstract should include a few keywords or phrases, as this will help readers to find it. Avoid over-repetition of such phrases as this can result in a page being rejected by search engines.

In addition to written abstracts, papers may include a graphical abstract; see http://www.ieee.org/publications_standards/publications/authors/authors_journals.html. for options and format requirements.

IEEE supports the publication of author names in the native language alongside the English versions of the names in the author list of an article. For more information, see "Author names in native languages" (http://www.ieee.org/publications_standards/publications/authors/auth_names_native_lang.pdf) on the IEEE Author Digital Toolbox page.

Open Access. The publication is a hybrid journal, allowing either Traditional manuscript submission or Open Access (author-pays OA) manuscript submission. Upon submission, if you choose to have your manuscript be an Open Access article, you commit to pay the discounted $1,750 OA fee if your manuscript is accepted for publication in order to enable unrestricted public access. Any other application charges (such as overlength page charge and/or charge for the use of color in the print format) will be billed separately once the manuscript formatting is complete but prior to the publication. If you would like your manuscript to be a Traditional submission, your article will be available to qualified subscribers and purchasers via IEEE Xplore. No OA payment is required for Traditional submission.

Page Charges.

Voluntary Page Charges. Upon acceptance of a manuscript for publication, the author(s) or his/her/their company or institution will be asked to pay a charge of $110 per page to cover part of the cost of publication of the first ten pages that comprise the standard length (two pages, in the case of Correspondences).

Mandatory Page Charges The author(s) or his/her/their company or institution will be billed $220 per each page in excess of the first ten published pages for regular papers and six published pages for correspondence items. (**NOTE: Papers accepted to IEEE TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.) These are mandatory page charges and the author(s) will be held responsible for them. They are not negotiable or voluntary. The author(s) signifies his willingness to pay these charges simply by submitting his/her/their manuscript to the TRANSACTIONS. The Publisher holds the right to withhold publication under any circumstance, as well as publication of the current or future submissions of authors who have outstanding mandatory page charge debt. No mandatory overlength page charges will be applied to overview articles in the Society's journals.

Color Charges. Color figures which appear in color only in the electronic (Xplore) version can be used free of charge. In this case, the figure will be printed in the hardcopy version in grayscale, and the author is responsible that the corresponding grayscale figure is intelligible. Color reproduction charges for print are the responsibility of the author. Details of the associated charges can be found on the IEEE Publications page.

Payment of fees on color reproduction is not negotiable or voluntary, and the author's agreement to publish the manuscript in these TRANSACTIONS is considered acceptance of this requirement.

*EDICS Webpages:

IEEE TRANSACTIONS ON SIGNAL PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/tsp/TSP-EDICS/
IEEE TRANSACTIONS ON IMAGE PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/image-processing/tip-edics/
IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE / ACM:
http://www.signalprocessingsociety.org/publications/periodicals/taslp/taslp-edics/
IEEE TRANSACTIONS ON INFORMATION, FORENSICS AND SECURITY:
http://www.signalprocessingsociety.org/publications/periodicals/forensics/forensics-edics/
IEEE TRANSACTIONS ON MULTIMEDIA:
http://www.signalprocessingsociety.org/tmm/tmm-edics/
IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING:
http://www.signalprocessingsociety.org/publications/periodicals/tci/tci-edics/
IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS:
http://www.signalprocessingsociety.org/publications/periodicals/tsipn/tsipn-edics/

2016 IEEE SIGNAL PROCESSING SOCIETY MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

❖ IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: SEND MAIL TO: ☐ Home Address **OR** ☐ Business/School Address
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE   FIRST OR GIVEN NAME   MIDDLE NAME   SURNAME/LAST NAME

HOME ADDRESS

CITY   STATE/PROVINCE   POSTAL CODE   COUNTRY

2. Are you now or were you ever a member of IEEE? ☐ Yes ☐ No
If yes, please provide, if known:

MEMBERSHIP NUMBER _____

Grade _____ **Year Membership Expired:** _____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position   Years in Current Position

Years in the Profession Since Graduation   ☐ PE State/Province

Street Address

City   State/Province   Postal Code   Country

4. EDUCATION
A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A. _____

Baccalaureate Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

B. _____

Highest Technical Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

5. _____
Full signature of applicant

6. DEMOGRAPHIC INFORMATION   – ALL APPLICANTS -
Date Of Birth _____   ☐ Male ☐ Female
Day   Month   Year

7. CONTACT INFORMATION

Office Phone/Office Fax   Home Phone/Home Fax

Office E-Mail   Home E-Mail

8. 2016 IEEE MEMBER RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$197.00 ☐	$98.50 ☐
Canada (incl. GST)	$173.35 ☐	$86.68 ☐
Canada (incl. HST for PEI)	$186.58 ☐	$93.29 ☐
Canada (incl. HST for Nova Scotia)	$188.05 ☐	$94.03 ☐
Canada (incl. HST for NB, NF and ON)	$185.11 ☐	$92.56 ☐
Canada (incl. GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%, 13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)
Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)
U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

	16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*	$ 22.00	$ 11.00

Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).
Add $17 to enhance SPS Membership and also receive: $17.00 $ 8.50
IEEE Signal Processing Magazine (print) and **SPS Digital Library:** online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Communications, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.

Publications available only with SPS membership:

		Full	Half
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.   Amount Paid $_____

9.

IEEE Membership Affiliate Fee (See pricing in Section 8) $_____
Signal Processing Society Fees $_____
Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only   Tax $_____
AMOUNT PAID WITH APPLICATION   TOTAL $_____
Prices subject to change without notice.
☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express**   ☐ **VISA**   ☐ **MasterCard**
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card   Date

10. WERE YOU REFERRED?
☐ Yes ☐ No   If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

2016 IEEE SIGNAL PROCESSING SOCIETY STUDENT MEMBERSHIP APPLICATION

(Current and reinstating IEEE members joining SPS complete areas 1, 2, 8, 9.)
Mail to: IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

◆ IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: SEND MAIL TO: ☐ Home Address OR ☐ Business/School Address
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE   FIRST OR GIVEN NAME   MIDDLE NAME   SURNAME/LAST NAME

HOME ADDRESS

CITY   STATE/PROVINCE   POSTAL CODE   COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ Yes ☐ No
If yes, please provide, if known:

MEMBERSHIP NUMBER _____

Grade_____ Year Membership Expired:_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position   Years in Current Position

Years in the Profession Since Graduation   ☐ PE State/Province

Street Address

City   **State/Province**   **Postal Code**   **Country**

4. EDUCATION
A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A. _____
Baccalaureate Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

B. _____
Highest Technical Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

5.
Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -
Date Of Birth _____
    Day   Month   Year
☐ Male   ☐ Female

7. CONTACT INFORMATION

Office Phone/Office Fax   Home Phone/Home Fax

Office E-Mail   Home E-Mail

8. 2016 IEEE STUDENT MEMBER RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$32.00 ☐	$16.00 ☐
Canada (incl. GST)	$33.60 ☐	$16.80 ☐
Canada (incl. HST for NB, NF, and ON)	$36.16 ☐	$18.08 ☐
Canada (incl. HST for Nova Scotia)	$36.80 ☐	$18.40 ☐
Canada (incl. HST for PEI)	$36.48 ☐	$18.24 ☐
Canada (incl. GST and QST Quebec)	$36.79 ☐	$18.40 ☐
Africa, Europe, Middle East, Latin America, Asia, Pacific	$27.00 ☐	$13.50 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%,13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)
Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)
U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS STUDENT MEMBER RATES

	16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*	$11.00 ☐	$ 5.50 ☐

Fee includes: IEEE Signal Processing Magazine (electronic and digital), Inside Signal Processing eNewsletter (electronic) and IEEE Signal Processing Society Content Gazette (electronic).

Add $8 to enhance SPS Membership and also receive: $ 9.00 ☐ $ 4.50 ☐
IEEE Signal Processing Society Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.

Publications available only with SPS membership:

		Full	Half
Signal Processing, IEEE Transactions on:	Print	$105.00 ☐	$ 52.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$ 80.00 ☐	$ 40.00 ☐
Image Processing, IEEE Transactions on:	Print	$104.00 ☐	$ 52.00 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$ 90.00 ☐	$ 45.00 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$ 88.00 ☐	$ 44.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 18.00 ☐	$ 9.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 28.00 ☐	$ 14.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 33.00 ☐	$ 16.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Big Data	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Molecular, Biological, & Multi-Scale Communications	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Internet of Things Journal	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 15.00 ☐	$ 7.50 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Biometrics Compendium:	Online	$ 15.00 ☐	$ 7.50 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 37.00 ☐	$ 18.50 ☐
	Electronic	$ 27.00 ☐	$ 13.50 ☐
	Print & Electronic	$ 45.00 ☐	$ 22.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 21.00 ☐	$ 10.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 17.00 ☐	$ 8.50 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 13.00 ☐	$ 6.50 ☐
	Print & Electronic	$ 20.00 ☐	$ 10.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
IEEE Sensors Journal:	Electronic	$ 20.00 ☐	$ 10.00 ☐
Smart Grid, IEEE Transactions on:	Print	$ 50.00 ☐	$ 25.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 60.00 ☐	$ 30.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$ 64.00 ☐	$ 32.00 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 64.00 ☐	$ 32.00 ☐
IEEE Wireless Communications Letters:	Electronic	$ 10.00 ☐	$ 5.00 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.   Amount Paid $_____

9.

IEEE Membership Fee (See pricing in Section 8)	$_____
Signal Processing Society Fees	$_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only   Tax $_____

AMOUNT PAID WITH APPLICATION   TOTAL $_____
Prices subject to change without notice.
☐ Check or money order enclosed Payable to IEEE on a U.S. Bank
☐ American Express   ☐ VISA   ☐ MasterCard   ☐ Diners Club

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card   Date

10. WERE YOU REFERRED?
☐ Yes ☐ No   If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

2016 IEEE SIGNAL PROCESSING SOCIETY AFFILIATE MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: SEND MAIL TO: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE   FIRST OR GIVEN NAME   MIDDLE NAME   SURNAME/LAST NAME

HOME ADDRESS

CITY   STATE/PROVINCE   POSTAL CODE   COUNTRY

2.
Are you now or were you ever a member of IEEE? ☐ Yes ☐ No
If yes, please provide, if known:
MEMBERSHIP NUMBER _____ | | | | | | | |

Grade_____Year Membership Expired:_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position   Years in Current Position

Years in the Profession Since Graduation   ☐ PE State/Province

Street Address

City   State/Province   Postal Code   Country

4. EDUCATION
A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.
Baccalaureate Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

B.
Highest Technical Degree Received   Program/Course of Study

College/University   Campus

State/Province   Country   Mo./Yr. Degree Received

5.
Full signature of applicant

6. DEMOGRAPHIC INFORMATION   – ALL APPLICANTS -

Date Of Birth _____   ☐ Male ☐ Female
             Day   Month   Year

7. CONTACT INFORMATION

Office Phone/Office Fax   Home Phone/Home Fax

Office E-Mail   Home E-Mail

8.   2016 IEEE AFFILIATE RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$73.50 ☐	$36.75 ☐
Canada (incl. GST)	$77.18 ☐	$38.59 ☐
Canada (incl. HST for PEI)	$83.79 ☐	$41.89 ☐
Canada (incl. HST for Nova Scotia)	$84.58 ☐	$42.26 ☐
Canada (incl. HST for NB, NF and ON)	$83.06 ☐	$41.53 ☐
Canada (incl. GST and QST Quebec)	$84.51 ☐	$42.25 ☐
Africa, Europe, Middle East	$73.50 ☐	$36.75 ☐
Latin America	$73.50 ☐	$36.75 ☐
Asia, Pacific	$73.50 ☐	$36.75 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 13%, 14% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)
Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)
U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

	16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*	$ 22.00 ☐	$ 11.00 ☐

Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).

Add $17 to enhance SPS Membership and also receive:	$17.00 ☐	$ 8.50 ☐

IEEE Signal Processing Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.

Publications available only with SPS membership:

Publication			
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.   Amount Paid $_____

9.
IEEE Membership Affiliate Fee (See pricing in Section 8) $_____

Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only   Tax $_____

AMOUNT PAID WITH APPLICATION   TOTAL $_____
Prices subject to change without notice.
☐ Check or money order enclosed Payable to IEEE on a U.S. Bank
☐ **American Express**   ☐ VISA   ☐ MasterCard
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card   Date

10. WERE YOU REFERRED?
☐ Yes ☐ No   If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____