Tao Mei
Microsoft Research Asia

A Light-Field Journey to Virtual Reality

Jingyi Yu
University of Delaware

With Facebook's acquisition of the VR startup Oculus (www.oculus.com) in 2014 for approximately US$2 billion, VR reemerged from research laboratories better prepared for success in the public marketplace. Since then, other tech giants—including Microsoft, Google, Sony, HTC, and Samsung—have released popular VR and augmented reality (AR) headsets—such as the Sony PlayStation VR (www.playstation.com) and HTC Vive (www.vive.com)—to provide fun and exciting experiences. Furthermore, smaller startup companies, such as 8i (https://8i.com), Otoy (https://home.otoy.com), and Lytro (www.lytro.com), have started focusing on producing ultra-high-quality content, recognizing that this will be the next grand challenge in the VR industry.

Producing high-quality VR content is difficult, because our eyes are very good at distinguishing between what is and isn't real. For example, the binocular stereoscopy of human eyes is largely missing in traditional 360-degree panoramic video-capture solutions. More advanced imaging systems, such as ones equipped with depth cameras, can recover certain aspects of 3D geometry to generate stereoscopic pairs. However, partial 3D is still insufficient to provide motion parallax, which is essential for providing a realistic viewing experience. In addition to stereo and motion parallax, human eyes also achieve 3D vision through refocusing, where only the object in the focal plane is clear and objects at other distances are blurry. The real world presents extremely high-dimensional data, comprising geometry, lighting, surface reflectance, and so on. We must record every piece of information in this high-dimensional dataset to faithfully reproduce reality.

Consequently, to create a virtual reality that even the human eye cannot distinguish from the real world, we must achieve the perfect immersive viewing experience, such that human viewers feel they can walk into the scene. This is known as the virtual *walk-in effect,* and it requires *light-field technology*—3D imaging technology that emerged from the field of computational imaging/photography to capture the light rays that people perceive from different locations and directions. When combined with computer vision and machine learning, light-field technology provides a viable path for producing low-cost, high-quality VR content, positioning this technology to be the most profitable segment of the VR industry. In fact, as VR technology enters everyday life, the frontier of light-field VR will become increasingly attractive by emulating real 3D worlds, including complex objects, humans, and large environments.

Human Visual Perception

VR has rapidly become popular because of its ability to create an immersive experience in which people can observe and feel the content. To achieve this, the virtual scene must be presented in the form of a 3D panoramic image or video, and the scene should change rapidly to accommodate human head motion. It should also provide sufficient depth information for the human eye and brain to determine the range of objects and simultaneously change focus.

Human beings acquire most of their information about the world through their eyes. Shaped by physiological evolution and extensive daily training, the human eye has cultivated an effective mechanism to perceive the external world, and this mechanism of perception is ultra-sensitive. For example, the human eye is very sensitive to depth information, which is perceived through clues such as binocular parallax, motion parallax, occlusion, and convergence. More specifically, when people

1070-986X/17/$33.00 © 2017 IEEE

Published by the IEEE Computer Society

observe an object in the real world, their eyes are constantly changing focus and, via the vestibular reflex, the resulting jitter and blur are eliminated. The real world is visualized via numerous light rays, which the human eye captures from different locations and directions by changing its focus. With head motion, human eyes can always focus on the object of interest, even if the object is surrounded by a changing environment.

The collection of light rays that people perceive from different locations and directions in their surroundings is the light field.

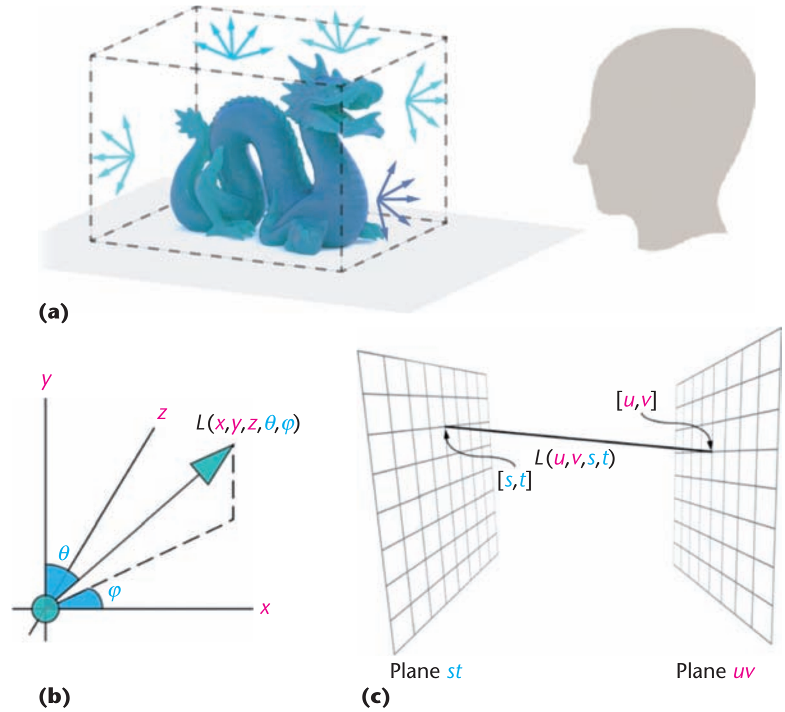

Figure 1. The concept of the light field, (a) which is a vector function that describes the amount of light flowing in every direction through every point in space. The human eye perceives the light field emitting from the surrounding environment. (b) The direction of each ray is given by the 5D plenoptic function, and the magnitude of each ray is given by the radiance. (c) The two-plane parameterization can be used to reduce the plenoptic function to 4D.

Why Light Fields?

When wearing a VR headset, the human eye is actually focused on the viewing screen inside the headset, while the virtual reality is trying to "deceive" the human eye and brain, as if the eyes were focused at different depths. Once there is some error or delay in displaying the content, the human eye will be attuned to perceiving these artifacts, which subsequently could cause uncomfortable symptoms such as headaches, dizziness, or fatigue. In contrast, if human eyes can directly perceive the VR content in the form of a light field, many of these visual discomforts would be significantly reduced if not completely eliminated.

In the past decade, researchers from both computer vision and computational photography have developed different means to record and reproduce the complete light field. The most notable examples are commodity light-field cameras, such as those from Lytro and Raytrix (www.raytrix.de), and light-field camera arrays,[1,2] which enable easy acquisition of light fields at various scales. Besides imaging solutions, recent efforts have focused on applying light fields to resolve challenging problems in computer vision and robotics, including stereo matching and 3D reconstruction, stereoscopy synthesis, saliency detection, surveillance, and recognition. There has also been useful research in several geometrical aspects of light-field cameras, such as calibration. In 2014, my colleagues and I organized the first Light Field for Computer Vision (LF4CV) workshop (www.eecis.udel.edu/~yu/LF4CV/index.html), in conjunction with the European Conference on Computer Vision (ECCV 14) and a special issue on the topic in the journal, *Computer Vision and Image Understanding*.[3] This year, we're organizing the second LF4CV workshop in conjunction with the conference on Computer Vision and Pattern Recognition (CVPR 17).

Understanding Light Fields

The phrase "light field" was first proposed by Andrey Gershun in 1936, describing the radiance of light at a certain point given its direction in 3D space.[4] More than 50 years later, Edward Adelson and James Bergen proposed using the plenoptic function to represent the light field,[5] which encodes the 3D spatial and 2D directional information of a light ray. The plenoptic function is a 5D parametric representation of the light field. Then, in 1996, Marc Levoy and Pat Hanrahan observed that the radiance of a light ray remains unchanged along its propagation direction in free space.[6] As a result, they proposed a method to represent the light field using the coordinates of the points at which a light intersects two arbitrary planes. In this manner, the 5D plenoptic function is reduced to 4D, as shown in Figure 1.

The light field records all the spatial information and direction information of the light, so it's easy to reconstruct images of different focus

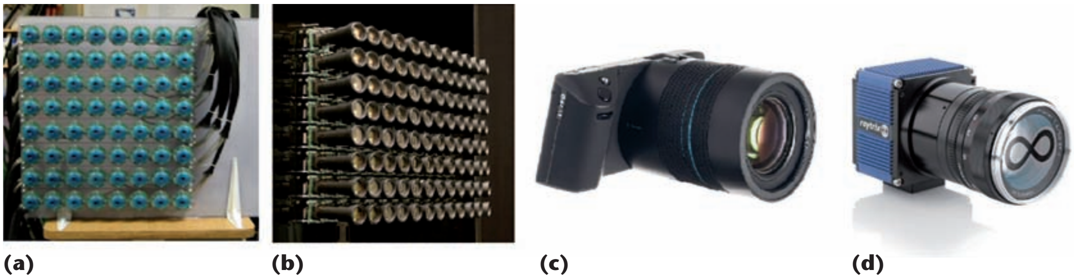

Figure 2. Light-field imaging systems: (a) the MIT camera array; (b) Stanford camera array; (c) Lytro light-field camera, Illum; and (d) Raytrix light-field camera.

depths, render images at different viewing angles, and even synthesize an image at a new position. In VR, stereoscopic images or video can be easily constructed from the light-field data, so the resulting stereo image is more realistic than the image produced using traditional binocular-vision reconstruction methods. This is because the traditional binocular-vision synthesis of a 3D image is essentially flat, whereas through the light field, the synthesized image is a real 3D image.

The development of light-field technology has brought great advances and changes to VR applications. It meets almost all the requirements for VR: stereoscopic parallax can bring people a realistic experience, motion parallax allows people to feel more natural when walking around, and refocusing can make people feel more comfortable with VR content.

Recording Light Fields

The core issue in light-field VR technology is how to record the light-field data and reproduce the light field for the human eye without any loss of either geometric or photometric information. Light fields can be captured using either a camera array or a light-field camera.

Stanford University and the Massachusetts Institute of Technology built multicamera array light-field capture systems, as shown in Figure 2a and 2b. By using a camera array, a high-resolution image can be captured from different viewing perspectives, such that each image corresponds to a 2D slice of the 4D light field. If enough different 2D slices are captured, the 4D light field can be completely reconstructed. The resulting light-field data can be conveniently used to render images from a new perspective, render images at different focus depths, or even increase the resolution and dynamic range of an image or the frame rate of a video.

The other method for obtaining a light field is to use a light-field camera based on a microlens array. A light-field camera is a new type of camera, which provides a new refocusing-after-capture feature. The most notable example is Lytro, which has released two models of the light-field camera. Figure 2c is a second-generation light-field camera, Illum. The German company, Raytrix, provides an alternative light-field camera design, where the microlenses have different focal lengths, as shown in Figure 2d.

In essence, the single-lens-based light-field camera samples the angle information through the microlens array. As shown in Figure 3a, each microlens covers a certain number of pixels. A cone of light from a subaperture of the main lens passes through different positions in each microlens imaging sensor, and this type of position change can be used to obtain directional information about the light. The main lens aperture is divided into numerous sampling units, and each microlens corresponds to the different position of the light field. By putting the entire microlens image together, you can obtain the 4D light-field records.

The light-field camera can refocus on any depth plane by rearranging the pixels on the sensor, because it records both position and angle information from the light field. Using Levoy's two-plane light-field parameterization,[6] you can render an image at a new focal depth according to the geometric relationships.[7] For each 3D point in the scene, you can calculate the coordinates of the points at which the light path intersects with the main lens and the sensor, and the final refocused image can be rendered by integrating the entire corresponding light ray from the light field. Because this virtual refocus plane can be at an arbitrary depth, you can achieve a specific depth of field within any plane by applying this refocusing function.

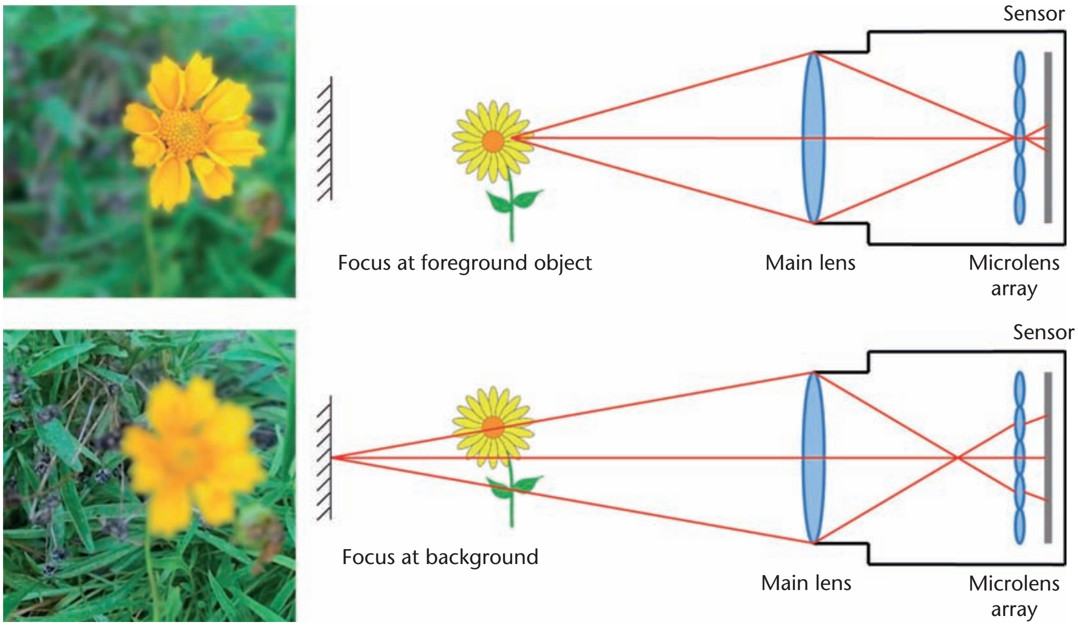

Figure 3. Post-capture refocusing on light-field cameras. A microlens array emulates the camera array to acquire a light field in a snapshot. Rays can then be combined based on hypothesized depth and aperture size to synthesize dynamic refocusing effects on the (a) foreground or (b) background object.

Using Light Fields for VR

To produce light-field VR content, recorded light-field data must be processed by both hardware and software. The reproduction of the light field includes two important technical parts: rendering and display. Because the light field contains important directional information, in any given focal plane and viewing direction, the desired scene can be rendered through a simple integration method. However, the entire set of light-field data is very high-dimensional; thus, the recorded light field usually suffers from insufficient sampling, resulting in image aliasing, especially when the geometric information is not correct. A ghost image will appear in the rendered image, as shown in Figure 4a.

This ghost problem seriously affects the quality of the rendered image, and the solution is to add auxiliary geometric information. Actually, only coupling with very simple geometrical information (such as depth) can significantly reduce the aliasing problem in the light-field rendering, as shown in Figure 4b.[8]

Obtaining this auxiliary geometry proxy turns out to be straightforward: you can directly estimate the depth through the light-field data via techniques such as light-field stereo matching.[9] First, because the light-field camera can provide a series of images of different views of the same scene, you can roughly estimate the depth of the scene using multiview geometry. Simultaneously, you can render a series of images focused at different depths by directly rearranging the pixels in the light-field camera, which is equivalent to focal-plane scanning of the scene. This virtual focal-plane scanning process can help us improve the depth-reconstruction process.

Secondly, in the optical imaging system, the intensity distribution near the focal plane is symmetrical. Specifically, the intensity of the light near the focal plane is symmetrically distributed. With this property, you can calculate the depth information of different points by detecting the intensity distribution for the images at different focal depths. This approach is much more robust to noise and undersampling than traditional methods.

In addition, occlusion is another important cue for depth estimation.[9] If there is no occlusion in the scene, the color distribution for the scene should be uniform; in contrast, at an occluded area, color mixing will appear. In a light-field camera, images rendered from different viewing angles have different statistical characteristics. If you apply occlusion detection based on such surface statistics and use it as an indicator of the consistency of the judgment, you can calculate the depth information. All

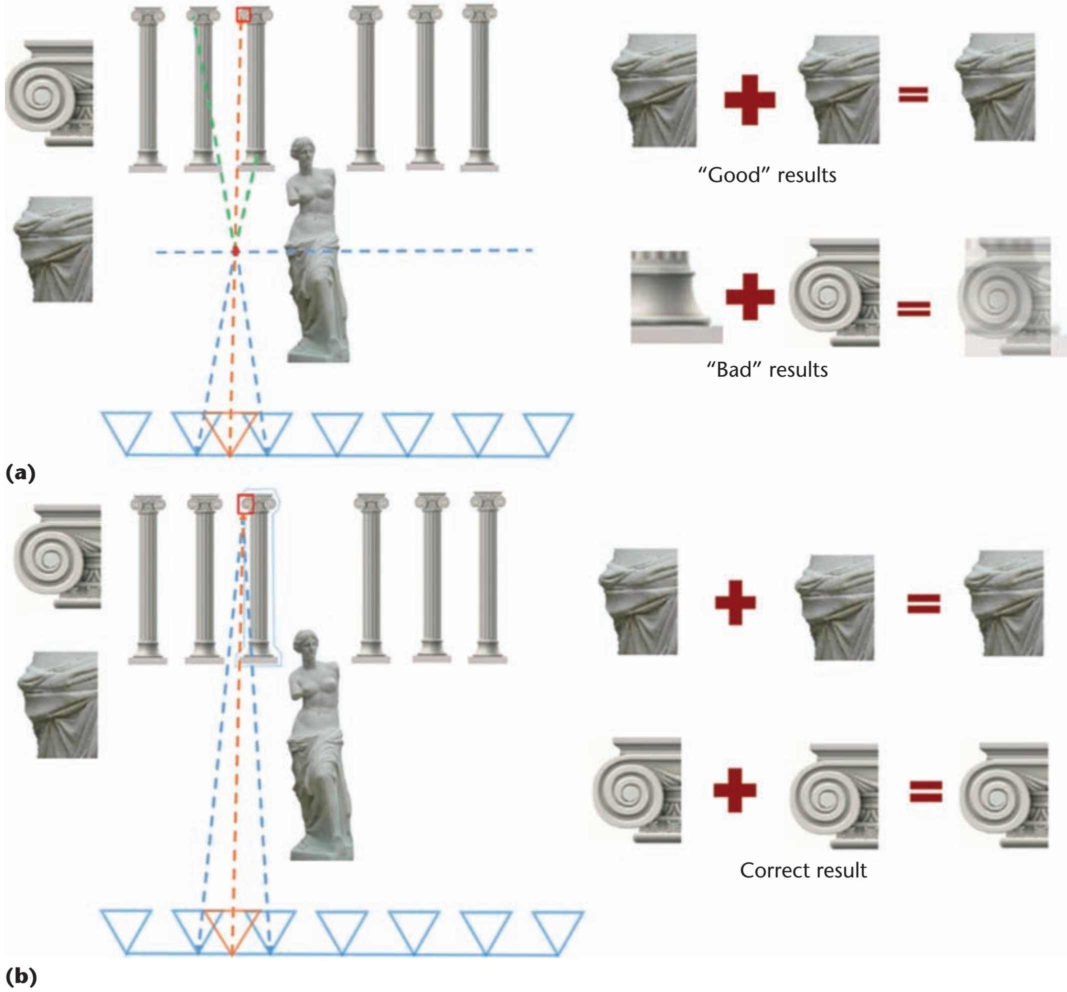

Figure 4. Ghosting effect in light-field rendering: (a) ghosting effects will appear in the rendered image when the geometrical information is incorrect; (b) the quality of the rendered image can be significantly improved by adding some simple geometry proxies, a key concept in lumigraph.[8]

the methods for reconstructing geometrical information in an image can help improve the geometrical information and render a comfortable image for the human eye. This rendered image displayed in a head-mounted display (HMD) can deliver all the light-field information; a human will naturally be able to watch VR content, leading to a strong sense of immersion.

Directly displaying the light field is an interesting direction in VR HMDs. Because of the reversibility of the optical path, the light-field display has a duality relation with the light-field imaging. The solution for light-field imaging can be directly applied to light-field display. An intuitive approach is to put a microlens in front of a monitor to display the light-field image, such that the scene can be viewed from the perspective of a different viewing angle or depth in different directions—this is the "naked-eye" 3D effect.

In recent years, light-field displays have undergone rapid development in both academia and industry. One type of light-field display uses a multilayer mask to control the outgoing light field. To design this mask, researchers adopted a method called *tensor decomposition,* which projects a 2D image into a 3D space to produce a high-resolution light-field image and motion parallax, occlusion, and translucency for naked-eye viewing.

Another type of light-field display device is the near-eye light-field display developed by NVIDIA (www.nvidia.com). This device is built with a microlens array and uses GPUs to render the light field. People can freely adjust the convergence and observe the correct near-eye accommodation, binocular parallax, and depth

(a) (b) (c) (d)

Figure 5. Panoramic imaging devices: (a) the Google Jump system (https://vr.google.com/jump); (b) Black Magic camera array (www.blackmagicdesign.com); and (c) Dual Panoramic Camera Array (DPCA) used for (d) live 360 3D VR live broadcasting for the play War Horse.

disparity. The main idea is that you cover the organic light-emitting diode (OLED) display with a microlens array, and the device synthesizes a virtual 3D light field, which is quickly adjusted according to the human eye distance. A downside of this approach is loss of resolution: the display is of a fixed resolution and when it needs to accommodate many views, the effective resolution of each view is significantly reduced.

An alternative design is to decompose the 4D light fields into layers of 2D images. Such a design, commonly referred to as the "layered" light-field displays, might be the key technology that VR giant Magic Leap (www.magicleap.com) is exploring. This design can break the curse of resolution loss, because each layer can maintain a very high resolution. However, it would be inevitably bulky due to the use of multiple display layers and interlayer wiring.

Light-Field VR Startups

Traditional entertainment media companies have been investing considerable resources in VR, developing experiences ranging from advertisements, news documentaries, and film tie-ins to live events such as sports and concerts. Yet compelling VR content is still severely lacking, and these recorded experiences have thus far been using monoscopic panoramas created using techniques dating from the original Quicktime VR format in 1995. Many people have developed homebrew camera systems using multiple cameras, ranging from GoPros (360Rize) at the low end to BlackMagic (www.blackmagicdesign.com) and Red (www.red.com) at the high end. There are an almost uncountable number of companies developing all-in-one multicamera systems to capture this content, starting with low-end systems that have two cameras, such as the Ricoh Theta (https://theta360.com) and Samsung Gear 360 (www.samsung.com/global/galaxy/gear-360), or six cameras, such as the GoPro Omni (https://vr.gopro.com). There are also systems with eight cameras, such as the Nokia Ozo (https://ozo.nokia.com), or, at the high end, 24 cameras, such as Jaunt One (www.jauntvr.com/jaunt-one). On the software side, Video Stitch (www.video-stitch.com) and Kolor (www.kolor.com) provide tools for stitching images.

There is a limitation with panoramic or 360 videos (see Figure 5). The first is that most of the current content is monoscopic so there's no sense of depth. Some people, such as HypeVR (https://hypevr.com) and NextVR (www.nextvr.com/#/), use stereo camera pairs to generate stereoscopic panoramas. HypeVR uses an additional laser scanner to capture depth. A few companies are trying to reconstruct stereo. Stereoscopic panoramas can be viewed as a special 3D light field: one can use a 1D array of cameras positioned on a circle or a dome facing outward to capture a subset of the light field. Google Jump (see Figure 5a) and Facebook use this type of configuration.

The next problem to solve is motion parallax or the ability to move around. Some companies are using traditional methods from games, such as static geometry capture and motion capture or authored animation. For a truly recorded experience, some companies (8i and Microsoft) use multiple cameras looking inward from a dome or room configuration to capture human actors. These actors are reconstructed in 3D and placed in virtual scenes. Generally, only the actors are captured and not the complete environment. To reconstruct the environment, Matterport (https://matterport.com) has developed a depth+color device that can scan and reproduce geometry of static scenes like rooms and buildings.

Light fields can provide huge improvements for immersive VR experience, transforming the entire VR market by solving both the depth and

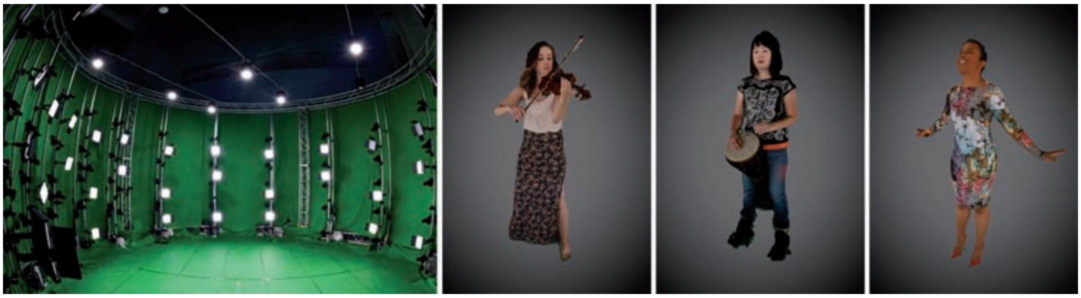

Figure 6. Acquiring inward-looking light fields: (a) the PLEX inward light-field capture dome, which comprises 140 cameras and can capture live human performance from 360 degrees. (b) Three performers captured, reconstructed, and re-rendered using this light-field solution.

motion parallax problems. There is considerable interest from the film community because of this potential. Lytro is currently developing new cameras aimed at VR capture, and Otoy has demonstrated light-field rendering in the cloud as well as other light-field prototypes. In fact, just recently, Lytro closed its venture round D of fundraising with $60M,[10] and 8i closed its round B of fundraising with $27M.[11]

Plex VR (www.plex-vr.com), a light-field startup that my colleagues and I founded, focuses on developing and commercializing a comprehensive array of light-field technologies for end-to-end VR content production. For example, we have employed new computational imaging solutions for efficient light-field acquisition. The Dual Panoramic Camera Array (DPCA) mounts stereoscopic camera pairs on a circle (see Figure 5c). Each camera pair is rectified to produce nearly perfect stereoscopic images. A special GPU-based solution can be employed to cover the gaps between camera pairs to perform real-time stitching and data streaming. In January 2016, Plex VR debuted the DPCA solution for acquiring the first stereoscopic panoramic VR video in China for the renowned play *War Horse* (Figure 5d). The camera system was positioned between the performing stage and the audience and online users were able to view the live stereoscopic contents on various types of VR headsets at home.

The DPCA can be viewed as an outward looking light-field capture device. To acquire inward-looking light fields, we have built a large camera dome composed of 140 cameras—80 static and 60 dynamic (see Figure 6). This light-field dome can capture light fields of human-size subjects at 60 frames per second. Special 3D reconstruction schemes and light-field rendering algorithms are applied to reproduce photorealistic humans in action.

In particular, we have partnered with the Julliard School for Performing Arts in the Lincoln Center in New York City on acquiring and displaying virtual performers in VR settings. The users can wear either an AR headset, such as Microsoft Hololens (www.microsoft.com/microsoft-hololens), or a VR headset, such as HTC Vive, to view the performers from any viewpoint. The work is part of the effort by the United Nations Educational, Scientific, and Cultural Organization to produce a world VR concert, which allows participants from any continent to join remotely through an immersive VR experience.

A smaller version of the dome further enables acquisition of both the geometry and surface reflectance of static objects. To that end, we have partnered with e-commerce giant Alibaba (www.alibaba.com) in China to provide immersive VR shopping experiences. Figure 7 demonstrates the power of light-field rendering in revealing complex reflectance of leather goods and scattering on Chinese Sancai pottery.

Light-field VR technology might lead to profound changes in the VR industry, and significant research efforts are being made on both hardware and software fronts. For example, due to the tradeoff between spatial resolution and angular resolution, there is a limitation of insufficient resolution for the light-field camera. Determining how to obtain a high-resolution result from a light-field camera from redundant light-field data using super-resolution techniques remains a problem. In recent years, new signal sampling and reconstruction theories based on compressive sensing have shown great

(a)

(b)

Figure 7. A light-field VR experience: (a) viewing a leather handbag and (b) a Sancai in an immersive VR experience. Light fields were used to acquire, reconstruct, and re-render the subjects.

success in recovering the original signal from a very small volume of samples.

The latest compressive light-field imaging techniques exploit this property to recover high-dimensional data.[12] The new solutions, however, require modifying existing optical systems as well as establishing new methodologies for capturing the compressed light field and rebuilding high-resolution light-field signals from this compressed data. More recently, from the perspective of data analysis, deep-learning technology further provides a viable path for light-field super-resolution. Applying the latest deep-learning technology to light-field imaging could be another feasible solution to improve quality while reducing the data size of light fields. Using these newest techniques to improve light-field technology will surely be the new driving force for the development of light-field VR. **MM**

References

1. J.C. Yang et al., "A Real-Time Distributed Light Field Camera," *Proc. 13th Eurographics Workshop on Rendering* (EGRW), 2002, pp. 77–86; http://dl.acm.org/citation.cfm?id=581907.
2. B. Wilburn et al., "High Performance Imaging Using Large Camera Arrays," *ACM Trans. Graphics*, vol. 24, no. 3, 2005, pp. 765–776.
3. *Computer Vision and Image Understanding*, special issue on light field for computer vision, April 2016; http://www.sciencedirect.com/science/journal/10773142/145.
4. A. Gershun, "The Light Field," *J. Mathematics and Physics*, vol. XVIII, MIT, 1939, pp. 51–151.
5. E.H. Adelson and J.R. Bergen, "The Plenoptic Function and the Elements of Early Vision," *Computation Models of Visual Processing*, M. Landy and J.A. Movshon, eds., MIT Press, 1991, pp. 3–20.
6. M. Levoy and P. Hanrahan, "Light Field Rendering," *Proc. 23rd Ann. Conf. Computer Graphics and Interactive Techniques* (SIGGRAPH), 1996, pp. 31–42; http://dl.acm.org/citation.cfm?id=237199.
7. R. Ng et al., *Light Field Photography with a Hand-Held Plenoptic Camera*, tech. report, Computer Science Technical Report (CSTR), 2005.
8. S.J. Gortler et al., "The Lumigraph," *Proc. 23rd Ann. Conf. Computer Graphics and Interactive*

Techniques (SIGGRAPH), 1996, pp. 43–54; http://dl.acm.org/citation.cfm?id=237200.
9. H. Lin et al., "Depth Recovery from Light Field Using Focal Stack Symmetry," *Proc. IEEE Int'l Conf. Computer Vision* (ICCV), 2015; http://ieeexplore.ieee.org/document/7410751.
10. C. Giardina, "Light Field Camera Developer Lytro Raises $60M," *Hollywood Reporter*, 15 Feb. 2017; www.hollywoodreporter.com/behind-screen/light-field-camera-developer-lytro-raises-60m-976475.
11. "8i Announces $27 Million Series B Funding to Bring Holograms to Consumers with New Mixed Reality App 'Holo,'" *Business Wire*, 13 Feb. 2017; www.businesswire.com/news/home/20170213006043/en/8i-Announces-27-Million-Series-Funding-Bring.
12. K. Marwah et al., "Compressive Light Field Photography Using Overcomplete Dictionaries and Optimized Projections," *J. ACM Trans. Graphics—SIGGRAPH 2013 Conf. Proc.*, vol. 32, no. 4, 2013, article no. 46; http://dl.acm.org/citation.cfm?doid=2461912.2461914.

Jingyi Yu is a professor at the University of Delaware and at ShanghaiTech University. He is currently on leave, working at Plex VR, a startup company that he co-founded with Jason Yang, Xu Hong, and Yi Ma. Contact him at jingyi.yu@plex-vr.com.