Review of light field technologies

Shuyao Zhou¹,², Tianqian Zhu¹, Kanle Shi¹, Yazi Li¹, Wen Zheng¹ and Junhai Yong³*

Abstract

Light fields are vector functions that map the geometry of light rays to the corresponding plenoptic attributes. They describe the holographic information of scenes by representing the amount of light flowing in every direction through every point in space. The physical concept of light fields was first proposed in 1936, and light fields are becoming increasingly important in the field of computer graphics, especially with the fast growth of computing capacity as well as network bandwidth. In this article, light field imaging is reviewed from the following aspects with an emphasis on the achievements of the past five years: (1) depth estimation, (2) content editing, (3) image quality, (4) scene reconstruction and view synthesis, and (5) industrial products because the technologies of light fields also intersect with industrial applications. State-of-the-art research has focused on light field acquisition, manipulation, and display. In addition, the research has extended from the laboratory to industry. According to these achievements and challenges, in the near future, the applications of light fields could offer more portability, accessibility, compatibility, and ability to visualize the world.

Keywords: Light field imaging, Holographics, Human-machine graphic interaction

Introduction

A light field is the totality of light rays or radiance in three-dimensional (3D) space through any position and in any direction, as defined by Gershun [1] in 1936. Formally,

\[ L : g \rightarrow c \]

where a light field \( L \) is defined by mapping the geometry of a light ray \( g \) to the attributes of the corresponding light \( c \). Here, \( c \) is a vector that describes the intensity of every component of the light such as red, green, and blue (RGB). Geometrically, \( g \) has various definitions in different light field models. A plenoptic function describes all visual information [2]. Gershun [1] defined a five-dimensional (5D) plenoptic function \( L(x, y, z, \theta, \phi) \in \mathbb{R}^5 \) for the light field because each ray can be parameterized by three coordinates \((x, y, z)\) and two angles \((\theta, \phi)\). Compared with the previous 5D representation, Levoy and Hanrahan [3] assumed in their four-dimensional (4D) representation \( L(u, v, s, t) \in \mathbb{R}^4 \) that the light field is composed of oriented lines in free space, successfully reducing the redundancy of the total dataset and simplifying the reconstruction of the plenoptic function.

Meanwhile, Levoy and Hanrahan introduced light fields to the computer graphics field. In addition, if one describes a light field captured by a camera moving on a sphere centered on the target object, then geometry \( g \) can be defined as \((\theta, \phi, s, t) \in \mathbb{R}^4\), where \((\theta, \phi) \in \mathbb{R}^2\) is a spherical surface and \((s, t) \in \mathbb{R}^2\) is a plane surface with light projecting to it. “Ray space” is a synonym of “light field” [4, 5] to describe rays in a 3D space. The light field is the same as the orthogonal ray space. In the field of free-viewpoint televisions [6, 7], the term “ray space” is often used to describe ray-based 3D information display systems.

In 2005, the light field made the transition from mainly pure research to large-scale industrial applications. For example, Ng et al. [8] developed the first handheld plenoptic camera. It was not until 2010 that...
light field technology was commercialized to capture a light field. With the development of commercial light field cameras [9], plenoptic cameras that make it possible to refocus provide many benefits, and they have been widely used in light field applications. Subsequently, the commercial potential of the light field has been greatly illustrated in image editing, holographically perceived light fields, augmented reality, and physical entities such as plenoptic cameras. The number of publications has increased geometrically as light fields have gained increasing attention from researchers (see Fig. 3 for the timeline of light field imaging).

Light field acquisition is the preliminary light field imaging process. Wu et al. [10] comprehensively highlighted methods and devices for light field acquisition in their survey, including (1) multisensor capture (using multiple cameras to capture a light field at one time, with most of them being camera arrays [11–17]).
(2) time-sequential capture (using one camera to capture a light field with multiple exposures, which is time consuming [18–23]), and (3) multiplexed imaging [encoding high-dimensional data into a simpler two-dimensional (2D) image, which is the most popular method [8, 24–37].

Herein, this paper aims to review light field imaging, revealing the current deficiencies and exploring the future possibilities (see Fig. 4 for an overview). Five aspects have been reviewed: current depth estimation methods (Depth estimation Section), light field editing techniques (Editing Section), light field enhancements with an emphasis on increasing the quality of images (Enhancement Section), 3D reconstruction and view synthesis (Reconstruction and view synthesis Section), and the light field industry, which is categorized into light field acquisition and light field displays (Section 6).

**Depth estimation**

Depth estimation involves inferring 3D information from 2D images, which is a foundation for light field editing and rendering. Light field data record the spatio-angular information of light rays; thus, a light field image contains many depth cues to make depth estimation possible. Conventionally, depth cues include, but are not limited to, correspondence cues, defocus cues, binocular disparity, aerial perspective, and motion parallax. Occlusion often occurs when two or more objects come too close and, therefore, hide some information from each other. Specifically, when people want to see an occluded object, they usually move slightly to avoid the occluder. This commonsensical solution explains why light fields have special benefits in solving the depth map with occlusion. Therefore, researchers have mainly focused on traditional approaches and convolutional neural network (CNN) approaches for depth estimation with examinations of occlusion handling: (1) constraint-based estimation, (2) epipolar plane image (EPI)-based estimation, and (3) CNN-based estimation.

Constraint-based estimation utilizes different constraints of the light field structure to estimate the depth. Bishop and Favaro [38] estimated the depth from multiple aliased views and demonstrated that this could be done at each pixel of a single light field image. Williem and Lee [39] utilized the correspondence cue and defocus cue, which were robust against both occlusion and noise, and they introduced two data costs: the constrained angular entropy cost and constrained adaptive defocus cost. Zhu et al. [40] addressed a multioccluder occlusion by regularizing the depth map with an antioclusion energy function. Some researchers have considered the relationship between occluders and natural light reflections. For example, Baradad et al. [41] estimated the 4D light field of a hidden scene from 2D shadows cast by a known occluder on a diffuse wall by determining how light, which naturally reflected off surfaces in the hidden scene, interacted with the occluder. Chen et al. [42] detected partially occluded boundary regions (POBRs) by using superpixel-based regularization. After a series of shrinkage and reinforcement operations on the labeled confidence map and edge strength weights over the POBR, they produced a depth estimate with a low average disparity error rate and high occlusion boundary precision-recall rate. To proceed with occlusion handling from image to video, Lueangwattana et al. [43] examined the structure from motion to improve light field rendering and, hence, addressed fence occlusion in videos while preserving background details.

However, the EPI, proposed by Bolles et al. [44] in 1987, simplifies depth measurement by restricting...
motion to straight lines and working with a series of closely spaced images, thereby reducing the 3D problem into a set of 2D problems. Some studies on EPI representations are highlighted here. Matoušek et al. [45] suggested a dynamic-programming-based algorithm find correspondences in EPIs by extracting lines with similar intensities in an EPI separately for each row. In other research, Criminisi et al. [46] worked with EPI volume, a dense horizontally rectified spatio-temporal volume that results from a linearly translating camera, for automated layer extraction. They relied on an EPI tube, which is a collection of EPI lines of the same depth.

Unlike the above works that refine EPI representations, the following works established how to apply EPIs in depth estimation. Wanner and Goldluecke [47] used the dominant directions of EPIs from the structure tensor method to estimate depth. However, this method is sensitive to noise and occlusion. In addition, estimation based on a 2D EPI is vulnerable to noise and sometimes fails because of very dark and bright image features.
Multiorientation EPIs are epipolar plane images in all available directions and provide rich light field angular information. To achieve a better depth map from light field images, Tao et al. [48] computed dense depth estimation by combining defocus and correspondence depth cues based on full 4D EPI. Moreover, defocus cues perform better in repeating textures and noise, and correspondence cues are robust in terms of bright points and features. Tao et al. [49] obtained defocus cues by computing the spatial variance after angular integration and correspondence depth cues by computing the angular variance. Furthermore, they performed depth estimation on glossy objects with both diffuse and specular reflections and one or more light sources by exploiting the full EPIs. Similarly, Zhang et al. [50] proposed a spinning parallelogram operator (SPO) to locate lines and calculate their orientations in an EPI for local depth estimation, which further handled occlusions and was more robust to noise. In addition, Sheng et al. [51] combined a multiorientation SPO with edge orientation to improve depth estimation around occlusion boundaries. They proved that the direction of the optimal EPI was parallel to the boundary of the occlusion. In contrast to the work of Sheng et al. [51], Schilling et al. [52] incorporated both depth and occlusion using an inline occlusion-handling scheme, OBER-cross+ANP, to improve object boundaries and smooth surface reconstruction.

Deep CNNs have been extensively applied to depth estimation because they have a better balance between accuracy and computational cost. In 2017, Heber et al. [53] extended the previous work [54] in which the network operated on EPIs, and they replaced all 2D operators with 3D counterparts. Then, they used a CNN that predicted disparity based on RGB EPI volumes, and the proposed network learned to recover depth information for shape estimation. For supervised training, researchers require large labeled datasets. Shin et al. [55] solved the data insufficiency problem by incorporating a multistream network, which encoded each EPI separately for depth estimation, into their CNN model EPINET. Tsai et al. [56] proposed an attention-based view selection network that exploited the priorities of light field images and the correlations between them to reduce redundancy and computation time. In 2021, Chen et al. [57] applied an attention-based multilevel fusion network. They grouped four directions (0°, 45°, 90°, and 135°) of light fields into four branches. Then, they combined the branches with two feature fusion methods to generate depth maps: intrabranch feature fusion based on channel attention and interbranch feature fusion based on branch attention. Researchers usually employ deep CNNs for accurate depth estimation, combining them with traditional approaches to produce better results.

Depth estimation has been a focus of much research. Researchers have worked on constraint-based methods and have explored different depth cues and their combinations. They also simplified the estimation by using EPIs and applying learning-based methods. There have been other studies that evaluate depth estimation methods. The work of Johannsen et al. [58] covers more depth estimation algorithms before and including 2017. The key to enhancing other light field-related applications, such as refocusing or rendering, is to develop more-precise and more robust depth estimation methods.

**Editing**

Because most light field datasets contain redundancy, researchers are interested in fully using redundancy and manipulating the light field images. Editing light fields is challenging [59] because (1) the light fields are 4D, whereas most tools on the market are for 2D, (2) local edits need to preserve the redundancy of the 4D light field, and (3) the depth information of the 4D light field is implicit. Light field image editing can be divided into (1) refocusing, (2) removing the occlusion, (3) segmenting the light fields to make the editing experience as smooth as editing a 2D image (e.g., removing the scene objects or changing their color), and (4) improving the user interface of the light field editing.

Because light field images contain not only textural information but also geometrical information, researchers can explore refocusing after capturing that cannot be accomplished with 2D images. In 2015, Dansereau et al. [60] demonstrated that a hyperfan-shaped passband can achieve refocusing over a wide range of depths, which they called “volumetric refocusing.” However, the approach only worked for a single volumetric region. To overcome this, Jayaweera et al. [61] proposed a simultaneous refocusing approach for multiple volumetric regions in light fields. They employed a 4D sparse finite-extent impulse response filter, which is a series of two 2D filters composed of multiple hyperfan-shaped passbands. Noncrucial parts of images produced by digital single-lens camera arrays often experience blurs (bokeh). Wang et al. [62] proposed a light field refocusing method to improve bokeh rendering and image quality. They first estimated the disparity map and rendered the bokeh on the center-view sub-image. The rendered bokeh image was then used as a regularization term to generate refocused images. Moreover, Yang et al. [63] proposed a refocusing framework that produced coordinates for interpolation, and they aligned the images onto the focal plane.

Occlusion removal is another typical task in light field editing. The nature of light field sub-aperture images (SAIs) provides complementary information so that
hidden scenes can be seen from other views. Yang et al. [64] partitioned an image into multiple visibility layers and propagated the visibility information through layers. The visibility layer is defined as all the occlusion-free rays in any camera, computed by energy minimization. For CNN methods, Wang et al. [65] suggested a deep encoder-decoder network for automatically extracting foreground occlusions by analyzing scene structures. The SAIs were first encoded with spatial and angular information and then decoded for center-view reconstruction. They applied a 5D tensor to explore spatial connections among SAIs. To prove this, Li et al. [66] proposed another CNN-based encoder-decoder network for automatically extracting superpixel segmentation on a 4D light field performed better in representing the proximity regions.

For existing user interfaces, users can employ tools to edit the light field, even though the depth map is imperfect. Horn and Chen [75] designed Lightshop to manipulate light fields by operating on view rays. Image warping is a process of distorting an image [76]. When a user defines how the view rays warp, the Lightshop renderer composites and renders multiple light fields by executing the user-defined ray-shading program. However, the system has limitations when compositing different light fields because of the fixed illumination of a light field. In 2014, Jarabo et al. [77] provided an overview of different light field editing interfaces, tools, and workflows from a user perspective. Some products, such as the single-lens 3D-camera with extended depth of field presented by Perwaß and Wietzke [78], can refocus or apply predefined filters to light field images. Building on previous work [75], Mihara et al. [59] were the first to use a graph-cut approach for 4D light field segmentation based on a learning-based multilabel segmentation scheme. The user needs to specify a target region so that the algorithm can identify the appropriate regions and evaluate whether each ray is included in the selected region. Moreover, they defined appropriate neighboring relationships to preserve redundancies.

Overall, light field editing methods remain largely unexplored compared with 2D image editing. Researchers have experimented further on light field images, such as adding mosaic or special effects or filters, editing body features, combining several images into a short movie, and changing background, which are promising.

Enhancement

Light field enhancement optimizes the quality of light field images. Modern light field research mainly focuses on deblurring and super-resolution (SR).

Motion deblurring has two approaches: blind motion deblurring and nonblind deblurring. Blind motion deblurring has been extensively studied using 2D images. In 2014, Chandramouli et al. [79] first investigated motion deblurring for light field images and assumed constant depth and uniform motion for the simplicity of the model. Jin et al. [80] explored bilayer blind deconvolution that removed the motion blur in each layer to recover 2D textures from a motion-blurred light field image. Moreover, Srinivasan et al. [81] introduced a light field deblurring algorithm by analyzing the motion-blurred light field in the primal and Fourier domains. Unlike in previous studies, they recovered a full 4D light field image. Lee et al. [82] addressed six-degree-of-freedom (6-DOF) blind deblurring by considering the 3D orientation change of the camera. Lumentut et al. [83] proposed a deblurring deep neural network with 16,000
times faster speed than in prior work [81] and deblurred a full-resolution light field in less than 2 s. Dansereau et al. [84], unlike in the above approaches, adopted a nonblind algorithm for 6-DOF motion blur, assuming that the ground truth camera motion was known.

Levin et al. [85] illustrated that there is a trade-off between spatial and angular resolutions. Therefore, many researchers have endeavored to improve the spatial resolution of the captured light field using an SR algorithm by exploiting additional information from the available data. For instance, in 2009, Bishop et al. [86] studied SR algorithms using a depth map. They characterized the point-spread function of a plenoptic camera under Gaussian optics assumptions for a depth-varying scene and formulated the reconstruction of the light field in a Bayesian framework. Therefore, they restored the images at a resolution higher than the number of microlenses. Zhou et al. [87] applied the ray-tracing method to analyze the subpixel shifts between the angular images extracted from the defocused light field data and the blur in the angular images, and they obtained an SR result with a magnification ratio of 8. In contrast to restoring high-resolution images from low-resolution images, Lumdsaine and Georgiev et al. [88] rendered high-resolution images by adopting positional and angular information in captured radiance data. They rendered images from a 542-megapixel light field to produce a 106-megapixel final image. Zheng et al. [89] presented a convolutional deep neural network using cross-scale warping to the reference-based SR, which involves applying an extra high-resolution image as a reference to help super-resolve a low-resolution image that shares a similar viewpoint. In 2019, Cheng et al. [90] categorized the existing SR methods into projection-based [91–94], optimization-based [95–101], and learning-based [102–107]. Moreover, Farrugia and Guillemot [108] reduced the light field angular dimension using low-rank approximation and then applied CNNs to achieve peak signal-to-noise ratio (PSNR) gains of 0.23 dB over the second-best-performing method. Zhang et al. [109] proposed a residual convolutional network for higher spatial resolution. They first improved the spatial resolution of the central view image because it contained more sub-pixel information. Then, they trained the network to improve the spatial resolution of the entire light field image. Wang et al. [110] suggested a spatial-angular interactive network. They began by extracting the spatial and angular features independently from the input light field images, and then the information was processed by many interaction groups to achieve spatial-angular interaction features. Finally, the interacted features were fused to achieve high-resolution SAIs. Similar to previous researchers [110], who used feature collection and distribution, Wang et al. [111] proposed a deformable convolution network where all side-view features were aligned with the center-view feature and then aligned with the original features. Consequently, angular information is encoded into views for SR performance. Ivan and Williem [112] investigated an end-to-end encoder-decoder style for a joint spatial and angular light field SR model from only a single image without relying on physical-based rendering or secondary networks so that end users can experience the advantages of light field imaging. Jin et al. [113] proposed another learning-based light field spatial SR framework that uses deep combinatorial geometry embedding and structural consistency regularization. Their method improved the average PSNR by more than 1.0 dB and preserved more-accurate parallax details at a lower computational cost.

Researchers have devoted great effort to light field image deblurring and resolution trade-offs. For individual users, the capturing process involves more randomness, which requires stabilization, and they expect to achieve a high-resolution display similar to what they can obtain from a normal camera.

Reconstruction and view synthesis

Geometric reconstruction involves reconstructing an object from its geometric information. Sparsity in the Fourier domain is an important property that makes 4D light field reconstruction possible from a small set of samples. Shi et al. [114] proposed a method for recovering non-Lambertian light fields from a small number of 1D viewpoint trajectories optimized for sparsity in the continuous Fourier domain. In general, 3D reconstruction includes shape estimation, and some research has extended the results to holography. In addition, view synthesis creates new views from a given set of views.

For shape estimation, Lanman et al. [115] proposed surround structured lighting to achieve full 360° reconstructions using a single camera position, but it could only scan relatively small volumes. Subsequently, multi-view reconstruction methods have become increasingly popular. Heber et al. [116] proposed a variational multi-view stereo method, in which they used a circular sampling scheme inspired by a technique called “active wavefront sampling” (AWS) [117], where the AWS module is an off-axis aperture that moves along a circular path around the optical axis. In 2016, Heber and Pock [54] trained a CNN to predict 3D scene points from the corresponding 2D hyperplane orientation in the light field domain by using horizontal and vertical EPs and dividing each EPI into patches. Similarly, Feng et al. [118] focused on 3D face reconstruction from 4D light field images using a CNN. They constructed 3D facial curves, rather than a complete face, to make up a 3D face at once by combining all the horizontal and vertical curves of a face to form horizontal and vertical depth
maps separately. However, unlike Heber and Pock [54], they exploited a complete EPI for depth prediction. Zhang et al. [119] applied a light field camera as a virtual 3D scanner to scan and reconstruct 3D objects, which enabled dense surfaces to be reconstructed in real time. They illustrated that, with five light fields, the reconstructed 3D models were satisfactory.

Another application of light field 3D reconstruction is the holographically perceived light field because the light field can provide continuous focus cues. Overbeck et al. [120] developed “welcome to light fields” to enhance the virtual reality experience by setting up a 16-GoPro rotating rig, processing multiview depth maps, adopting disk-based light field rendering to make seamless connections among pictures, and varying compression levels with the movement of the eyeballs. Many studies have demonstrated high-quality scene rendering [121–123]. In 2017, Shi et al. [124] introduced near-eye light field computer-generated rendering with spherical waves for wide-field-of-view interactive 3D computer graphics. Nonetheless, Mildenhall et al. [125] proposed a deep-learning method for view synthesis from an irregular grid of sampled views that first expands each sampled view into a local light field by means of a multiplane image scene representation and then blends adjacent local light fields. They used up to 4000 times fewer views and presented a plenoptic sampling framework by clearly specifying how users should sample input images for view synthesis with portable devices.

Recent view synthesis methods have approached new view generation from only a few input images. The neural radiance field (NeRF) [126] is a nonconvolutional deep network representation that characterizes the volume space with a multilayer perceptron. It takes a continuous function \((x, y, z, \theta, \varphi) \in \mathbb{R}^5\) as input, and it outputs the volume density and view-dependent RGB color. Overall, it reconstructs the surface geometry and appearance from a small set of images. Subsequent studies have extended the NeRF in a variety of ways. Park et al. [127] modeled shape deformations by augmenting the NeRF. They can reconstruct free-viewpoint selfies from photographs. Nonetheless, the NeRF in the wild [128] considers the photometric and environmental variations between images to reconstruct real-world scenes. Compared with NeRF, Mip-NeRF [129] takes in a 3D Gaussian that represents the region over which the radiance field should be integrated. As a result, it can show the same scene at multiple levels of sharp detail. For video synthesis, Pumarola et al. [130] and Li et al. [131] extended the NeRF to dynamic objects with an additional parameter, time \(t\).

The reconstruction and view synthesis processes involve constructing, rendering, and displaying. Obtaining 3D reconstruction through light fields provides researchers with a bright outlook for real-time augmented reality and virtual reality. In addition, light fields provide insights into various view synthesis methods.

**Industrial applications**

As mentioned earlier, Lytro and Raytrix have industrialized light field acquisition devices. Other companies, such as FoVi3D [132] and Japan Display [133], have also produced light field projection solutions. Unlike for the Lytro camera, in which a microlens array is placed in front of an image sensor, Wooptix [134] reduced the resolution trade-off by using a liquid lens [135] of the optical chain in front of the sensor to make it possible to change the focal planes quickly, providing full resolution of the sensor in real time. Furthermore, Google published many patents related to light field capturing [136–138]. It also published “Capturing Light Field Images with Uneven and/or Incomplete Angular Sampling” [139] in 2018. It designed a camera to capture light field images with uneven and incomplete angular sampling. The results showed an improvement in not only the spatial resolution but also the quality of the depth data.

For light field display, Avegant [140], Leia [141], Light Field Lab [142], Dimenco [143], and Creal [144] facilitated realistic digital photographs with display screens. Likewise, Looking Glass Factory [145] created a light field image display providing 45 different viewpoints as long as the viewer is within a 58° viewing cone, and it is far less expensive than previous products and, therefore, more attainable. In 2020, Sony published two light field-related technologies: 3D Spatial Reality Display Technology [146] and Atom View [147]. The former tracks the position of the eyes of the user, enabling the user to see real-world images or creations in 3D. In addition, it achieves a relatively high resolution and glasses-free 3D using real-time light field rendering technology. Atom View is an application in volumetric virtual production through point-cloud rendering, editing, and coloring. It can digitize space instead of a physical set and, therefore, can reproduce locations and sets. These light field display solutions are physical products, and light fields can also be applied in immersive online experiences. For instance, OpenFuture [148] provides Hololux Light Field Solution, which focuses on 3D reconstruction. It can reconstruct complex materials at full angle, and it is closely working with e-commerce companies to enhance the shopping experience. Furthermore, Google has produced its glasses-free light field display technology, Project Starline [149], which is used for real-time communication. People can communicate with each other as if they are sitting across from each other. However, Project Starline relies on custom-built hardware and highly specialized equipment.
The above products solve problems in two important parts of the light field imaging pipeline: acquisition and display. In the near future, one can expect more portable devices for capturing light fields to emerge. In addition, the use of light field displays may extend from fixed screens to display extremely small or extremely large pictures and can further benefit medical microscopy or cinematic displays. Finally, the light field can contribute to closer-to-truth communication, which should make the “smart life” more attainable.

Conclusions
Depth estimation, which is essential for light field applications, was introduced. Then, the trend of light field applications was evaluated in terms of editing, enhancement, reconstruction, and current industrial products. The light field has been a research focus in computer graphics since 1996 and has progressed into the commercial market since 2010. Starting in 2010, the number of publications has increased rapidly, showing that many researchers are exploring the potential applications of light fields. These studies emphasize the critical role of the light field in enhancing visual experience. However, they require considerable expertise in utilizing light field technology. Therefore, there are still many human-light field interaction challenges to address not only in holography and augmented reality, but also in free image editing and interactive 3D reconstruction. Overall, light field imaging is commercially practical for businesses and individual users.

Abbreviations
1D: One-dimensional; 2D: Two-dimensional; 3D: Three-dimensional; 4D: Four-dimensional; 5D: Five-dimensional; AWS: Active wavefront sampling; CNN: Convolutional neural network; e-DOF: Six-degree-of-freedom; EPI: Epipolar plane image; NeRF: Neural radiance field; POBR: Partially occluded boundary region; PSNR: Peak signal-to-noise ratio; RGB: Red, green, and blue; SA: Sub-aperture images; SPO: Spinning parallelogram operator; SR: Super-resolution

Acknowledgements
The authors are grateful to Kuaishou Technology, University of California, Berkeley, and Tsinghua University.

Authors’ contributions
All authors read and approved the final manuscript.

Funding
The last author was supported by the National Key R&D Program of China, No. 2019YFB1405703.

Availability of data and materials
All data analysed during this study are included in this published article.

Declarations
Ethics approval and consent to participate
All authors give ethics approval and consent to participate.

Consent for publication
All authors give consent for publication.

Competing interests
The authors declare that they have no competing interests.

Author details
1Y-tech, Kuaishou Technology, Beijing 100085, China. 2EECS Department, University of California, Berkeley, CA 94720, USA. 3School of Software, BNRist, Tsinghua University, Beijing, China.

Received: 6 July 2021 Accepted: 29 October 2021

Published online: 03 December 2021

References
1. Gershun A (1939) The light field. J Math Phys 18(1–4):51–151. https://doi.org/10.1002/sapm1939151
2. Adelson EH, Bergen JR (1991) The plenoptic function and the elements of early vision. In: Landy M, Movshon JA (eds) Computational models of visual processing. MIT Press, Cambridge, pp 3–20
3. Levoy M, Hanrahan P (1996) Light field rendering. In: Abstracts of the 23rd annual conference on computer graphics and interactive techniques. Association for Computing Machinery, New York. https://doi.org/10.1145/237170.2371999
4. Fuji T (1994) A basic study on the integrated 3-D visual communication. Dissertation, The University of Tokyo
5. Fuji T, Kimoto T, Tanimoto M (1996) Ray space coding for 3D visual communication. Paper presented at the international picture coding symposium, Electronic Imaging, Melbourne, 13-15 March 1996
6. Tanimoto M (2005) Overview of free viewpoint television. Signal Process Image Commun 21(6):454–461. https://doi.org/10.1016/j.image.2005.03.009
7. Kubota A, Smolic A, Magnor M, Tanimoto M, Chen T, Zhang C (2007) Multiview imaging and 3DTV. IEEE Signal Process Mag 24(6):10–21. https://doi.org/10.1109/MSP.2007.905873
8. Ng R, Levoy M, Bredif M, Duval G, Horowitz M, Hanrahan P (2005) Light field photography with a hand-held plenoptic camera. Stanford University Computer Science Tech Report
9. Raytrix. 3D light-field camera technology. https://raytrix.de/. Accessed 31 May 2021
10. Wu GC, Masia B, Jarabo A, Zhang YC, Wang LY, Dai QH et al (2017) Light field image processing: an overview. IEEE J Sel Top Signal Process 11(7):926–954. https://doi.org/10.1109/JSTSP.2017.2747126
11. Wilburn BS, Smulski M, Lee H-HK, Horowitz MA (2002) Light field video camera. In: Media Processors 2002, vol 4674, pp 29–36. https://doi.org/10.1117/12.451074
12. Yang JC, Everett M, Buehler C, McMillan L (2002) A real-time distributed light field camera. In: Abstracts of the 13th eurographics workshop on rendering, Eurographics Association, Pisa
13. Zhang C, Chen T (2004) A self-reconfigurable camera array. In: Abstracts of the ACM SIGGRAPH 2004 sketches. Association for Computing Machinery, New York. https://doi.org/10.1145/1186233.1186412
14. Chan SC, Ng KT, Gan ZF, Chan KL, Shum HY (2005) The plenoptic video. IEEE Trans Circuits Syst Video Technol 15(12):1650–1659. https://doi.org/10.1109/TCSVT.2005.8988616
15. Liu YB, Dai QH, Xu WL (2006) A real time interactive dynamic light field transmission system. In: Abstracts of the 2006 IEEE international conference on multimedia and expo. IEEE, Toronto. https://doi.org/10.1109/ICME.2006.262686
16. Venkataraman K, Lelescu D, Duparré J, McMahon A, Molina G, Chatterjee P et al (2013) PiCam: an ultra-thin high performance monolithic camera. ACM Trans Graph 32(6):166. https://doi.org/10.1145/2508363.2508900
17. Lin X, Wu J, Zheng G, Dai Q (2015) Camera array based light field microscopy. Biomed Opt Express 6(9):3179–3189. https://doi.org/10.1364/BOE.6.003179
18. Unger J, Wenger A, Hawkins T, Gardner A, Debevec P (2003) Capturing and rendering with incident light fields. In: Abstracts of the 14th eurographics workshop on rendering, Eurographics Association, Leuven
19. Ihrke I, Stich T, Gottschlich M, Magnor M, Seidel HP (2008) Fast incident light field acquisition and rendering. J WSCG 16(1–3):25–32.
20. Liang CX, Lin TH, Wong BY, Liu C, Chen HH (2008) Programmable aperture photography: multiplexed light field acquisition. ACM Trans Graph 27(8):1–10. https://doi.org/10.1145/1399904.1360654
21. Taguchi Y, Agrawal A, Ramalingam S, Veeraraghavan A (2010) Axial light field for curved mirrors: reflect your perspective, widen your view. In: Abstracts of the 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, San Francisco. https://doi.org/10.1109/CVPR.2010.5540172
22.  Kim C, Zimmer H, Pritch Y, Sorokin-Hornung A, Gross M (2013) Scene reconstruction from high spatio-angular resolution light fields. ACM Trans Graph 32(4):73. https://doi.org/10.1145/2461912.2461926

23.  Dansereau DG, Schuster G, Ford J, Wetzstein G (2017) A wide-field-of-view monoenocentric light field camera. In: Abstracts of the IEEE conference on computer vision and pattern recognition. IEEE, Honolulu. https://doi.org/10.1109/CVPR.2017.400

24.  Adelson EH, Wang JYA (1992) Single lens stereo with a plenoptic camera. IEEE Trans Pattern Anal Mach Intell 14(2):99–106. https://doi.org/10.1109/34.121783

25.  Georgiev T, Zheng XC, Curless B, Salesin D, Nayar S, Intwala C (2006) Spatio-angular resolution tradeoffs in integral photography. In: Abstracts of the 17th europgrafics conference on rendering techniques. Eurographics Association, Aire-la-Ville

26.  Levy M, Ng R, Adams A, Footer M, Horowitz M (2006) Light field microscopy. ACM Trans Graph 25(3):924–934. https://doi.org/10.1145/1141991.1141976

27.  Veeraraghavan A, Raskar R, Agrawal A, Mohan A, Tumblin J (2007) Dappled photography: mark enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans Graph 26(3):60–63. https://doi.org/10.1145/1276377.1276463

28.  Lannman D, Raskar R, Agrawal A, Taubin G (2008) Shield fields: modeling and capturing 3D occluders. ACM Trans Graph 27(5):131. https://doi.org/10.1145/1457515.1409384

29.  Horstmeyer R, Euliss G, Athale R, Levoy M (2009) Flexible multimodal light field imaging for three-dimensional photography. ACM Trans Graph 32:46. https://doi.org/10.1145/157515.1409384

30.  Lanman D, Raskar R, Agrawal A, Taubin G (2008) Shield fields: modeling and capturing 3D occluders. ACM Trans Graph 27(5):131. https://doi.org/10.1145/1457515.1409384

31.  Horstmeyer R, Euliss G, Athale R, Levyo M (2009) Flexible multimodal light field imaging for three-dimensional photography. ACM Trans Graph 32:46. https://doi.org/10.1145/157515.1409384

32.  Manakov A, Restrepo JF, Klehm O, Hegedus R, Eisemann E, Seidel HP et al (2013) A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. ACM Trans Graph 32(4):47. https://doi.org/10.1145/2461912.2461937

33.  Marwah K, Wetzstein G, Bando Y, Raskar R (2013) Compressive light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graph 32(4):46. https://doi.org/10.1145/2461912.2461914

34.  Cohen N, Yang S, Andalman A, Broxton M, Grosenick L, Deisseroth K et al (2014) Enhancing the performance of the light field microscope using waveform coding. Opt Express 22(20):24817–24839. https://doi.org/10.1364/ OE.22.024817

35.  Wang YP, Wang LC, Kong DH, Yin BC (2015) High-resolution light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graph 32:46. https://doi.org/10.1145/2461912.2461914

36.  Cohen N, Yang S, Andalman A, Broxton M, Grosenick L, Deisseroth K et al (2014) Enhancing the performance of the light field microscope using waveform coding. Opt Express 22:24817–24839. https://doi.org/10.1364/OE.22.024817

37.  Wang YP, Wang LC, Kong DH, Yin BC (2015) High-resolution light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graph 32:46. https://doi.org/10.1145/2461912.2461914

38.  Heber S, Pock T (2015) Convolutional networks for shape from light field. In: Abstracts of the IEEE conference on computer vision and pattern recognition. IEEE, Sydney. https://doi.org/10.1109/CVPR.2015.800

39.  Taiy J, Li Y, Ouyoung M, Chuang YY (2020) Attention-based view synthesis for light field via dynamic group attention. In: Abstracts of the CVPR conference on computer vision and pattern recognition. IEEE, Vancouver. https://doi.org/10.1109/CVPR.2020.00499

40.  Chen J, Hou JH, Ni Y, Chau LP (2018) Accurate light field depth estimation with superpixel regularization over partially occluded regions. IEEE Trans Image Process 27(10):4889–4900. https://doi.org/10.1109/TIP.2018.2839524

41.  Baradad M, Ye V, Yedidia AB, Durand F, Freeman WT, Wornell GW et al (2018) Inferring light field images from shadows. In: Abstracts of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, Salt Lake City. https://doi.org/10.1109/CVPR.2018.00666
61. Jayaweera SS, Edusooonya CU, Wijenayake C, Agathoklis P, Bruton LT (2020) Multi-volumetric refocusing of light fields. IEEE Signal Process Lett 28:31–35. https://doi.org/10.1109/LSP.2020.3043990
62. Wang YQ, Yang JG, Guo YL, Xiao C, An W (2019) Selective light field refocusing for camera arrays using bokeh rendering and superresolution. IEEE Signal Process Lett 26(1):204–208. https://doi.org/10.1109/LSP.2018.285213
63. Yang JG, Xiao C, Wang YQ, An CJ, An W (2020) High-precision refocusing method with one interpolation for camera array images. IET Image Process 14(5):3899–3908. https://doi.org/10.1049/iet-ippr.2019.0081
64. Yang T, Zhang YN, Yu JY, Li JJ, Ma WG, Tong XM et al (2014) All-in-focus synthetic aperture imaging. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) ECCV 2014: computer vision - ECCV 2014, 13th European conference on computer vision, Zurich, Switzerland, September 2014. Lecture notes in computer science, vol 8694. Springer, Cham, pp 1–15. https://doi.org/10.1007/978-3-319-10599-4_1
65. Wang YQ, Wu TH, Yang JG, Wang LG, An W, Guo YL (2020) DeOccNet: learning to see through foreground occlusions in light fields. In: Abstracts of the IEEE winter conference on applications of computer vision. IEEE, Snowmass. https://doi.org/10.1109/WACV54572.2020.9093448
66. Li YJ, Yang W, Xu ZB, Chen Z, Shi ZB, Zhang Y et al (2021) Mask4D: 4D convolution network for light field occlusion removal. In: Abstracts of the IEEE international conference on acoustics, speech and signal processing. IEEE, Toronto. https://doi.org/10.1109/CASSP39728.2021.9433449
67. Berent J, Dragotti PL (2007) Unsupervised extraction of coherent regions for image based rendering. In: Abstracts of the British machine vision conference. BMVA Press, Coventry. https://doi.org/10.5244/C.21.28
68. Sethan JA (1996) Theory, algorithms, and applications of level set methods for propagating interfaces. Acta Numer 5:309–395. https://doi.org/10.1017/S0962492900002671
69. Wanner S, Straeler C, Goldscheider B (2013) Globally consistent multi-label assignment on the ray space of 4D light fields. In: Abstracts of the IEEE conference on computer vision and pattern recognition. IEEE, Portland. https://doi.org/10.1109/CVPR.2013.135
70. Batra D, Kowdle A, Parikh D, Luo JB, Chen T (2010) iCoseg: interactive co-segmentation with intelligent scribble guidance. In: Abstracts of the 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, San Francisco. https://doi.org/10.1109/CVPR.2010.5544080
71. Xu YC, Nagahara H, Shimada A, Taniguchi RI (2015) TransCut: transparent object segmentation from a light-field image. In: Abstracts of the IEEE international conference on computer vision. IEEE, Santiago. https://doi.org/10.1109/ICCV.2015.339
72. Ren XF, Valk J (2003) Learning a classification model for segmentation. In: Abstracts of the 9th IEEE international conference on computer vision. IEEE, Nice. https://doi.org/10.1109/ICCV.2003.12311_6
73. Cheng Z, Xiong ZW, Chen C, Liu D (2019) Light-field super-resolution: a benchmark. In: Abstracts of the IEEE/CVF conference on computer vision and pattern recognition workshops. IEEE, Long Beach. https://doi.org/10.1109/CVPRW.2019.00231
74. Georgiev T, Chuev G, Lumsdale A (2011) Superresolution with the focused plenoptic camera. In: Abstracts of the computational imaging IX. SPIE, San Francisco Airport. https://doi.org/10.1117/12.872666
75. Liang CK, Ramamoorthi R (2015) A light transport framework for lenslet light field cameras. ACM Trans Graph 34(2):16–19. https://doi.org/10.1145/2666075
76. Lim J, Ok H, Park B, Kang J, Lee S (2009) Improving the spatial resolution based on 4D light field data. In: Abstracts of the 16th IEEE international conference on image processing. IEEE, Cairo.
77. Navaux FP, Luke JP (2009) Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In: Abstracts of the 2009 3DTV conference: the true vision-capture, transmission and display of 3D video. IEEE, Potsdam.
78. Alain M, Smolic A (2018) Light field super-resolution via LBFSMD sparse coding. In: Abstracts of the 25th IEEE international conference on image processing. IEEE, Athens. https://doi.org/10.1109/CIP.2018.851162
79. Bishop TE, Favaro P (2012) The light field camera: extended depth of field, aliasing, and superresolution. IEEE Trans Pattern Anal Mach Intell 34(5):972–986. https://doi.org/10.1109/TPAMI.2011.168
80. Farag S, Veličković V (2018) A novel disparity-assisted block matching-based approach for super-resolution of light field images. In: Abstracts of the 2018-3DTV-conference: the true vision-capture, transmission and display of 3D video. IEEE, Helsinki. https://doi.org/10.1109/3DTV.2018.8478627
81. Mitr K, Veeraraghavan A (2012) Light field denoising, light field superresolution and stereo camera based refocusing using a GMM light field patch prior. In: Abstracts of the 2012 IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, Providence. https://doi.org/10.1109/CVPRW.2012.6239346
82. Ross M, El Gheche M, Frossard P (2018) A nonsmooth graph-based approach to light field super-resolution. In: Abstracts of the 25th IEEE international conference on image processing. IEEE, Athens. https://doi.org/10.1109/CIP.2018.8451127
83. Ross M, Frossard P (2017) Graph-based light field super-resolution. In: Abstracts of the 19th international workshop on multimedia signal processing. IEEE, Luton. https://doi.org/10.1109/MMSP.2017.8122224
Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.