Light Field Image Processing: An Overview

Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang,
Qionghai Dai, *Senior Member, IEEE,* Tianyou Chai, *Fellow, IEEE,* and Yebin Liu, *Member, IEEE*

Abstract—Light field imaging has emerged as a technology allowing to capture richer visual information from our world. As opposed to traditional photography, which captures a 2D projection of the light in the scene integrating the angular domain, light fields collect radiance from rays in all directions, demultiplexing the angular information lost in conventional photography. On the one hand, this higher-dimensional representation of visual data offers powerful capabilities for scene understanding, and substantially improves the performance of traditional computer vision problems such as depth sensing, post-capture refocusing, segmentation, video stabilization, material classification, etc. On the other hand, the high-dimensionality of light fields also brings up new challenges in terms of data capture, data compression, content editing and display. Taking these two elements together, research in light field image processing has become increasingly popular in the computer vision, computer graphics and signal processing communities. In this article, we present a comprehensive overview and discussion of research in this field over the past 20 years. We focus on all aspects of light field image processing, including basic light field representation and theory, acquisition, super-resolution, depth estimation, compression, editing, processing algorithms for light field display, and computer vision applications of light field data.

Index Terms—Light field imaging, light field processing.

✦

1 INTRODUCTION

A s the most important medium for people to perceive the world, light rays carry abundant information of our 3D environment. Unlike conventional images, which record the 2D projection of the light rays by angularly integrating the rays at each pixel, a light field describes the distribution of light rays in free space. This allows to capture richer information from our world. An early model that describes the distribution of light rays was first defined by Gershun [1] in 1936 and further completed by Adelson and Bergen [2] in 1991. This model is known as the plenoptic function, and describes light in a scene as a function of position, angle, wavelength and time.

However, obtaining the full plenoptic function for a scene is challenging, due to the high dimensionality of the data. Alternatively, modern light fields [3] mainly focus on the distribution of light rays as a function of position and angle. With this lower dimensionality and the progress in both hardware and software, today's light field capture devices are portable and commercially available [4], [5], and can even be integrated in cell phones [6]. Moreover, they have been integrated in microscopy [7], allowing rapid scan-less volumetric photographs of biological specimens. The widespread of light field cameras has allowed several new applications, ranging from their initial purpose (photorealistic image-based rendering [3]), to current computer vision applications that make use of their rich encoded information; these include 3D reconstruction, segmentation and

matting, saliency detection, object detection and recognition, tracking, and video stabilization.

Despite this large advance in the last 30 years, many scientific challenges remain for researchers to address. From an acquisition point of view, light field imaging requires the ability to capture higher dimensional data as opposed to simply recording a 2D projection in conventional photography (i.e., a light field camera should be designed to capture the directional light distribution at each location on the sensor). However, acquiring high-dimensional data typically imposes a resolution trade-off between the dimensions. Moreover, processing and analysis techniques for light fields are affected by their high dimensionality, which increases the computational complexity and imposes more challenging conditions on the design of algorithms. For example, segmentation in conventional 2D images aims to separate foreground and background within a single image, whereas when performing segmentation in light fields the photo-consistency must be kept.

This paper aims to provide an overview on the field of light field imaging and processing, while simultaneously revealing the challenges within. We first provide the theoretical description of light fields (Sec. 2). Then, we dive into different light field processing areas, which can be roughly divided in three modules (see Fig. 1): *low-level hardware-based acquisition*, which focuses on the trade-offs needed to effectively capture a light field (Sec. 3); *mid-level processing*, including different techniques proposed to mitigate the resolution trade-off (Sec. 4), depth estimation (Sec. 5), and light field evaluation approaches and compression schemes (Sec. 6); and *high-level user interface or application-related algorithms*, including light field applications in vision and graphics (Sec. 7), editing techniques (Sec. 8) and algorithms for display (Sec. 9). We additionally provide an overview of different databases and code publicly available for research (Sec. 10).

The classification of light field research modules is not

- G. Wu is with both Department of Automation at Tsinghua University and State Key Laboratory of Synthetical Automation for Process Industries at Northeastern University.
- B. Masia and A. Jarabo are with Graphics and Imaging Lab at Universidad de Zaragoza.
- L. Wang and T. Chai are with State Key Laboratory of Synthetical Automation for Process Industries at Northeastern University.
- Y. Zhang is with Shanghai Jiao Tong University.
- Q. Dai and Y. Liu are with Department of Automation at Tsinghua University. E-mail: liuyebin@mail.tsinghua.edu.cn
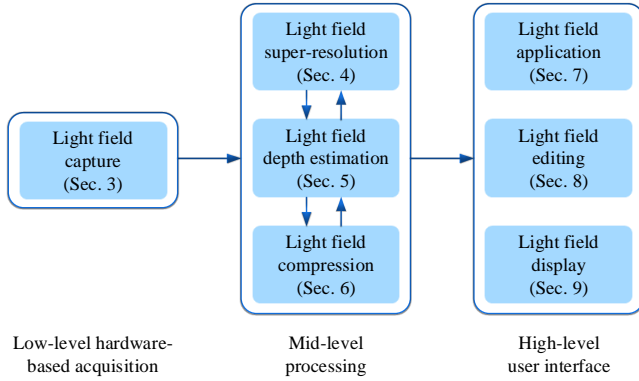
Fig. 1. Organizational framework of the paper.

unique, and we adopt the classification above based on the following considerations. First, depth estimation is classified as a mid-level processing technique because depth information is the most important intermediate result for the deploy of editing and display algorithms,as well as computer vision applications. Second, light field super-resolution (or reconstruction) takes place after light field acquisition, and is often iteratively optimized with depth estimation; therefore, we classified it as a mid-level processing technique as well. Third, light field compression techniques are well related with light field super-resolution (reconstruction) approaches, and are thus also classified as middle-level processing techniques. Last, light field editing and display are two visualization problems, which belong to applications of computer graphics, while the applications we described in the paper are mainly within the field of computer vision.

Partially overlapping works can be found at an early introductory-level overview [8], at a more conceptual and theoretical approach [9], and in a broad survey on computational plenoptic imaging techniques [10]. The difference in this overview is that we focus on the 4D light field as a scene representation structure, and provide comprehensive overviews and discussions of all the possible processing techniques related with it, i.e., from the low-level hardware-based acquisition, through mid-level data processing, and all the way to the so-called high-level techniques related to visualization or editing.

2 LIGHT FIELD THEORY

In this section, we present a brief overview of light field theoretical principles. This includes the representation of a light field as a multi-dimensional function, the basics of light field visualization and rendering, and an analysis of light field sampling in the Fourier domain.

2.1 Light Field Representation

The plenoptic function [2] is a multidimensional function describing the set of light rays traveling in every direction through every point in 3D space, from a geometric optics perspective. To obtain such a function, one must measure the light rays at every possible location (x, y, z), from every possible angle (θ, ϕ), at every wavelength γ and at every time t. The plenoptic function is then a 7D function denoted as $L(x, y, z, \theta, \phi, \gamma, t)$.

However, such high dimensional data is difficult to record and handle in practice. Thus, the light field model has been simplified twice for practical usage. In the first simplification, the measured function is assumed to be monochromatic and time-invariant. The wavelength γ of each light ray is recorded independently in different color channels. And the time sequence t can be recorded in different frames for a dynamic light field (i.e., a light field video). In this way, we can remove the wavelength γ and time t dimensions from the plenoptic function, reducing the model from seven to five dimensions.

The second simplification was made by Levoy and Hanrahan [3] and Gortler et al. [11], who realized that the 5D representation still contained some redundancy and could be reduced to 4D by assuming that the light field was measured in free space. In such cases, light ray radiance remains constant along a straight line, making one dimension redundant in the 5D plenoptic function. The extra dimension increases the complexity of measuring and reconstructing the plenoptic function.

When parameterizing a 4D light field, there are three key issues [3]: computational efficiency, control over the set of rays, and uniform sampling of the light field space. Based on these issues, the most common solution to the representation of a 4D light field is to parameterize the light rays by the coordinates of their intersections with two planes placed at arbitrary positions. The coordinate system is denoted by (u, v) for the first plane and (s, t) for the second plane. An oriented light ray defined in the system first intersects the uv plane at coordinate (u, v) and then intersects the st plane at coordinate (s, t), and is thus denoted by $L(u, v, s, t)$ (the schematic diagram will be presented in Sec. 2.3). Thus, the plenoptic function describing a light field is reduced from 7 to 4 dimensions, and parameterized by four coordinates (u, v, s, t).

2.2 Light Field Visualization

Although the function $L(u, v, s, t)$ is a simplified light field model, it is still hard to imagine this 4D representation. In this subsection, we will visualize this 4D light field in two different ways: integral light field structure, and 2D slices.

In the two-plane light field model $L(u, v, s, t)$, we can consider the st plane as a set of cameras with their focal plane on the uv plane. Two different perspectives can be taken to understand this model. First, that each camera collects the light rays leaving the uv plane and arriving at a point on the st plane (the collection of light rays from a certain viewpoint). Thus the 4D light field can be represented as a 2D array of images, such as the one shown in Fig. 2(a). Each image recorded by the camera is called a sub-aperture image, also known as a pinhole view. Second, that a certain point on the uv plane represents the light rays bound for all points on the st plane (the collection of light rays from different viewpoints projected onto a certain point, i.e., the same point as seen from different viewpoints, see Fig. 2(b)). Because the number of samples in the st plane depends on the number of viewpoints, while that in the uv plane depends on the camera resolution, the s and t dimensions are referred to as the angular dimensions and u and v dimensions are referred to as the spatial dimensions.
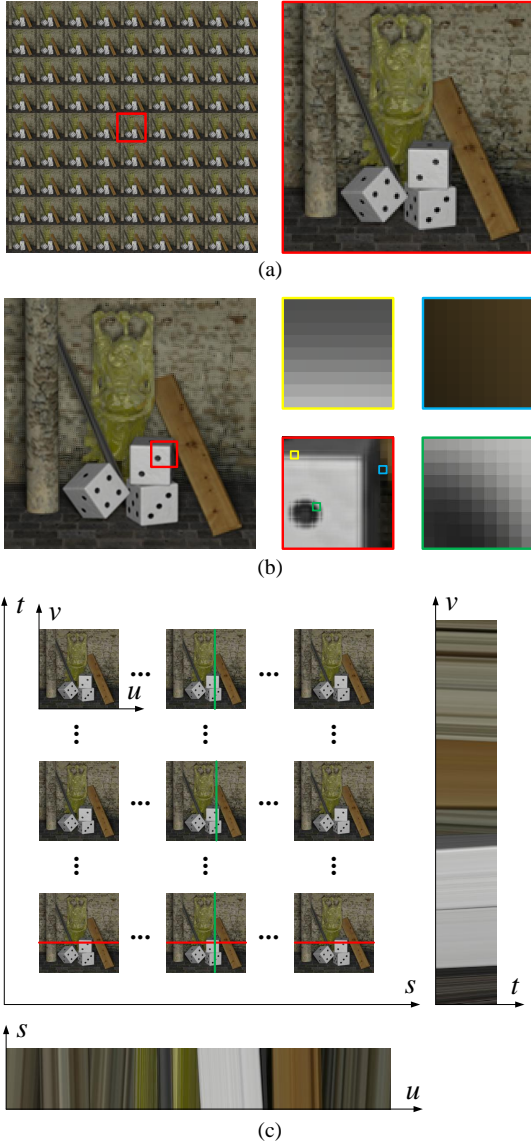
(a)

(b)

(c)

Fig. 2. Different ways of visualizing a light field. (a) A sub-aperture image $I_{s^*,t^*}(u,v)$ can be acquired by gathering the light field samples with fixed st coordinates (s^* and t^*). (b) A light field subview $I_{u^*,v^*}(s,t)$ can be acquired by gathering the samples with fixed uv coordinates (u^* and v^*). (c) The epipolar-plane image (EPI) is obtained by fixing the coordinates in both a spatial and an angular dimension. For example, the bottom EPI $E_{v^*,t^*}(u,s)$ is obtained by fixing v and t, and the right EPI $E_{u^*,s^*}(v,t)$ is obtained by fixing u and s.

Specifically, for a 4D light field $L(u,v,s,t)$, the 2D slice $I_{s^*,t^*}(u,v)$ can be acquired by gathering the samples at the fixed st coordinates s^* and t^*. The slice $I_{s^*,t^*}(u,v)$ (sub-aperture image) can be considered as a photograph captured by a camera located at (s^*, t^*). The right part in Fig. 2(a) visualizes a sample sub-aperture image. Similarly, the slice $I_{u^*,v^*}(s,t)$ is acquired by gathering the samples at the fixed uv coordinates u^* and v^*. The slice $I_{u^*,v^*}(s,t)$ (often referred to as the light field subview) is formed by gathering, at each point, the rays from different viewpoints, as visualized in Fig. 2(b) for three points (marked in yellow, green and blue).

The 2D slices described above are acquired by gathering either two spatial dimensions or two angular dimensions. By gathering the light field samples with a fixed spatial
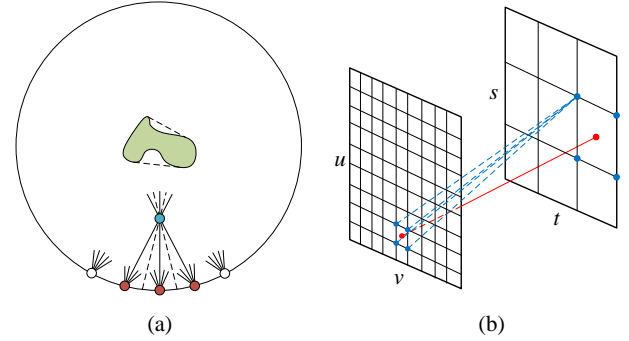

(a) (b)

Fig. 3. Rendering a virtual view using the recorded light field. (a) For a virtual view located at the blue dot that is closer to the object but outside its convex hull (the dashed line around the center object), some rays can be resampled in the existing views (e.g., the rays in the solid lines can be resampled by the rays captured by the cameras that are located at the red dots. However, some rays are captured by none of the cameras, such as the rays in the dashed lines. (b) For a virtual ray, the nearest 16 sampled rays are used to perform an interpolation (only four are depicted in the figure for illustration purposes).

coordinate v and an angular coordinate t (or u and s), one can produce a slice $E_{v^*,t^*}(u,s)$ (or $E_{u^*,s^*}(v,t)$). The resulting map is called the epipolar-plane image (EPI), which is a well-known term from multi-view computer vision [12]. Unlike a pinhole image or a light field subview, the EPI contains information in both spatial and angular dimensions. Consider a certain point on the Lambertian surface of an observed object with depth Z; when changing the coordinate s, the coordinate u will also change according to $\Delta u = \frac{f}{Z}\Delta s$ (where f is the distance between the planes), forming a line on the EPI. Points with different depths can be visualized as lines with different slopes in the EPI, see Fig. 2(c). Conversely, the slopes of lines in the EPI reflect the depth of the scene captured by the light field. This particular structure is widely exploited to infer scene geometry, as will be introduced in Sec. 5.

2.3 Light Field Rendering

We now know that a 4D light field $L(u,v,s,t)$ can be considered as a set of views captured by cameras defined by the two parallel planes, st and uv. In practice, the st and uv space can also be spherical. Compared with planar planes, spherical planes sample light rays more uniformly in directional space and provide a fixed distance to the object. Fig. 3 illustrates a light field sampling system with spherical planes [13]. An object is located at the very center of a sphere with cameras distributed on its surface. The cameras capture a 4D light field of the object as long as the sphere is larger than its convex hull.

With a sufficiently dense set of cameras, one can create virtual renderings of the light field at any position of the sphere surface—or even closer to the object—by resampling and interpolating the light rays [3], rather than synthesizing the view based on geometry information [14]. Fig. 3(a) shows a virtual view (located at the blue dot) that is closer to the object (the viridis shape) than the existing views. Because the radiance of the rays remains constant in free space, some of the light rays can be resampled from the existing views. However, there are still light rays that none of the cameras capture, and that must therefore be inferred through interpolation (e.g., the rays shown by dashed lines
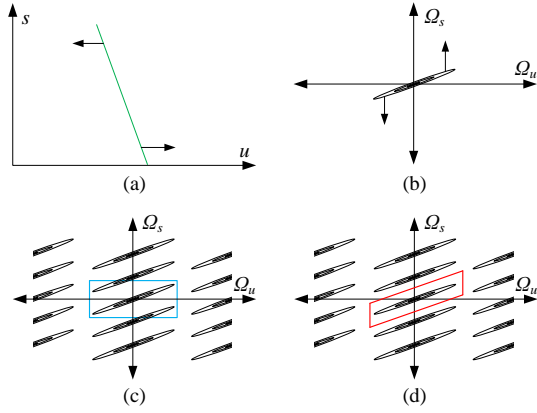
Fig. 4. (a) An EPI with a single line located slightly off the u-axis. (b) The power spectrum of the EPI in (a). The slope of the power spectrum with Ω_u-axis is determined by the slope of the line with the u-axis in the EPI. (c) Sampling the light field produces copies of the power spectrum in the Fourier domain. Using the reconstruction filter shown by the blue box will cause high frequency leakage from the neighboring copies of the power spectrum, leading to aliasing in the form of ghosting. (d) Using an alternative reconstruction filter (shown by the red box), we can accurately reconstruct the original EPI in (a).

in Fig. 3(a)). To render such rays, first, their intersection coordinates with the two planes are computed. Then, the nearest 16 sampled rays are used to interpolate the virtual ray. Fig. 3(b) depicts the two-plane representation of the 4D light field introduced in Sec. 2.1, in which four of the sampled rays are applied for interpolation. The above idea is called light field rendering [3].

In light field rendering, insufficient samples will cause ghosting effects in the novel views. However, it is impractical to acquire too many samples of a light field. For example, the synthetic *hallway* light field rendered by Levoy and Hanrahan [3] uses four light slabs, each of which involves 64×32 images with a raw size of 1608 MB. Chai et al. [15] and Lin et al. [16] investigated the minimum number of samples needed for light field rendering and concluded that the contributing pixels must at least touch each other to render novel views without ghosting. In other words, the maximum disparity between the neighboring views must be less than 1 pixel, a value closely related to the camera resolution and scene depth; the closer together the contributing pixels are (i.e., the less disparity), the sharper the interpolated point will be.

When the geometric information of the scene is known, the required number of samples can be greatly reduced, and even an irregularly sampled light field is appropriate for novel view rendering. Buehler et al. [17] presented a generalized model that combines light field rendering and depth image-based rendering using irregularly sampled images (i.e., unstructured input). To render a novel view, the input views were combined using a "camera blending field", that was related to the angular difference, estimates of undersampling, and the field of view (FOV).

2.4 Fourier Analysis

In this subsection, we focus on the analysis of light field sampling in the Fourier domain. Consider an EPI with a single line located slightly off the u-axis as shown in Fig. 4(a). The power spectrum of the EPI is shown in Fig. 4(b),
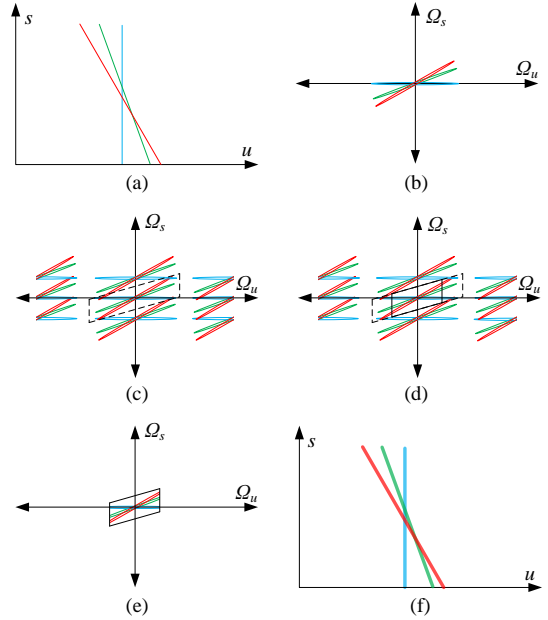

Fig. 5. (a) An EPI with three intersecting lines. (b) The power spectrum of the EPI in (a). (c) In the power spectrum of the sparsely sampled light field, the neighboring copies are overlap in the high frequency portion. Using the shear reconstruction filter shown by the box with dashed lines will still cause high frequency leakage. (d) The high frequency portion is filtered properly, producing the reconstruction filter shown by the box in the solid line. The filtered power spectrum is shown in (e), and the reconstructed EPI is shown in (f).

where the intersection angle with the Ω_u-axis is determined by the slope of the line in the EPI. Sampling the light field produces copies of the spectrum shown in Fig. 4(c). For a light field with sufficiently dense sampling in the angular dimensions, there will be no overlap between the copies along the Ω_s axis in the Fourier domain. Using the reconstruction filter shown by the blue box in Fig. 4(c) will cause high frequency leakage from the neighboring copies of the power spectrum, producing ghosting effects in the reconstructed views [36]. An alternative is to use a shear reconstruction filter (shown by the red box in Fig. 4(d)). In this way, the EPI shown in Fig. 4(a) can be accurately reconstructed.

However, issues in practical cases are usually more complicated. Fig. 5(a) shows an EPI containing three lines with different slopes intersecting each other. The resulting power spectrum of the sparsely sampled light field is shown in Fig. 5(c), where the neighboring copies overlap in the high frequency spectrum. Using the shear reconstruction filter shown by the box with dashed lines in Fig. 5(c) will still cause high frequency leakage from other copies of the power spectrum, producing ghosting effects. To reconstruct the light field without those ghosting effects, the high frequencies should be filtered properly (shown in Fig. 5(d)), producing a band-limited reconstructed light field [37]. The resulting power spectrum after filtering (shown in Fig. 5(e)) is close to the one before the sparse sampling. The reconstructed EPI shown in Fig. 5(f) is slightly blurrier than the original EPI due to the high frequency loss.

In addition, Fourier analysis of light fields has been extensively used in light transport [38], [39] and wave optics [40], [41], [42]. Ren [43] derived a 4D "Fourier slice

Category	Approach	Year	Implementation	Resolution	Capture speed	Notes
Multi-sensor	Yang et al. [18]	2002	8×8 camera array	$320 \times 240 \times 8 \times 8$	15-20fps	-
	Zhang and Chen [19]	2004	6×8 camera array	$320 \times 240 \times 6 \times 8$	15-20fps	-
	Wilburn et al. [20]	2005	10×10 camera array	$640 \times 480 \times 10 \times 10$	30fps	-
	ProFusion 25 [21]	-	5×5 camera array	$640 \times 480 \times 5 \times 5$	25fps	Portable device
	PiCam [22]	2013	4×4 camera array	$1000 \times 750 \times 4 \times 4$	-	Integrated unit
	Lin et al. [23]	2015	5×5 camera array	$1024 \times 768 \times 5 \times 5$	30fps	Microscopy
Time-Sequential	Light field gantry [24]	2002	Gantry	$1300 \times 1030 \times 62 \times 56$	5h/slab	-
	Kim et al. [25]	2013	Linear stage	$5616 \times 3744 \times 100$	-	3D light field
	Liang et al. [26]	2008	Programmable aperture	$3039 \times 2014 \times 5 \times 5$	above 0.5s	-
	Taguchi et al. [27]	2010	Mirror ball	-	-	Axial light field
Multiplexed (spatial)	Ng et al. [28]	2005	Microlens array	$292 \times 292 \times 14 \times 14$	16ms	Portable device
	Lanman et al. [29]	2006	Spherical mirror array	-	-	-
	Georgiev et al. [30]	2006	Lens-prism pairs	$700 \times 700 \times 4 \times 5$	-	-
	Manakov et al. [31]	2013	Camera add-on	$1800 \times 1200 \times 3 \times 3$	-	-
	Lytro Illum [4]	2014	Microlens array	$625 \times 434 \times 15 \times 15$	3fps	Portable device
	Levoy et al. [7]	2006	Microlens array	$120 \times 120 \times 17 \times 17$	1/15s	Microscopy
Multiplexed (frequency)	Veeraraghavan et al. [32]	2007	Attenuation mask	$228 \times 181 \times 9 \times 9$	-	-
	Marwah et al. [33]	2013	Attenuation mask	$480 \times 270 \times 5 \times 5$	-	Learning-based
	Antipa et al. [34]	2016	Diffuser plate	$170 \times 170 \times 11 \times 11$	-	-

TABLE 1

A summary of typical light field acquisition approaches. Note that for multi-sensor and time-sequential capture, the spatial resolution refers to the sensor resolution; and for multiplexed imaging, the spatial resolution refers to the resolution of a sub-aperture image after being extracted by the corresponding algorithm described in each reference. The spatial resolution for the Lytro Illum [4] refers to that of a sub-aperture image extracted with the Light Field Matlab Toolbox [35].
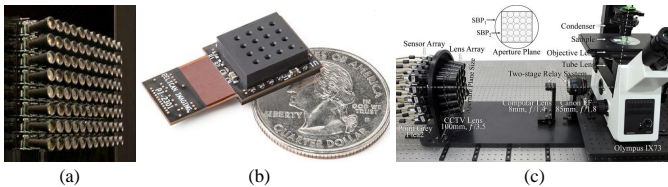

Fig. 6. Devices for multi-sensor capture. (a) A camera array system composed of 8×12 video cameras developed by Wilburn et al. [20]. (b) The Picam developed by Venkataraman et al. [22] was an ultra-thin monolithic 4×4 camera array that could be integrated into a cell phone. (c) Lin et al. [23] used a 5×5 camera array system to capture the light field of microscopic objects.

photography theorem", and indicated that a photograph formed with a full lens aperture is a 2D slice in the 4D light field. This mathematical theory was applied to analyze the performance of digital refocusing, and further used to yield a Fourier-domain approach for digital refocusing using light fields. Zhang and Levoy [41] indicated that the light field is equivalent to a smoothed Wigner distribution, which approaches the Wigner distribution under a geometric optics approximation. They further explored the correspondence between Fourier slice photography and wavefront coding using the Wigner distribution and light fields.

3 LIGHT FIELD ACQUISITION

This section focuses on existing devices or methods for light field acquisition. A conventional camera captures a 2D projection of a light field on its sensor plane by integrating the light rays that hit each pixel from every direction. In contrast, the devices or methods for light field acquisition measure the distribution of light rays in a directionally-resolved manner, avoiding angular integration.

However, sensors can only measure the information from two dimensions (usually two spatial dimensions) of

a scene at a single moment. To acquire a 4D light field, we need to capture multiple samples along the angular dimensions. Existing light field acquisition approaches can be divided into three fundamental categories [8], [9], [10], [44]: multi-sensor capture, time-sequential capture and multiplexed imaging. In the following, we will provide a detailed review of the existing light field acquisition approaches according to those three categories.

3.1 Multi-Sensor Capture

The multi-sensor capture approach requires an array of image sensors distributed on a planar or spherical surface to simultaneously capture light field samples from different viewpoints [3]. The spatial dimensions (u and v) of the light field are determined by the sensors, while the angular dimensions (s and t) are determined by the number of cameras and their distribution. Therefore, the 4D light field is recorded by the combination of the captured images.

In 2002, Yang et al. [18] described a design using an array of 8×8 video cameras for dynamic light field capture. To overcome data bandwidth problems, they employed a distributed rendering algorithm. In the same year, Wilburn et al. [45] applied 6 CMOS image sensors to record a synchronized video dataset. Each camera had a processing board that implemented MPEG2 compression to achieve a scalable capture. They further expanded the system to 125 video cameras and used it to capture multi-thousand frame-per-second (fps) videos [20], [46] (one configuration of this system is shown in Fig. 6(a)). Zhang and Chen [19] presented a system consisting of 6×8 cameras able to render novel views and reconfigure the camera positions to achieve better rendering quality. Chan et al. [47] used a 1×8 array of video cameras to capture and render a dynamic image-based representation called "plenoptic video". The
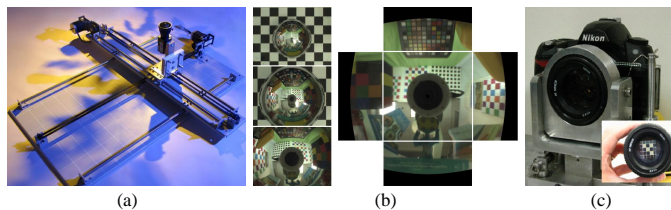
Fig. 7. Devices for time-sequential capture. (a) A motorized linear stage for mounting a single camera to capture 4D light fields by Unger et al. [49]. (b) The *axial light field* proposed by Taguchi et al. [27], consisting of a mirror ball and a single camera. The light field was captured by moving the camera along the mirror's axis of rotation. (c) A programmable aperture presented by Liang et al. [26] was applied to capture light fields through multiple exposures.

system was able to stream plenoptic video (256×256 spatial resolution) at 15 fps over the network. Liu et al. [48] also developed an 8×8 dynamic light field streaming system over broadband network.

Typical camera array systems are bulky and expensive and are thus unsuitable for most commercial uses. The Pro-Fusion 25 [21], manufactured by ViewPLUS Inc., includes an array of 5×5 VGA cameras packaged inside a small box and is able to capture light field video at 25fps. Venkataraman et al. [22] presented an ultra-thin monolithic camera array, which they called PiCam (Pelican Imaging Camera-Array, Fig. 6(b)). The PiCam is a 4×4 camera array; each camera is able to capture 1000×750 pixels. The entire device is smaller than a coin, and can be integrated into a cell phone [6].

Moving to smaller scene scales, Lin et al. [23] presented a camera array-based light field microscopy setup to measure the distribution of light passing through a specimen volume. The system, shown in Fig 6(c), comprised a 5×5 sensor array operating at a resolution of 1024×768 at 30fps.

The multi-sensor capture approach is able to capture a light field instantaneously, and is thus competent to record a light field sequence. Early multi-sensor systems were inevitably bulky and expensive due to large amounts of video cameras. However, we can see the potential of this light field acquisition method through the recent cheaper and more portable designs.

3.2 Time-Sequential Capture

In contrast with the multi-sensor approach, a time-sequential capture approach uses a single image sensor to capture multiple samples of the light field through multiple exposures. The typical approach uses a sensor mounted on a mechanical gantry to measure the light field at different positions [3], [11].

Two gantry systems were presented by the Computer Graphics Laboratory at Stanford University [24]: one was a computer-controlled gantry with four degrees of freedom, translation in X and Y, nod and shake; another was a *Lego Mindstorms* gantry in which the motors have rotatory controllers enabling the camera to move along accurate and well-defined paths. Unger et al. [49] used a single camera mounted on a motorized linear stage with two degrees of freedom, translation in X and Y, to capture light fields. Fig. 7(a) shows the system by Unger et al. [49]. A similar design with only one degree of freedom was presented by Kim et al. [25]. The system was applied to capture 3D outdoor light

fields with high angular resolution. To acquire a light field with large FOV, Dansereau et al. [50] presented a compact optical design combining a monocentric lens with microlens arrays (which will be introduced in the following section). The system was mounted on a mechanical arm that can be rotated around a fixed axis to enable large FOV capture.

The approaches above capture images at different viewpoints by moving the image sensor, which requires high-precision control and is time consuming. Fortunately, some fast time-sequential capture approaches have also been introduced. Ihrke et al. [51] proposed using a planar mirror and a high dynamic range video camera to record light fields. The moving mirror produces different viewpoints to be captured by the camera. Taguchi et al. [27] proposed a system consisting of a mirror ball and a single camera, as shown in Fig. 7(b). The light field, which they called "axial light field", was captured by moving the camera along the mirror's axis of rotation.

Rather than applying a reflective surface, Liang et al. [26] presented a programmable aperture for light field acquisition, shown in Fig. 7(c). The aperture patterns were encoded into an opaque slit of paper or a liquid crystal array, allowing the camera to capture the light rays from certain angles. By synchronizing the aperture patterns with the camera exposures, the light field can be acquired at full sensor resolution. The entire capturing procedure required 25 patterns for a light field with 5×5 angular resolution, and required 10 to 20ms for each exposure. The spatial resolution of each image was 3039×2014, but was downsampled to 640×426 to achieve high computational efficiency.

Compared with multi-sensor systems, time-sequential capture systems require only a single sensor, which reduces the cost of the entire system. In addition, time-sequential capture systems are able to capture light fields at dense angular resolutions, which multi-sensor systems cannot do because of the high associated cost. However, the capture process in time-sequential systems is always time-consuming; thus, they are appropriate only for static scenes.

3.3 Multiplexed Imaging

The last approach aims to encode the 4D light field into a 2D sensor plane, by multiplexing the angular domain into the spatial (or frequency) domain. It allows dynamic light field capture with a single image sensor, but imposes a trade-off between the spatial and angular resolutions (i.e., one can obtain densely sampled images in the spatial domain with sparse samples in the angular domain, and viceversa). Multiplexed imaging can be further divided into spatial multiplexing and frequency multiplexing.

3.3.1 Spatial Multiplexing

In spatial multiplexing, an interlaced array of elemental images representing samples from different 2D slices of the light field are captured by the sensor. Most spatial multiplexing approaches are implemented using an array of microlenses or a lenslet array mounted on the image sensor. Interestingly, this is one of the first approaches for light field imaging: in 1908, Lippmann used this approach in the development of his "integral photography" [52].

In 1992, Adelson and Wang [53] described a novel camera for "single lens stereo" that incorporates a main lens
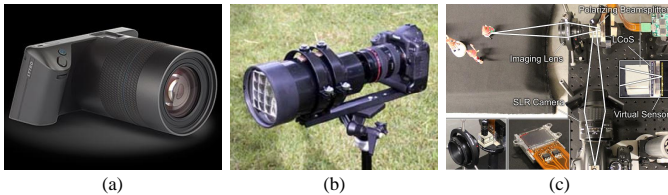
Fig. 8. Devices for multiplexed imaging. (a) Lytro Illum [4], a "plenoptic camera 1.0" that is currently commercially available. It uses a lenslet array to interlace the images of different views on a single sensor. (b) An integral camera that uses a 4×5 array of lenses and prisms to capture a low angular resolution light field developed by Georgiev et al. [30]. The angular resolution can be upsampled by view interpolation techniques. (c) Marwah et al. [33] described an attenuation mask-based scheme for frequency multiplexing capture. Different samples in the light field were encoded into different frequency bands.

and a lenticular array placed before the sensor. The device could capture a 5×5 angular resolution light field with 100×100 pixels in each image. Ng et al. [28] presented a hand-held light field camera by inserting a 296×296 lenslet array between the sensor and main lens. This device looked and operated like a conventional camera and could capture 14×14 angular resolution light fields.

The spatial multiplexing devices described above belong to the type referred to as "plenoptic camera 1.0"; in it, each microlens captures one position in the scene. Two commercially available cameras that fall under the "plenoptic camera 1.0" category are those by Raytrix [5] and Lytro [4]. They are designed for industrial and consumer use, respectively; Fig. 8(a) shows the commercially available "Lytro Illum". Wei et al. [54] proposed a novel design that introduced sample irregularities and lens aberrations into a light field camera to improve its quality and usability. Each microlens in a "plenoptic camera 1.0" captures the angular distribution of the radiance. By gathering pixels in the same coordinate of each subview, we can obtain an image located at a certain viewpoint. An alternative version uses multiple microlenses to capture the same position in a scene, referred to as "plenoptic camera 2.0", and the angular information is spread across each microlens [55].

Instead of using a lenslet array, filters and mirror arrays are also employed in some light field capture approaches. Horstmeyer et al. [56] used multiple filters placed in the pupil plane of the main lens and a pinhole array on the sensor plane to capture light fields. The filters could be exchanged and reconfigured, providing a high degree of flexibility. Similarly, Manakov et al. [31] developed a single reconfigurable add-on that enables plenoptic imaging with standard cameras including multi-spectral, high dynamic range and light field imaging. Similar to the "axial light field" described in Sec. 3.2 that captured a light field reflected by a spherical mirror, Unger et al. [49] and Lanman et al. [29] presented an array of spherical mirrors to capture light fields. A slight difference is that their designs used multiple mirrors and thus turned the time-sequential capturing into spatial multiplexing.

In addition to measuring macroscopic objects or scenes, spatial multiplexing has also been applied to capture specimens at micro-scales [7], [57], [58], [59]. Levoy et al. [7], [57] inserted a microlens array into the optical path of a conventional microscope to record light fields of biological

specimens on a single sensor plane. Compared with light field microscopy systems based on camera arrays [23] discussed in Sec. 3.1, Levoy's system captures light fields with high angular resolution (17×17), but at reduced spatial resolution. Debise and Ihrke [60] developed a light field microscope using a consumer light field camera, Lytro, and an optical matching system, and achieved a maximum spatial resolution of about $6.25\mu m$.

Spatial multiplexing has been the most extensively used approach to capture light fields. It enables capturing a light field in a single image with a single exposure. However, there is a problem inherent to spatial multiplexing approaches, and that is the trade-off between angular and spatial resolution at the image sensor. To handle this problem, Georgiev et al. [30] sacrificed angular resolution for higher spatial resolution in light fields. They presented a 4×5 array of lenses and prisms cut into squares placed in front of a conventional camera (shown in Fig. 8(b)) to capture a low angular resolution light field. Then, interpolation techniques were applied to synthesize a dense angular resolution light field. Many researchers have focused on super-resolving a light field in the spatial and/or angular domain; please refer to Sec. 4 for more details on super-resolution.

3.3.2 Frequency Multiplexing

Unlike the spatial multiplexing approach, which interlaces the 2D light field slices on the sensor plane, the frequency multiplexing approach encodes different 2D slices of the light field into different frequency bands. Typically, frequency multiplexing approaches use a modulation mask to achieve a certain property in the Fourier domain [44].

Veeraraghavan et al. [32] described a theoretical framework for 4D light field acquisition using an attenuation mask in the optical path of a conventional image sensor; they termed the approach "Dappled Photography". Rather than blending the light rays, the patterned mask attenuated and encoded them on the image sensor. The Fourier transformed image is rearranged into 4D planes and then an inverse Fourier transform is applied to restore the light field. As indicated by Lanman et al. [61], the patterned masks are equivalent to a truncated Fourier series approximation of a pinhole array for high angular sampling rates. Ashok and Neifeld [62] further described two separate architectures for light field imaging allowing compressive light field imaging in either the angular or spatial dimensions. In their designs, an amplitude mask was employed behind the main lens for angular compressive light field imaging; and an amplitude mask was placed over each microlens, for spatial compressive light field imaging. Antipa et al. [34] encoded multiplexed spatial and angular information using a diffuser plate. Compared with an attenuation mask, the diffuser plate allows higher light throughput, and provides an inexpensive and flexible means for frequency multiplexing. Pégard et al. [63] applied the compressive light field imaging technique to capture specimens at microscopic scales, while Cohen et al. [64] applied a wavefront coding technique that produces a more uniform distribution across depth to enhance the performance of the light field microscope.

Taking advantage of learning techniques, Marwah et al. [33] described an attenuation mask-based scheme and used an over-complete dictionary to reconstruct the light field

from its coded 2D projection (shown in Fig. 8(c)). Wang et al. [65] further improved the quality of reconstructed light fields using a random convolution CMOS sensor, which is able to maintain more information by means of correlation.

Although frequency and spatial multiplexing are two different multiplexed imaging approaches, they actually have a close relationship in terms of their multiplexed patterns. Ihrke et al. [44] indicated that patterns derived for Fourier space can be reconstructed by spatial interpolation, and spatial patterns can also be reconstructed with a Fourier space algorithm. In addition, Fourier reconstruction for frequency multiplexing is equivalent to spatial reconstruction with a sinc filter kernel, and the resulting ringing artifacts can be significantly removed by a spatial filter, such as cubic interpolation or an edge-preserving filter.

3.4 Discussion

We have introduced various light field acquisition approaches in this section, and Table 1 provides a summary of some of the typical ones. As we can see in the table, most multi-sensor approaches are able to capture light field video but require a large number of cameras. Meanwhile, time-sequential approaches use a single camera, and are able to capture high spatial and angular resolution light fields, but its time-consuming capturing procedure makes them impractical for dynamic scenes. Multiplexed imaging enables the capability of high-speed light field capture with a single camera, but leading to the problem of a resolution trade-off in the spatial and angular dimensions. Current light field acquisition devices at the commercial level use spatial multiplexed imaging. However, due to the large volume of data that needs to be recorded, it is hard to meet the capture speed needed for light field video capture; for instance, Lytro Illum [4] can only record light field video at 3fps. For frequency multiplexing, the modulation mask will inevitably lose luminous flux, leading to longer exposure times for each frame. It is worth mentioning that, as sensors get smaller and cheaper, multiple sensors can be assembled into a portable box (e.g., ProFusion 25 [21]), or even integrated onto a small monolithic camera array (e.g., PiCam [22]). Each sensor is able to record video at a high frame rate (e.g., ProFusion 25 [21] can record light field video at 25fps). With this trend, light field cameras could find its way into mobile devices, such as mobile phones and tablet computers, in a near future.

4 LIGHT FIELD SUPER-RESOLUTION

Due to their portability and low cost, plenoptic cameras have sparked renewed interest in light field imaging [4], [5]. However, limited by the sensor resolution and the processing speed, it is still very challenging to capture a high spatio-angular resolution light field and high frame-rate light field video. Moreover, the spectral range and resolution of current plenoptic cameras is in most cases restricted to that of three (RGB) channels.

To mitigate these problems on the basis of existing hardware, many researchers have focused on light field super-resolution in different dimensions. In this section, we investigate light field super-resolution mainly in terms of
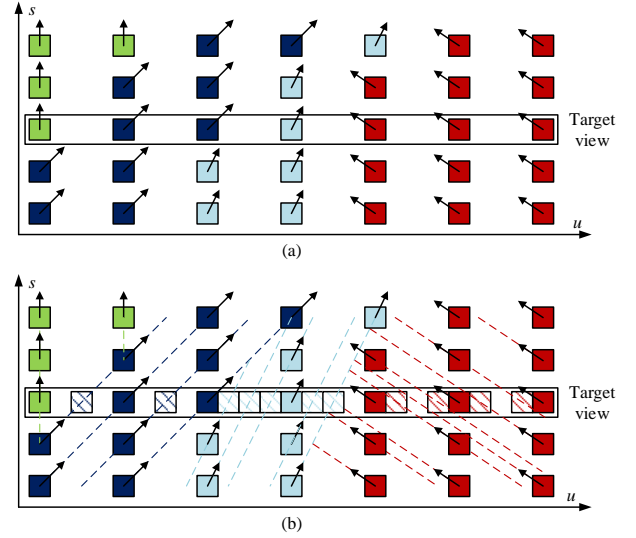

Fig. 9. Principle of light field super-resolution in the spatial dimensions. (a) Each pixel in an EPI is assigned a disparity label (indicated by the arrows) with subpixel accuracy. (b) The intensity values in other views are propagated to the target view. Due to the non-integral shifts between two corresponding pixels (the slope of the dashed lines in the figure), some pixels propagate to positions between the neighboring pixels (the pixels in the hatched-boxes), resulting in a super-resolved view.

the spatial and angular dimensions, which have received the most attention. We will nevertheless, although more briefly, also look into approaches for super-resolving the spectral and temporal dimensions.

4.1 Spatial Super-resolution

The light field captured by the Lytro Illum [4] is 7728×5368 pixels—higher than most current DLSR cameras; however, the resolution of a single sub-aperture image extracted by Light Field Matlab Toolbox [35] is only 625×434. A straightforward way to super-resolve the light field in the spatial dimensions is to employ single-image super-resolution techniques to upsample each of the views in the light field. However, this type of method treats each image in the light field as an individual entity, which of course results in underuse of the full potential of the information in the light field. As indicated in [55], [75], the true spatial resolution of a sub-aperture image is not limited by lenslet resolution, i.e., one can extract a sub-aperture image by taking advantage of the properties of a light field. In a 4D light field captured by plenoptic cameras, each image is a 2D sample from a viewpoint with only a slight difference with respect to its neighboring views, leading to non-integral shifts between two corresponding pixels in the images. By taking advantage of these non-integral shifts, pixels in the neighboring views can be propagated to the target view [55] (see Fig. 9).

Following the general principle described above, typical spatial super-resolution methods first estimate the depth of the scene to infer those non-integral shifts, and then super-resolve the images using various optimization frameworks. Depth estimation will be discussed in Sec. 5. For the inference of non-integral shifts, Chan et al. [76] presented a mathematical model to simulate the super-resolution of light fields. By explicitly introducing Lambertian reflectance

Approach	Main feature	Domain	Evaluation (PSNR)					
			Synthetic data			Real-world data		
			Buddha	*Mona*	*Still life*	*Maria*	*Couple*	*Tarot*
Wanner et al. [66]	Variational optimization	Spatial (2×)	34.50	-	24.93	35.18	25.54	26.66
PaSR [67]	Hybrid input, patch-based	Spatial (8×)	-	-	25.11	34.30	27.38	30.77
iPADS [68]	Hybrid input, patch- and depth-based	Spatial (8×)	-	-	26.45	36.10	28.65	31.63
Wanner et al. [66]	Variational optimization	Angular (using 8 views)	42.84	-	30.13	40.06	26.55	28.71
Pujades et al. [69]	Robust to depth	Angular (using 8 views)	42.37	-	30.45	40.10	28.50	28.88
DAPS [70]	Phase-based, micro-baseline	Angular (1 × 2 to 9 × 9)	44.71	-	34.81	40.81	32.07	-
Wu et al. [71]	Learning-based	Angular (4×)	43.20	44.37	-	-	-	-
LFCNN [72]	Learning-based	Spatio-angular (2×)	36.86	37.56	-	-	-	-
Wang et al. [73]	Learning-based	Temporal (10×)	-	-	-	-	-	-

TABLE 2

Comparison of light field super-resolution approaches. The evaluation provides numerical results of super-resolved light fields using the listed implementations. For angular super-resolution approaches by Wanner et al. [66] and Pujades et al. [69], 8 views were used to synthesize each novel view. The datasets *Buddha*, *Mona*, *Still life*, *Maria* and *Couple* are from the HCI light field dataset [74], and *Tarot* is from the (New) Stanford light field archive [24].

priors in the image formation model, Bishop et al. [77] formulated the reconstruction of the light field in a Bayesian framework, allowing more information to be recovered in the super-resolved light field. Mitra and Veeraraghavan [78] proposed a patch-based approach using a Gaussian mixture model (GMM). They used a fast subspace projection technique to generate a disparity map and then modeled the light field patch as Gaussian random variables conditioned on its disparity value. Wanner et al. [66] used a structure tensor method to compute a continuous depth map on the EPI, then applied a variational optimization framework to super-resolve the images based on the depth map.

In contrast with the approaches above, which use the inherent information acquired by the plenoptic cameras, Boominathan et al. [67] proposed a hybrid light field imaging system that uses a single high resolution image captured by a DSLR camera to super-resolve the low resolution images in the light field. They extracted patches from the high resolution image with the lowest matching cost to those in the low resolution light field, and then blended the high resolution patches using a weighted average. Based on this approach, Wang et al. [68] applied an additional depth-based view synthesis method to warp the high resolution image to the position of each low resolution image, and built an iterative refinement pipeline to improve the warped images. Compared with the early work by Boominathan et al. [67], higher frequency information can be reconstructed in the low resolution images.

Taking advantage of convolutional neural networks (CNNs), Yoon et al. [72] presented a CNN-based approach to perform spatial super-resolution without depth estimation. Farrugia et al. [79] trained linear projections between subspaces of reduced dimension in which patch-volumes extracted from the light field reside.

4.2 Angular Super-resolution

Many studies have focused on angular super-resolution using a small set of views with high spatial resolution. These views can be modeled as a reconstruction of the plenoptic function using limited samples. The existing angular super-resolution approaches can be divided into two categories: those that use depth estimation, and those that do not.
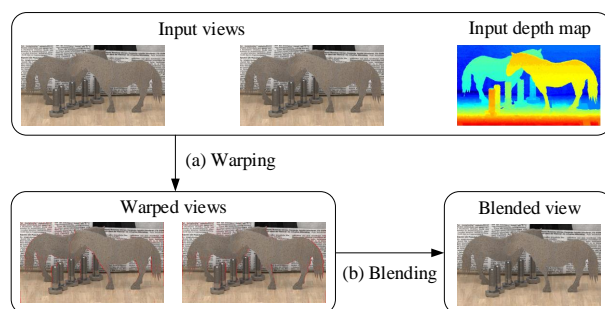

Fig. 10. Typical pipeline of depth image-based view synthesis. (a) Using the input depth map to warp the input images to novel views. In the novel, warped views, the red pixels are blank areas, also called "holes", caused by the warping operation. (b) Blending the warped views into a single image. The dataset in this figure is the *Horses* scene from the HCI database [74].

4.2.1 Depth Image-Based View Synthesis

Depth image-based view synthesis approaches typically first estimate the depth information and then warp the existing images to the novel view based on the estimated depth. The warped views are blended in a specific way. Fig. 10 illustrates this process.

Georgiev et al. [30] employed a segmentation-based optical flow method to compute the flow between an image and two neighboring views. Then, the novel view was generated by weighting the three warped views. Pearson et al. [80] introduced a depth layer-based method for synthesizing an arbitrary novel view. They assigned each pixel to a particular layer and then warped the pixels using a probabilistic interpolation approach. Wanner et al. [66], [81] formulated the view synthesis problem as an energy minimization one. Chaurasia et al. [82] presented a novel warping scheme using a superpixel technique, and then blended the warped views with weights specified by camera orientation and the reliability of depth information. Pujades and Devernay [69] proposed a view synthesis approach by optimizing a novel cost function with a Bayesian formulation that is robust to errors in the estimated depth. Zhang et al. [70] presented a phase-based approach using a micro-baseline stereo pair. They also introduced a disparity-assisted phase-based synthesis strategy to integrate the disparity (depth)
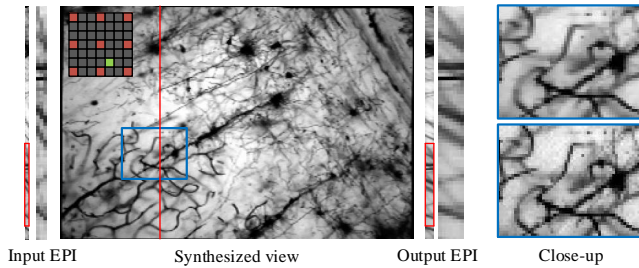
| Input EPI | Synthesized view | Output EPI | Close-up |

Fig. 11. Angular super-resolution on the Stanford microscope light field data *Neurons* $20\times$ [7] using 3×3 input views (output is 7×7 views). The angular super-resolution approach was proposed by Wu et al. [71], and models the reconstruction as a CNN-based angular restoration of the EPI (see the input EPI and the output EPI in the figure). The views in the blue boxes show the close-up version of a synthesized view (top) and corresponding ground truth view (bottom).

information into the phase term when warping the input image to the novel view. Zhang et al. [83] proposed a layered patch-based synthesis method for multiple applications.

Typical depth-based view synthesis approaches rely heavily on the estimated depth, which is sensitive to textureless and occluded regions. In addition, they often focus on the quality of depth estimation rather than the quality of the synthetic views. In recent years, some studies based on CNNs aimed at maximizing the quality of the synthetic views have been presented. Flynn et al. [84] proposed a deep learning method to synthesize novel views using a sequence of images with wide baselines. Kalantari et al. [85] used two sequential CNNs to simultaneously estimate depth and color by minimizing the error between the synthetic views and ground truth images.

4.2.2 Light Field Reconstruction without Depth

Depth image-based view synthesis approaches involve depth estimation, which tends to fail in occluded regions, as well as in glossy or specular ones. An alternative approach is based on sampling and consecutive reconstruction of the plenoptic function. For a light field that meets the requirement of minimum sampling described in Sec. 2.3, novel views can be produced using light field rendering techniques [3].

For sparsely sampled light fields, direct interpolation will cause ghosting effects in the rendered views. To mitigate this effect, some studies have investigated light field reconstruction in the Fourier domain. Levin and Durand [86] proposed a linear view synthesis approach using a dimensionality gap prior to synthesize novel views from a set of images sampled with a circular pattern. Shi et al. [87] considered light field reconstruction as an sparsity optimization in the continuous Fourier domain. They sampled a small number of 1D viewpoint trajectories formed by a box and 2 diagonals to recover the full light field. Didyk et al. [88] used the phase information from a complex steerable pyramid decomposition to synthesize novel views with a small parallax. However, those reconstruction approaches always require the light field to be sampled in specific patterns, which imposes a restriction for practical usage. Instead, Vagharshakyan et al. [89], [90] considered angular super-resolution as an inpainting problem on the EPI, and used an adapted discrete shearlet transform to super-resolve the angular resolution of a light field.

Also, some learning-based methods have been presented for depth-free angular super-resolution. Yoon et al. [72] proposed a CNN-based approach that used two neighboring views in the vertical (horizontal) angular dimension to generate the middle view. However, they underused the potential of the entire angular information, and could only achieve a fixed super-resolution rate. Instead, Wu et al. [71] took advantage of the clear texture structure of the EPI and modeled the angular super-resolution as a CNN-based angular restoration of the EPI. Before feeding the EPI to the network, the spatial low-frequency component was extracted to avoid aliasing effects in the angular dimension of the EPI. Fig. 11 shows an angularly super-resolved result on a microscope light field [7]. In this case, a light field of 3×3 views was input to the network, resulting in a 7×7 views super-resolved light field.

4.3 Temporal Super-resolution

Due to the high dimensional data in a light field, current commercial light field cameras usually fail to capture light field video with a satisfactory frame rate. For example, Lytro Illum [4] captures a light field containing 7728×5368 pixels in a single shot, and can only achieve 3fps in the continuous shooting mode. Most researchers have focused on video frame interpolation [91], [92]. But the problem becomes much more challenging when it comes to a light field video for the following reasons: First, the frame rate of a light field video can be extremely low; second, 4D light fields rather than 2D images need to be considered.

To super-resolve the temporal dimension of a light field, a hybrid imaging system composed of a plenoptic camera (Lytro Illum [4]) and a DSLR camera was developed by Wang et al. [73]. In this system, the plenoptic camera captured a light field sequence at 3fps and the DSLR camera captured the corresponding 2D video at 30fps. Then, a learning-based approach was designed to propagate the sequences containing a light field to the sequences containing only 2D images, and finally output a light field video at 30fps. Specifically, a spatio-temporal CNN and an appearance CNN were designed, where the former was trained to warp the input images from the 2D video and the light field images to the target angular view, and the latter was trained to combine the warped images into the final image.

4.4 Spectral Reconstruction

Rather than directly capturing a hyperspectral light field, which would sacrifice the resolution of other dimensions, practical approaches split the task into 4D light field imaging and hyperspectral imaging [93], [94], [95], and then reconstruct the 5D light field in specific ways.

Wu et al. [93] presented a hyperspectral light field microscope that contains a 5×5 camera array, where each camera was mounted with a color filter (25 filters in total) to capture the scene at a certain spectral band. A 4D deconvolution algorithm is applied to reconstruct the hyperspectral light field from the spatial-spectral coupled sampling. Instead of using a camera array and taking advantage of a hybrid imaging approach, Xiong et al. [94] developed a hyperspectral light field imaging system using a plenoptic camera and

a coded aperture snapshot spectral imager (CASSI). In this system, the incident light was divided by a beam splitter and then captured by the plenoptic camera and CASSI, where the plenoptic camera recorded a 4D light field and RGB spectral information, and the CASSI encoded a 2D image with 27 spectral bands. To reconstruct the final hyper-spectral light field, self-learned dictionaries were trained to exploit the large correlations across the angular and spectral dimensions.

4.5 Discussion

Resolution trade-offs being an intrinsic problem of light field acquisition, researchers have devoted great attention to light field super-resolution. Table 2 lists numerical evaluation (PSNR) of some representative light field super-resolution approaches. From these numerical results and the discussed techniques, we can draw the following conclusions:

- Light field super-resolution can greatly benefit from hybrid imaging systems; these have several merits, such as much smaller data size and compressibility. Typical approaches, such as PaSR [67] and iPADS [68], have an extremely large super-resolution rate.
- Approaches based on learning techniques (e.g., CNNs, or sparse coding) can achieve better performance, especially when taking advantage of properties of light fields structure, such as the EPI structure. Among numerous super-resolution techniques, the learning-based approaches, e.g., Wu et al. [71] and Kalantari et al. [85], currently constitute the state of the art.
- Super-resolution of non-Lambertian regions can be a great challenge, especially for a microscopy light field, where the depth information is difficult to obtain.

5 LIGHT FIELD DEPTH ESTIMATION

In this section, we present techniques devoted to depth map estimation from light field data. The four-dimensional light field representation essentially contains multiple views of the scene, making depth map estimation possible. In comparison to stereo vision-based methods, light field-based methods do not require camera calibration, making it convenient for data acquisition in real-world scenes. Moreover, the peculiar structure of the light field expands the disparity space to a continuous space [66], making depth estimation more robust and precise. However, processing this high dimensional data requires large computing times and resources, which is still a barrier for real-time depth estimation from light fields.

Typical light field depth estimation approaches first estimate an initial depth map in specific ways, then global optimization frameworks or local smoothing approaches are used to refine the depth map. Fig. 12 demonstrates this process. In this section, we will investigate light field depth estimation following this scheme.

5.1 Initial Depth Estimation

The baseline between adjacent views in a light field is narrow, which makes it difficult to recover disparity from two

(a) Light field

(b) Ground Truth
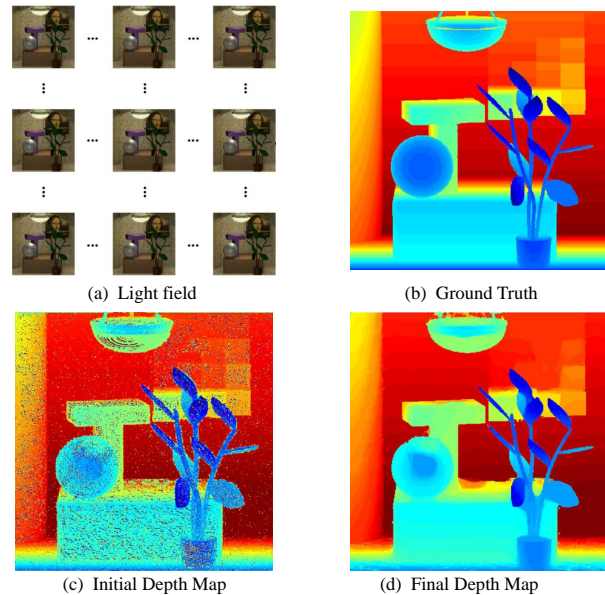
(c) Initial Depth Map

(d) Final Depth Map

Fig. 12. Typical depth estimation pipeline using light field data. (a) Input light field. (b) The ground truth used for comparison with the final result. (c) Initial estimated depth map (please refer to Sec. 5.1), which contains outliers due to noise, occlusions and inherent matching uncertainty caused by textureless regions. (d) Final depth map after depth refinement (please refer to Sec. 5.2). The depth map is refined by means of a MRF framework (see the discretization effect in the background). The scene in this figure is *monasRoom* from the HCI database [74].

views using traditional stereo matching methods. Therefore, instead of using stereo matching methods, constraints and cues which take advantage of all the views together are used to estimate the depth map from a light field image. Existing initial depth estimation approaches can be divided into three categories: sub-aperture image matching-based methods, EPI-based methods and learning-based methods.

5.1.1 Sub-Aperture Image Matching-based Methods

As described in Sec. 2.2, a sub-aperture image is recorded by light rays arriving at one point in the st plane. Each pair of sub-aperture images has a very narrow baseline, making the disparity range of sub-aperture images also quite narrow. For example, the disparity range between adjacent sub-aperture views of the Lytro camera is smaller than 1 pixel [96]. With such narrow baselines, the sub-pixel shift in the spatial domain usually involves interpolation with blurriness, which leads to poor performances on correspondence-matching approaches. Moreover, the matching costs of stereo correspondence are highly ambiguous.

Therefore, instead of using stereo matching, different constraints that leverage the structure of light fields are used to estimate the initial depth map. Different from stereo matching based on two views, all views are contained in the constraints. Yu et al. [96] explored the geometric structure of 3D lines in ray space to improve the light field triangulation and stereo matching. Since the method uses stereo matching to first estimate each pixel's disparity, it does not perform well on the datasets with small disparity range. To solve the problem, Heber et al. [97] proposed a novel principal component analysis (PCA) matching term for multi-view stereo reconstruction. They assumed that if one were to warp the images of various views to a common
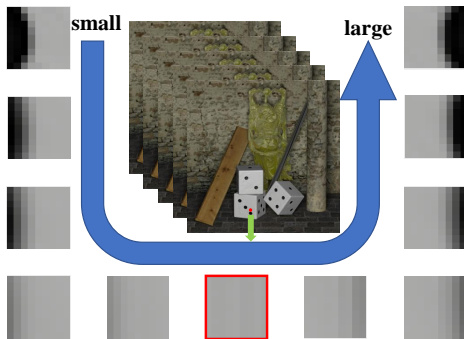
Fig. 13. We display a patch, consisting of all views recorded for a point, as would appear under different depth candidates. The red point marked in the light field view is the one selected. In counter-clockwise order, the depth values of the patches range from smaller to larger depths. Among these candidates, since the pixels in the patch marked with a red frame have the largest similarity among them, it is that depth the one that has the largest probability of being the actual depth of that point.

warping center and then consider each warped image as one row in a matrix, the matrix should be low rank. In order to match sub-aperture images with an extremely narrow baseline, Jeon et al. [98] presented an algorithm based on the cost volume. The cost volume is computed to evaluate the matching cost of different disparity labels, which is defined using the similarity measurement between the sub-aperture images and the center view sub-aperture image shifted at different sub-pixel locations. To generate a better result on occlusions, Wang et al. [99] developed a novel occlusion scheme for their occlusion-aware depth estimation algorithm. By ensuring photo-consistency in only the occluded view region, their method enabled identification of occluded edges, a technique that could be useful in other applications. Based on this occlusion model, Zhu et al. [100] extended the approach above to the context of multi-occluder occlusion. By using the concept of the surface camera (SCam), which models angular radiance distribution with respect to a 3D point, Chen et al. [101] introduced a bilateral consistency metric on SCams for light field depth map estimation.

Also, various other cues can be exploited in depth estimation, the defocus cue being the most prominent one. Ng's work [28] noted that a light field image can be refocused on any area by shearing the light field using a combination of digital refocusing and Fourier refocusing. After refocusing at a new depth, the light rays from any in-focus scene point will be recorded in the same spatial location—a concept also called angular coherence (shown in Fig. 13). We can refocus the light field at different depth candidates. The absolute difference between the central view and the other angular views reflects the depth probability.

Based on Ng's work [28], Tao et al. [102] combined the defocus cue and the correspondence cue to estimate the depth of scenes. In addition, they extended their analysis, using the additional cue of shading, to refine fine details in the shape [103]. For light fields including a scattering medium, such as underwater scenes or scenes with fog, Tian et al. [104] combined the correspondence and defocus cues with a novel transmittance-based depth cue to remove the influence of backscatter and attenuation. Williem et al. [105] proposed two novel cues, which are an angular entropy metric and the adaptive defocus response, to improve the robustness of depth estimation to occlusion and noise. Lin

et al. [106] presented two features for depth estimation: one feature is the property that non-occluding pixels exhibit symmetry along the focal depth dimension centered at the in-focus slice; the other is a data consistency measure based on the analysis of a synthesized view focused at a hypothesized depth. Lee and Park [107] proposed a depth estimation approach by accumulating binary maps, which are obtained by separating the foreground and background with respect to a series of focused planes.

For the particularly challenging case of depth estimation for glossy or specular surfaces, Tao et al. [108] presented a novel theory of the relationship between light-field data and reflectance in the dichromatic model. Wang et al. [109] derived a spatially-varying (SV) BRDF-invariant theory for recovering 3D shapes and reflectance from light field cameras to estimate the depth of non-Lambertian planes. They proposed an equation relating depths and normals that could be derived to estimate the depth map.

5.1.2 EPI-based Methods

Instead of performing stereo matching on sub-aperture images, novel approaches considering the properties of the light field structure have been developed. Before the rise of light field cameras, epipolar geometry was used to achieve scene reconstruction from images taken from different views. After the introduction of the light field camera, EPIs could be generated directly from light field images; in them, the slopes of the lines are indicative of the depths of the different objects (as we mentioned in Sec. 2.2). Most EPI-based depth map estimation approaches rely on measuring the slopes in EPIs using different optimization techniques.

In 1987, Bolles et al. [12] presented a technique for building a 3D description of a static scene from a dense sequence of images. The technique utilized knowledge of the camera motion to form and analyze slices of the solids generated by these images. Criminisi et al. [115] addressed the problems of achieving a dense scene reconstruction from a sequence of images, as well as analyzing and removing specular highlights. The key was to directly exploit the high degree of regularity found in the EPI volume. Basha et al. [116] proposed a method for estimating the 3D structure and the dense 3D motion (scene flow) of a dynamic non-rigid 3D scene, using a camera array. Their core idea was to use a dense multi-camera array to construct a novel, dense 3D volumetric representation of the 3D space in which each voxel holds both an estimated intensity value and a confidence measure of that value.

Kim et al. [25] introduced a method to perform scene reconstruction of complex, detailed environments from high angular resolution EPIs. Their method computed the 3D depth information for all the visible scene points. They applied a confidence measure in EPI space to compute the reliability of the estimated depth. Wanner et al. [66] utilized the structure tensor in EPI space to estimate the local direction of a line. Based on their work, Li et al. [110] made use of the structure information to build a reliability map for EPIs. Krolla et al. [111] extended depth estimation using the structure tensor to spherical light fields. Diebold et al. [113] modified the structure tensor approach

Category	Approach	Main feature	Evaluation (MSE)					
			Buddha	MonasRoom	Still life	Horses	Medieval	Papillon
Matching-based	Wang et al. [99]	Occlusion-aware	2.06	0.87	7.24	1.76	3.62	0.43
	Jeon et al. [98]	Sub-pixel accuracy	2.75	0.43	1.52	4.91	5.34	1.9
	Yu et al. [96]	Line-assisted graph cut	5.03	18.8	12.88	18.8	28.45	17.29
	Heber et al. [97]	PCA matching term	-	-	-	-	-	-
	Tao et al. [102]	Defocus cues	-	-	-	-	-	-
EPI-based	Kim et al. [25]	Large scene reconstruction	-	-	-	-	-	-
	Li et al. [110]	Sparse linear optimization	0.64	0.64	3.25	0.95	2.1	2.28
	Krolla et al. [111]	Spherical light field	-	-	-	-	-	-
	Wanner et al. [112]	Total variation	-	-	-	-	-	-
	Diebode et al. [113]	Modified structure tensor	-	-	-	-	-	-
Learning-based	Johannsen et al. [66]	Sparse coding	0.57	0.65	2.95	3.26	0.84	1.85
	Heber et al. [114]	CNN-based	-	-	-	-	-	-

TABLE 3

Comparison of light field depth map estimation approaches. These approaches are classified into different categories according to their initial depth estimation approaches. The evaluation provides numerical results of MSEs using the listed implementations. The datasets *Buddha*, *MonasRoom*, *Still life*, *Horses*, *Medieval* and *Papillon* are from the HCI light field dataset [74].

for heterogeneous light fields in which the EPIs have non-constant intensity in each line. Ziegler et al. [117] extended depth estimation in EPI space to the 4D epipolar volume. To handle highly noisy and strong occlusions, Zhang et al. [118] integrated a spinning parallelogram operator in the 2D EPI, where the operator measures the slopes of an EPI by maximizing distribution distances between two parts of the parallelogram window.

5.1.3 Learning-based Methods

More recently, researchers have presented techniques for depth map estimation based on CNNs. Heber et al. [114] utilized CNNs to predict depth information for light field data learning an end-to-end mapping between the 4D light field and a representation of the corresponding 4D depth field in terms of 2D hyperplane orientations. With the development of sparse coding, Johannsen et al. [119] proposed a novel method for depth estimation in light fields that employed a specifically designed sparse decomposition to leverage the depth-orientation relationship in EPIs.

5.2 Depth Refinement

Due to noise, occlusions or inherent matching uncertainty caused by textureless regions, the initial depth map typically contains outliers, see Fig. 12(c). Therefore, a depth map refinement is crucial to smooth out the outliers and produce the final depth map. Typical methods are either within a Markov random field (MRF) framework or a variational framework.

MRF are widely used for problems that involve feature point matching [96], [97], [98], [99], [102]. The framework includes a unary term and a pairwise term. The unary term is first constructed using initial depth estimation approaches described above, then discretized to different depth values. Each unary term value reflects the probability of the corresponding depth. However, in textureless areas, the unary term is easily impacted by noise, making it inaccurate at estimating the depth of such areas. To address this, a pairwise term was added to refine noise points with their neighbors. The pairwise term is designed using the properties of the depth map, such as smoothness or color consistency. An energy function combines the unary term and the pairwise

term, and the depth map is estimated by minimizing it. Fig. 12(d) shows an example depth map refined with a MRF framework.

Although MRF-based refinement performs well, it results in *discrete* disparity values (see the background region in Fig. 12(d)). Consequently, this depth map refinement approach does not perform well for 3D reconstruction. To address the problem, Wanner and Goldluecke [66], [112] presented a continuous framework and described novel variational methods for disparity reconstruction. Their framework employed a total variation (TV) smoothing scheme. Based on their work, Heber et al. [120] proposed a model that combines the main idea of active wavefront sampling (AWS) with light fields. By tracking an image scene point over a sequence of sub-aperture images, they observed a virtual rotation of the scene point on the image plane, that was inversely related to the scene depth. To construct their model, they utilized an extension of the total generalized variation (TGV) approach.

5.3 Discussion

In this section, we have introduced depth estimation approaches based on light field data. Table 3 lists numerical evaluation values (MSE) of some representative approaches. These methods help to generate accurate depth maps based on light fields, but challenges remain. First, most methods are based on depth estimation of Lambertain surfaces; while for non-Lambertian ones, in which a change in viewing direction can imply a drastic change in appearance, traditional constraints and cues do not work so well. Second, the resulting depth map can be quantized, leading to errors on it. Third, due to the constraint on the angular resolution of the light field image, there exists a restriction for the depth range.

Compared with stereo pairs, a light field provides a more continuous disparity space due to a denser viewpoint sampling, i.e., higher angular resolution. Intuitively, a higher angular resolution of the light field benefits depth estimation. However, from the point of view of hardware design, higher angular resolution means denser sensor arrangement (in the case of multi-sensor capture), longer capture time (in the case of time-sequential capture), or lower spatial resolution (in the case of spatial multiplexing). Besides, dense

viewpoint sampling also requires accurate calibration tools [121] or mechanical controllers. In practice, researchers have found that the quality of the estimated depth can be greatly increased when using an angularly super-resolved light field [66]. The effectiveness can be more significant when the light field is super-resolved using a depth-free framework, such as learning-based methods [71], [72]. Conversely, a depth map of better quality, such as with better occlusion handling, produces better super-resolution results. One may model depth estimation and light field super-resolution in a joint optimization framework to obtain both a high quality depth map and a super-resolved light field.

6 LIGHT FIELD ASSESSMENT AND COMPRESSION

Light field quality assessment and light field compression are two critical aspects for the practical usage of light fields. Light field assessment is helpful to obtain a better understanding of the performance of light field acquisition, generation, and processing techniques. With regard to compression, a light field involves a large amount of data, but also records a scene with a set of images from different viewpoints, thus exhibiting data redundancy in both the spatial and angular dimensions [3]; examples of this redundancy are the smooth regions in each sub-aperture image (Fig. 2(a)) and light field subview (Fig. 2(b)). Therefore, the compression approaches for light fields offer much opportunity and are essential for light field storage, transmission and display. In this section, we introduce quality assessment approaches and compression schemes for light fields.

6.1 Quality Assessment

Light field quality assessment usually involves assessment of spatio-angular resolution, spatial quality, and angular consistency. Existing image resolution assessment methods adopt the ISO 12233 standard [122], which uses a slanted edge-based method to calculate the spatial frequency response. Fu et al. [123] showed that a light field camera exhibits more stabilized visual resolution in terms of depth of field (DOF) compared with a conventional camera due to the refocusing capability. For spatial quality assessment, peak signal to noise ratio (PSNR) and structure similarity index (SSIM) [124] are always applied. For instance, Shidanshidi et al. [125] applied PSNR to evaluate synthesized novel views using different ray interpolation approaches [3] with the corresponding simulated views. Angular consistency measures the visual coherence between views. Such criterion is always critical to the evaluation of light field processing approaches, such as angular super-resolution and light field editing. Current angular consistency assessment usually relies on subjective evaluations on an EPI or a pseudo sequence. Adhikarla et al. [126] presented an interactive light field viewing setup for the subjective evaluation of angular consistency. Besides, they extended SSIM [124] to a 3D context for light field-specific angular assessment, and evaluated the performance of existing quality assessment metrics.

6.2 Light Field Compression

In this subsection, we investigate lossy and lossless compression schemes for light fields; the lossy compression is further divided on the following taxonomy: progressive or transform coding, predictive coding and pseudo-sequence-based coding. Note that some techniques employ hybrid coding schemes; we classify these techniques according to their main aspect.

6.2.1 Lossy Compression

Transform coding approaches typically rely on a transform, such as the discrete cosine transform (DCT) or the discrete wavelet transform (DWT), to compress a light field. Classical coding schemes, such as JPEG (using DCT) and JPEG 2000 (using DWT), are applied to compress the light field raw image. However, these two schemes are not specifically designed for light field compression, and will thus not achieve optimal compression results. Aggoun [127] proposed a 3D-DCT to exploit the spatial redundancy within the light field subviews. Specifically, they first arranged a selected subview and its neighboring ones into a 3D brick, then the 3D-DCT was applied to produce the decorrelated subview group. Rather than 3D-DCT, Aggoun [128] further applied a 3D-DWT to a brick of sub-aperture images, and Magnor et al. [129] presented 4D-DWT to directly compress the 4D light field without arranging subviews or sub-aperture images.

Those approaches may produce misalignment between views due to large displacements between neighboring views, and the resulting high frequencies are expensive to code. To handle this problem, disparity compensation was applied to warp each view onto its view-dependent texture map. Xu et al. [130] first decomposed a light field into sub-bands using wavelet packet transform; the wavelet packet bases were divided into two groups: predictable bases and unpredictable bases. Finally, a disparity map, which was applied for sub-band prediction, was estimated using an optical flow-based approach. Chang et al. [131] incorporated disparity compensation into wavelet decomposition based on lifting. They first applied disparity-compensated wavelet decomposition to exploit similarity between different views in the light field, then the resulting coefficient images were encoded using shape-adaptive discrete wavelet transform.

Predictive coding approaches first select a set of images from the light field array to be coded as *Intra*, also known as I-images. Then the I-images serve as reference for coding the remaining light field images, also known as P-images. An early work proposed by Magnor and Girod [132] first separated each sub-aperture image into square blocks, and predicted the blocks in the P-images using the blocked disparity and blocks in the I-images. Similarly, Conti et al. [133] used blocked disparity for the coding of 3D light fields and 3D light field video. Based on Magnor and Girod's work, Kundu [134] used 2D homography to match P-images with I-images, and predicted P-images using homography-based warping and interpolation. Jiang et al. [135] further combine homography and low rank approximation, which are jointly optimized. By considering the similarity between adjacent views, Liu et al. [136] proposed a layered prediction scheme that predicts the P-images layer by layer, i.e., views in a lower layer were applied to predict views in a higher layer. For large compression rates, Li et al. [137] presented a coding scheme that uses a sparse set of subviews and associated disparities to compress the full light field. The
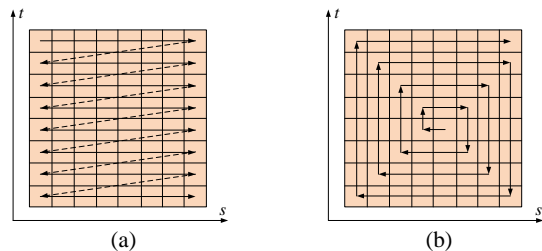
Fig. 14. Typical rearrangement paths for pseudo sequence-based coding approaches: (a) *zig-zag*; and (b) *spiral*.

coding scheme first applied the sparse set of subviews and disparities to reconstruct an intermediate light field, then utilized it to predict the original light field using a modified high efficiency video coding (HEVC) encoder. To improve coding efficiency, Jin et al. [138] employed a reshaping algorithm before the compression.

Pseudo sequence coding. A 4D light field can be considered as multiple frames with slightly different viewpoints. Based on this assumption, pseudo sequence-based coding approaches first rearrange light field elements (usually sub-aperture images) as a pseudo sequence, then use a video encoder to compress the sequence. Typical rearrange paths are *zig-zag* and *spiral* [139], [140], where the *zig-zag* scan path rearranges the views from left to right and top-down (as shown in Fig. 14(a)) and the *spiral* scan path begins from the center view and progresses from the inner to the outer ones spirally (as shown in Fig. 14(b)). For the compression, Dai et al. [139] and Vieira et al. [140] applied standard video encoders, such as H.264/AVC and HEVC, to compress the pseudo sequence. In their experiment results, the *spiral* scan path outperforms the *zig-zag* pattern due to a smoother rearrangement. Rather than using the scan paths described above, Li et al. [141] proposed a 2D hierarchical coding structure to reduce the reference buffer. They divided the views into four quadrants, and encoded one quadrant after another.

6.2.2 Lossless Compression

Perra [142] presented a lossless compression scheme on raw plenoptic data. They first divided the raw image into four color components and then encoded them separately using an adaptive prediction tool. Instead of encoding raw plenoptic data, Helin et al. [143], [144] implemented their compression scheme on rectified light field images, and thus, it was not limited to plenoptic data. By exploiting the high redundancy existing between the sub-aperture images, their lossless compression scheme first constructed consistent segmentations for the central view, and then propagated them to other views using a quantized depth map. An optimal sparse predictor was then applied to predict each region of the segmentations using the corresponding regions from the already coded views.

6.2.3 Discussion

A light field contains high dimensional data, posing a great challenge for practical usage, storage, transmission and display. Fortunately, light fields exhibit data redundancy in both the spatial and angular dimensions [3], which can be effectively removed using compression techniques. Some

events, such as JPEG Pleno activity [145] and ICME 2016 Grand Challenge on Light-Field Image Compression [146], were held to look for effective compression schemes. Current stat-of-the-art techniques can achieve more than 80% bitrate reduction [137] for lossy compression, and 60% bitrate reduction [143] for lossless compression. However, for practical usage of light fields, more progress is expected for light field compression schemes that aim at providing high compression quality with reasonable resource requirements in terms of data rates, computational complexity, and power consumption. Besides, in lossy compression schemes, the coding and decoding process for the residual will distort the structure of the input 4D light field to a certain extent. This structure distortion will definitely influence the subsequent light field processing, such as depth estimation or super-resolution. Thus, obtaining a decent bitrate reduction while preserving the light field structure is also a challenge for light field compression [147]. A more detailed evaluation of state-of-the-art light field compression approaches can be found in [148].

7 LIGHT FIELD APPLICATIONS

When the light field was first introduced to computer graphics in 1986 [151], researchers proposed only one application: novel view synthesis. However, as computing speed and computational resources increased, the light fields have grown in popularity in computer vision applications. In this section, we investigate various applications that benefit from light field data.

7.1 Synthetic Aperture Imaging

Isaksen et al. [36] first demonstrated the light field's ability to "see-through an occluder" by synthesizing a sufficiently large aperture. Because the object of interest and the occluder are at different depth layers and parallaxes, each view is able to capture a small portion of the object [152]. By propagating the portion of interest in each view, one can synthesize a virtual aperture image at a specific depth layer.

Levoy et al. [149] used synthetic apertures to reconstruct partially obscured objects (e.g., seen through a plant). A single camera was used to capture images from different viewpoints reflected by an array of planar mirrors. With an appropriate focal plane, a virtual view with a wide aperture and a shallow DOF was synthesized, where the object was in focus and occluders were blurred out (see Fig. 15(a)). Vaish et al. [153] studied four cost functions, including color medians, entropy, focus and stereo, for reconstructing occluded surfaces using a synthetic aperture. The cost functions, however, may fail under severe occlusions. Yang et al. [154] divided the scene into multiple visibility layers, and then propagated the visibility information among those layers using an optimization framework. Their approach was able to produce a synthetic aperture image with a wider DOF compared with previous works. Dansereau et al. [150] presented a volumetric focus approach that improves focus quality while maintaining sharp focus over a selected depth range using light fields (see Fig. 15(b)).

In addition, some studies for tracking objects through occlusions based on synthetic aperture techniques have been

| Captured light field | Synthetic aperture image | Refocus on a single plane | Refocus over a range of depths |

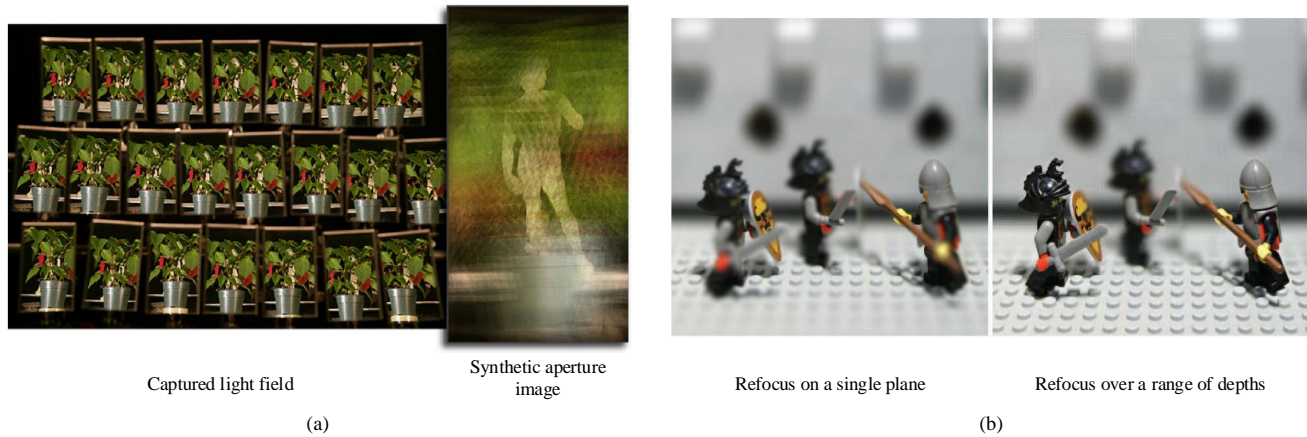(a)   (b)

Fig. 15. Applications using light fields. (a) The light field acquired by capturing an array of planar mirrors (left) was used to generate a synthetic aperture image (right) [149]. (b) A light field can be applied for post-capture refocusing on a single plane (left) or over a controllable range of depths (right) [150].

proposed [155], [156]. Occlusions are one of the obstacles for object tracking. Conventional single-camera tracking algorithms typically predict the motion parameters using filtering techniques such as Kalman filtering or particle filtering. However, the tracking tends to fail when the motion trajectory is unpredictable. Rather than tracking an object in a passive way, a camera array system (e.g., Joshi et al. [23] used 8 VGA video cameras) is able to provide clearer views of objects of interest using synthetic aperture techniques.

7.2 Segmentation and Matting

Typically, segmentation can be considered as a labeling problem. Unlike conventional segmentation approaches, such as co-segmentation and multiview segmentation, views in a light field are much more correlated (i.e., EPI volume), and thus can help in ensuring labeling consistency. However, due to the large number of views in a light field (usually a dozen [157] to a hundred views [158], [159]), segmentation and matting using light field data take abundant computing resources and running time during the energy minimization, which restricts their further promotion.

Campbell et al. presented multiple-view segmentation approaches using graph-cuts in voxel space [161] and image space [162]. However, the MRF framework is always time-consuming. To handle this problem, Hog et al. [163] proposed a graph structure to reduce the graph size of a light field. By taking advantage of structure information in the EPI space, Berent and Dragotti [164] modeled the segmentation problem as an energy minimization scheme. Wanner et al. [157] presented a variational framework for multi-label segmentation on the ray space of 4D light fields. Yücer et al. [158], [159] proposed a segmentation approach that uses unstructured 3D light fields captured by a hand-held video camera (see Fig. 16(a)). They computed the gradients in the light field volume and used local gradient filters to estimate the likelihood of edges belonging to the foreground. The likelihood was then propagated to the remaining areas in the volume. Finally, a Bayesian framework was employed to aggregate the segmentation likelihood of every image into the 3D volume. Based on the estimated depth, Zhu et al. [165] propagated the superpixel segmentation in a sub-aperture image to the whole light field to form a light

field superpixel segmentation. Also, light fields have been used to segment refractive surfaces or transparent objects into different superimposed layers. Johannsen et al. [166] used disparity and luminance information from a light field and formulated a linear generative model to segment the light field into individual layers, while Xu et al. [167] proposed a transparent object segmentation approach using the consistency and distortion properties of a light field.

To solve the matting problem using a light field, Joshi et al. [168] proposed an efficient algorithm for natural video matting using a camera array. By taking advantage of light field data, a synthetic aperture image focused on the foreground was computed to reduce the variance in the foreground. The variance measurement was then used to construct a trimap for the alpha matte. Cho et al. [169] built their framework on the EPI to minimize the linear composite error, which is subject to correspondence constraints. Fiss et al. [170] used light field matting to separate foreground occluders (e.g., dust on a window or a wire mesh) from the background target. They represented the 4D light field as a foreground layer composited over a background light field.

7.3 Detection, Classification and Recognition

Due to the inherent robustness to occlusions, light fields have an excellent performance on detection and recognition problems. Li et al. [171] demonstrated that light fields can benefit saliency detection. Their approach capitalized on the refocusing capability of light field imaging to measure likelihood scores of foreground and background. Compared with single-image saliency detection, their approach can more robustly handle challenging scenarios such as similar foreground and background.

Also, Pei et al. [172] used a camera array to synthesize a confocal view for multi-object detection. Ralph et al. [173], [174] introduced a novel face recognition approach using a light field system composed of 13 cameras. Raghavendra et al. [175] proposed a face detection approach from the best-focused image based on a wavelet energy analysis, and Raja et al. [176] used the best-focused image rendered by a light field camera for iris recognition.

Compared with object recognition, material recognition is a challenging problem because of the sheer diversity in

| Sub-aperture image (top) and EPI (bottom) | Segmentation result | Segmentation in 3D |

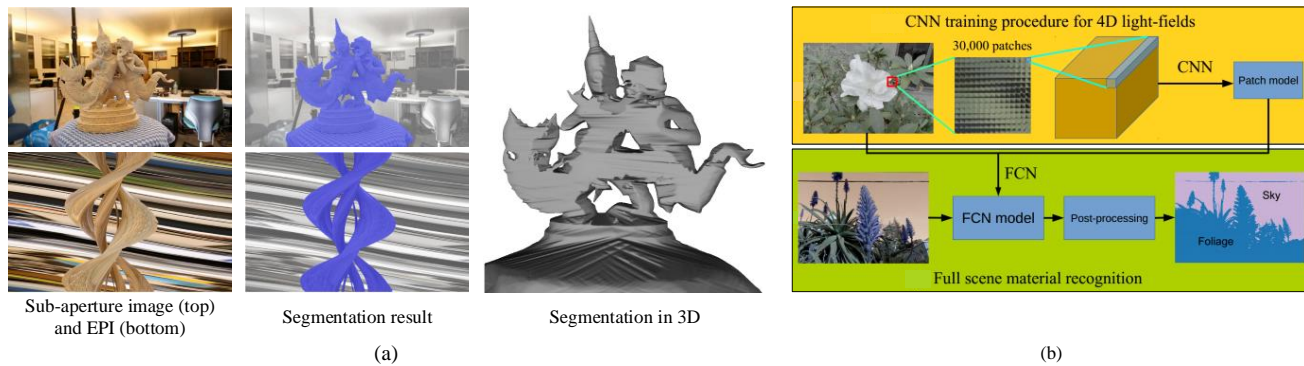(a)　　　　　　　　　　　　　　　　　　　　　　　　　　(b)

Fig. 16. Further examples of applications using light fields. (a) Object segmentation proposed by Yücer et al. [158], [159], where the likelihood of edges belonging to the foreground was estimated using the gradients in the 3D light field volume. The left column shows a sub-aperture image and an EPI of the light field, the middle column shows the segmentation results for them, and the right column shows the 3D segmentation result. (b) Material recognition proposed by Wang et al. [160], where a CNN was trained on a 4D light field dataset.

the appearance of materials. Using the recently developed learning techniques, Wang et al. [160] proposed a CNN-based material recognition method that used a 4D light field dataset (see Fig. 16(b)). They also presented several CNN architectures for the 4D light field inputs as well as a light field dataset for material recognition. Compared with conventional single (2D) image material recognition approaches, their approach achieved approximately a $6-7\%$ boost.

7.4 Handling Refractive and Transparent Object

Light field-based detection and reconstruction of non-Lambertian materials [177], [178], such as refractive and transparent objects, has been studied, a problem that is extremely ill-posed when only a single image is considered. Maeno et al. [179] presented a transparent object recognition approach using a light field camera. They observed that a transparent object deforms background scenes due to refraction, causing ray distortions in the 4D light field. They named this observation the "light field distortion feature", and used it to detect transparent objects. Anglin et al. [180] proposed a shift-invariant refocusing transformation for volumetric estimation, which can be further applied for reconstruction of special materials, such as fluid and flame. Ding et al. [181] used the distortion of a known pattern caused by refraction to reconstruct a fluid's surface. They applied a camera array to track distorted feature points over time and across cameras and built spatio-temporal correspondence maps for the fluid surface reconstruction. Sulc et al. [182] used a 4D light field to separate a dichromatic reflection component from diffuse object colors. They model the observed light field as a sparse linear combination of a constant-color specular term and a small finite set of albedos. Lu et al. [183] considered light field sampling as sampling a 4D BRDF, and presented a CNN-based architecture for BRDF identification.

7.5 Visual Odometry, Localization and SfM

Dansereau et al. [184] presented a six-degree-of-freedom visual odometry approach that used a camera array to estimate the pointwise plenoptic flow. Their approach achieved a higher performance than conventional approaches (e.g., stereo feature-tracking) in terms of both robustness and

tracking accuracy. Instead of performing stereo matching between images, Zeller et al. [185] presented a visual odometry approach that is able to find stereo correspondences in full sensor resolution, avoiding aliasing effects due to spatial undersampling. Schillebeeckx and Pless [186] used light fields to simplify the estimation of camera rotation. They applied a lenticular sheet with a color-coded backplane pattern to create a light field where the observed color depends on the direction of the light. Srinivasan et al. [187] developed an oriented 4D light field window for robust and accurate pixel comparisons, and applied it to scene flow recovery. Fahringer et al. [188] further reduced complexity necessary to acquire volumetric data and used the 3D flow for velocity field estimation.

7.6 Stitching, Deblurring and Stabilization

Guo et al. [189] proposed a multiple light fields stitching algorithm to produce a seamless navigation using a light rays transformation matrix, named "ray-space motion matrix". Similarly, Johannsen et al. [190] used ray-to-ray correspondence to produce a light field panorama. Srinivasan et al. [191] took advantage of the 3D capture ability of the light field camera, introducing a general model for 3D camera motion estimation and deblurring. The model produced higher quality light field deblurring results when compared with conventional 2D image deblurring, which typically assumes the blur kernel to be a 2D function. Smith et al. [192] described a video stabilization approach using an integrated camera array, ProFUSION-25C [21]. They computed a sequence of relative poses between the virtual camera and the camera array by employing spatio-temporal optimization, and produced synthesized views of these poses.

7.7 Discussion

Due to the particular structure and other specific properties of light fields, there is a growing number of applications in both computer vision and computer graphics that use them to try to solve long-standing problems. By taking advantage of spatio-temporal properties, we can imagine the immense potential of light field video. More related applications, such as tracking and motion capture, can be further exploited. Besides, the integration with more novel imaging techniques is expected to bring new capabilities to light field applications;

for instance, by combining light fields with hyperspectral imaging, the extraction of the material properties of the observed object can be greatly enhanced.

8 LIGHT FIELD EDITING

Editing light fields has in general been posed as a generalization of traditional 2D image editing. However, given the richer information encoded in the light field, this generalization cannot be considered trivial; as opposed to 2D images, light fields include complex directionally-varying effects such as glossy reflections, refractive objects or occlusions. These problems exist when editing other higher-dimensional media as well, such as video or image collections. We focus on works target to edit the most common structured light fields: We categorize these works on local and global edits, depending on whether they operate on local, precisely placed edits, or on the full light field. Later, we briefly discuss generalization to non-structured light fields and its relation with other types of media.

8.1 Local Editing

Local point-based editing. One of the first methods for light field editing was introduced by Seitz and Kutulakos [193], [194], based on creating a voxel-based reconstruction of the geometry encoded by the light field. This representation allowed the authors to perform operations such as painting or scissoring. However, the editing is highly dependent on the quality of the voxelization, imposing a Lambertian assumption on the surfaces, which limits its applicability to complex appearances (Fig. 17(a)). Zhang et al. [83] proposed a system for performing complex edits in the light field such as erasing, copy-pasting at different depths or cloning, based on patch-based methods. By using a precomputed depth map, the system was able to propagate the edits performed in a single view to the rest of the light field using patch-based synthesis.
Local deformations. Chen et al. [197] demonstrated a technique for deforming a light field in the context of photorealistic image-based rendering, based on splitting the light field into multiple sub-light fields. These were later deformed independently, and then rendered together to keep consistency. To preserve illumination consistency on the deformation, the system is limited to co-axial illumination. Birklbauer et al. [198] proposed a method to perform nonlinear edits on the spatial domain of the light field, keeping consistency on the angular domain. This allowed complex deformations on the spatial domain of the light field, while preserving light field capabilities such as refocusing.
Editing interaction. The previous works focus on specific editing tools, that require precise positioning of the edit in the three-dimensional scene implicitly encoded by a light field. Jarabo et al. [196] analyzed how novice users edited light fields, focusing on the two main positioning paradigms in the literature. These two paradigms used either focus or parallax to specify the position in depth within the light field. Later, Masia et al. [199] and Ortin et al. [200] further analyzed the workflows of the users during editing. An example edit from [196] can be seen in Fig. 17(c). These works analyzed the interaction using standard 2D displays,

although using more complex prototype light field displays (see Sec. 9) has been proposed, allowing light field editing by means of gesture tracking [201] or a 3D-positioned light pen [202].

8.2 Global Editing

Global edits are the most wide-spread form of light field editing, and they have been incorporated in commercial software, usually based on applying some predefined filters (e.g., *Lytro Desktop* [4] or *Lightfield Iris* [203]).

Horn and Chen [204] proposed a shader-based language to manipulate light fields, enabling light field compositing, or complex ray-warping with applications in non-physical defocus. Cossairt et al. [205] presented a novel image-based method for compositing real and synthetic objects in the same scene using two 4D light fields, one captured from and one projected onto the real scene. Zhang et al. [195] and follow-up work by Wang et al. [206] proposed a method for morphing two light fields, based on user-defined correspondences between both light fields (Fig. 17(b)).

Birklbauer and Bimber [207] proposed an approach for light field retargeting, allowing compression or expansion of the spatial domain, while preserving angular consistency without the need of explicit depth estimation. Garces et al. [208] recently proposed a method for obtaining the intrinsic decomposition of light fields into albedo and shading. As opposed to other multidimensional intrinsic decomposition methods (e.g. [209], [210]), they leveraged the coherence and structure of the light field.

In the context of coarse edit propagation, Jarabo et al. [211] extended *AppProp* [212] to the 4 domain of light fields. To efficiently handle the large scale of light fields, they proposed to operate on a downsampled light field, and then upsample the results following [213]. Several works have followed a similar approach, reparametrizing the light field to increase the locality of the propagation metric [214], or increasing the angular consistency on the propagation distance metric to improve spatio-temporal coherence [215]. A different approach is that based on filtering, by which the editing operation is posed as a multi-dimensional filtering operation over the light field. This is the approach taken by the work of Gryaditskaya et al. [216] for gloss editing in light fields.

8.3 Beyond Structured Light Fields

The works described so far target structured light fields, which have the nice property of a well-defined structure that can be leveraged for robust correspondence estimation for transfer edits (see Sec. 5). This structure is the key difference with other methods targeting non-structured media, such as unstructured light fields [217], [218], video [209], [219], [220], [221] or image collections [222], [223], [224]. These are more complex structures, and might require the use of techniques such as depth estimation, optical flow or feature matching, to transfer edits between images.

8.4 Discussion

While the first works on light field editing date from the late 90's, it has been recently when it has become a more active
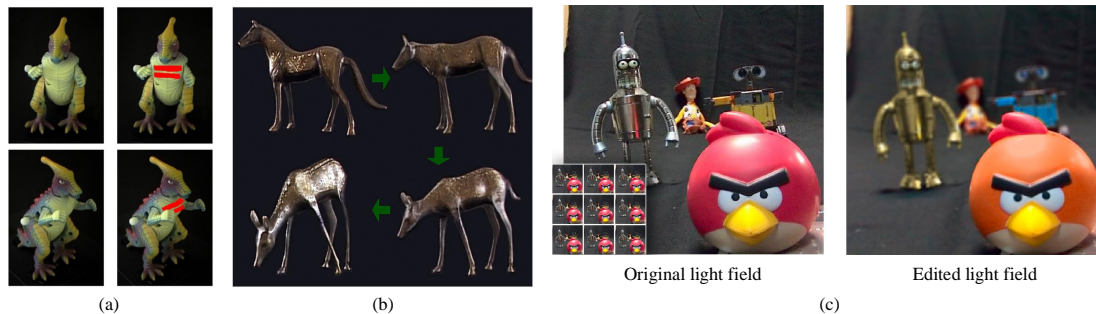
Fig. 17. Example edits on light fields. (a) Seitz and Kutulakos [193], [194] used a voxel-based representation of the light field, allowing drawing and scissoring. (b) Light field morphing between two input light fields [195]. (c) Sample light fields captured with the Lytro camera, edited using the different paradigms analyzed by Jarabo et al. [196].

field with the widespread of plenoptic imaging. In this context, most current works aim to transfer well-established editing tools in traditional 2D images to (structured) light fields (e.g. filters, edit propagation, retargeting or intrinsic decomposition). These works focus on overcoming the intrinsic difficulties of the additional angular domain, instead of leveraging the novel capabilities brought by light fields. This type of editing approaches also often rely on depth information. We have seen how depth from light fields is an active research area (Sec. 5), but a critical question that remains to be answered is how precise this recovered depth information needs to be in order for editing approaches to work well. Of course, this depends on the approach, but should be an aspect addressed by the novel editing approaches to come that rely on depth information.

This research direction is important, and together with principled analyses on the interaction with light fields, should allow for a seamless transition from 2D to 4D images. However, the gained complexity of light fields is still underused in terms of editing capabilities. Making use of all the encoded additional information to enable novel realistic edits on the captured scenes, such as high-quality relighting or material editing, or to increase the expressibility of the artistic editing operations on the final images is a complementary avenue for future research, which would further differentiate between traditional and plenoptic images.

9 ALGORITHMS FOR LIGHT FIELD DISPLAY

Displays capable of showing light field contents have been around since the beginning of the 20th Century [225], [226]. Since then, and especially in the last two decades, different technologies exist capable of displaying light fields. In terms of viewing experience, and in comparison to a conventional 2D display or 3D (stereoscopic) display, a light field display adds the ability to see the scene from different viewpoints. This entails that, in light field displays, the perception of depth is conveyed not only through binocular disparity—as in stereoscopic displays—and monocular cues, but also through motion parallax thanks to the different views; this can be done without the need to wear glasses or other hardware. To do this, they must display different views of the light field to both eyes and for each viewpoint within the FOV of the display.

The relevant work in terms of light field processing for display revolves around two main areas. The first is novel display architectures that require more or less heavy

processing of the light field contents prior to its display. The second, a number of algorithms for content preprocessing devoted to improving the visual quality of the displayed imagery, attempting to overcome the limitations that the hardware architecture imposes. In the following we review both areas.

9.1 Light Field Display Architectures

Traditional light field displays. Traditionally, light field displays rely on parallax barriers or lenslet arrays to display these different views to different eyes and viewpoints, e.g., [227]. A back panel displays the different images of the light field, while a front layer redirects light rays appropriately to different positions in front of the screen. This front layer can be formed by regularly-spaced slits which act as a mask in the case of a parallax barrier, or by a series of lenses redirecting rays in the case of a lenslet array-based display (also termed integral imaging); see Fig. 18(a) and 18(b). In any case, most of these systems require simple image processing, such as some sort of interlacing of the light field views before presentation (see, e.g., [228]) and are out of the scope of this survey. We refer the interested reader to comprehensive existing surveys [229], [230], [231], [232].

The main drawbacks of these systems are: a low light throughput (in the case of parallax barriers), a trade-off between spatial and angular resolution, and a limited FOV. Additionally, crosstalk between views, inducing ghosting artifacts, is a common problem. In order to alleviate some of these problems, a number of systems have been developed that employ head or eye tracking and actively steer the views accordingly [233], [234], [235]. However, these approaches require a significant engineering and calibration effort, and typically feature complex systems and/or additional gear. Therefore, more recently, researchers have looked for alternative display architectures, typically at the cost of a heavier processing of the content prior to display.

Multilayer light field displays. The majority of subsequent approaches to displaying light fields have relied on multilayer architectures of one kind or another. A light field is a 4D structure, while a conventional display is 2D. A way of addressing this mismatch is to stack several layers, such as several LCDs. The main goal of the initial multilayer displays was to provide a wider DOF [236]. Further, having a multilayer setup allows to improve accommodation cues, in an attempt to avoid the possible mismatch between focus (accommodation) cues of the eyes, and other depth cues

employed by our visual system to infer depth, such as vergence; this mismatch can cause discomfort in the viewer.

In these multilayer systems, an algorithm is required to distribute the input light field to the different layers of the display. Often, the distribution is done such that each ray is assigned to the depth layer closest to the ray origin [237]; this of course requires knowledge of scene depth. A more sophisticated approach was presented by Ranieri et al. [238], based on decomposing the light field into a series of components which are then rendered to the different display primitives: essentially, different parts, such as diffuse, occlusions, glossy areas, are decomposed and rendered differently. Later works achieved correct accommodation in multi-plane displays [239]; in them, the patterns for the different layers are obtained via primal-dual optimization in the frequency domain. All these approaches require an increase in computation and processing with respect to the traditional ones, but in turn allow to reduce some of their drawbacks. In the following we focus on a family of light field displays which also feature a multilayer architecture, but with a distribution among layers based on a low-rank light field decomposition. This enables them to improve spatial resolution, light throughput and contrast, at the cost of increased processing.

Compressive light field displays. The so-called *compressive* light field displays are multilayer displays consisting of stacked LCD panels (or attenuating layers if content is static) and backlighting (Fig. 18(c)). They can also feature temporal multiplexing of the contents of the different layers (then termed multilayer *multiframe* displays). Prior to display, the content of each of these layers has to be determined, and this is the relevant part in terms of light field processing: the most general formulation can be found in the work on tensor displays [240]; this work aims to provide a general framework for multilayer multiframe displays, encompassing a number of variants which had been appearing before.

Lanman et al. [241] showed that a low rank approximation of the light field could be used; this, combined with temporal multiplexing, allowed for display of light fields using a dual-layer display. Wetzstein et al. [242] posed the problem as a tomographic reconstruction of the light field. In both cases, displaying the light field requires solving a non-negative least squares problem which yields the optimal attenuation maps to be displayed in the different layers. A generalization to multiple layers using directional backlighting was presented later [240]. The inclusion of time multiplexing [243], and the support of near-correct accommodation cues [244] followed. These displays recreate the desired scene achieving higher resolution, larger DOF, and higher light throughput than previous approaches, at the cost of significant computation for prior processing of the input light field. In order to reduce the cost of this processing, Heide et al. [245] presented an approach that combines adaptive sampling and knowledge of display-specific characteristics to synthesize light fields for compressive displays at a much lower consumption of resources. On a related note, there have also been approaches devoted to fast rendering of light fields from the 3D description of the scene, based on exploiting the redundancy of the light field reusing radiance evaluation between pixels [246].

9.2 Light Field Display Content Preprocessing

In addition to processing strictly required for display on the different hardware systems, a number of works have strived to improve the quality of the imagery displayed by performing some preprocessing of the content to be shown. These software solutions essentially focus on addressing two of the main limitations that we have mentioned: crosstalk between views; and reduced resolution and the subsequent aliasing, which is particularly severe in the angular domain, in which resolution is most reduced. Additionally, there is a body of work concerned with retargeting of light field content for its display on devices of different characteristics.

Crosstalk. In both parallax barriers and lenslet arrays, light leakage occurs between adjacent views. There is a variety of causes of crosstalk, ranging from imperfect alignment of components to quality of the optical elements, and including the tilted nature of many parallax barriers and lenslet arrays [247]. Crosstalk not only severely diminishes image quality, but can also cause perceived depth to be incorrect [248], [249]. Consequently, a number of works have appeared that are devoted to minimizing the effects of crosstalk in these displays; we focus here on those which imply preprocessing of the input (light field) content. Most approaches incorporate some model of the crosstalk and manipulate the luminance of the input images so that, after being shown on the display with crosstalk, the resulting image is as similar as possible to the target one. To model the crosstalk, careful calibration is often required. Jain and Konrad simulated the crosstalk as a blurring filter, and Wiener deconvolution was employed to preprocess the images to minimize the effect of the crosstalk [250]; the effect of angular antialiasing filters was also taken into account in the process. A linear model of crosstalk was also employed by Wang et al. [251], thus modeling the process as a linear system of equations. In their case, the crosstalk coefficient matrix of the model was not measured experimentally, but derived from screen characteristics instead. They further proposed a method to solve the system posing it as a box-constrained integer least squares problem. Li et al. [252] focus on light leakage among subpixels: they also rely on inverse filtering for vertical neighboring subpixels, while horizontally blending weights for the neighboring subpixels are computed based on minimizing the mean square error. A different approach is that of Pei et al. [253], who propose a method to reduce crosstalk via light field decomposition: they propose to backproject rays to the LCD plane, during which multiple rays can be mapped to one LCD subpixel, these are weighted and the value used to compute the final emitted intensity.

Aliasing. A classical problem of conventional light field displays is spatial and angular *aliasing*. The former has been addressed in seminal works on light field sampling [36], [254]. The latter was tackled by works [255], [256] that derived theoretical formulations of the angular aliasing problem for different cases, and proposed antialiasing techniques based on prefiltering of the content. Some works have also looked into the generation of multiview content from stereoscopic video taking into consideration the angular aliasing problem and the limited depth range of the content that it

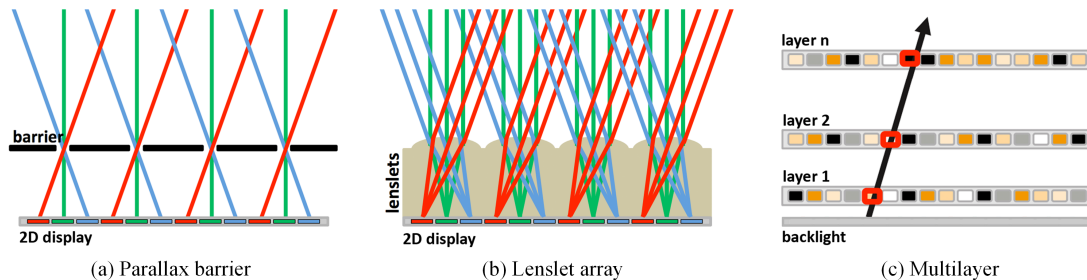| (a) Parallax barrier | (b) Lenslet array | (c) Multilayer |

Fig. 18. Three light field display architectures, featuring different ways of creating the final light field rendition. (a) Parallax-barrier based display, in which a mask blocks certain rays from the back layer LCD to show the different views from different directions. (b) Integral imaging display, where a lenslet array is used instead of the mask in (a), also redirecting rays but allowing for a higher light throughput. (c) Example architecture of a multilayer display (specifically, the one presented in [242]); light rays from the backlight go through different attenuation layers, forming the final displayed images. Images adapted from [240].

imposes [88], [257][1]. Image quality offered by these displays is also highly affected by the viewer's position; this was addressed by Du et al. [258], who proposed shearing (both globally and locally) the light field, followed by stitching, to hide transitions, improve continuity, and reduce hot-spotting issues. Building on this, a method to eliminate the depth-reversal that takes place in certain viewpoints based on adjusting light field alignment has recently been presented [259]. The problem of hot-spotting has also been addressed by tracking viewers eye positions and dynamically assigning appropriate views to them [260].

Retargeting. Retargeting refers to the process of modifying content to adapt it to the characteristics of a specific display [261]. In the most common case the concept refers to spatial retargeting. The first approach for light fields was that of Birklbauer and Bimber [207], in which they could expand or compress a light field without the need of depth information by applying seam carving on the z-stack ("an image stack that contains only the in-focus contributions in each of its slices while the out-of-focus contributions are removed"). Additionally, a large number of works have addressed the problem of *depth* retargeting. The reason is that the maximum allowed depth budget of a display is limited by the problem of angular aliasing. A traditional solution to angular aliasing, as mentioned before, is pre filtering the content. However, filtering for angular aliasing creates a depth-of-field effect by which only objects close to the screen plane in depth can be shown in focus, while in objects further away only the lower frequencies are preserved and thus appear blurred. To avoid this, depth is retargeted to a narrower depth budget, which essentially requires a subsequent warping of the light field according to the new depth that can be done using techniques targeted to stereo content [262], [263]. Depth retargeting techniques are extensive and not the main focus of this survey; we thus only gloss over them briefly: early works explore simple depth remapping functions [264], [265], while others rely on perceptual considerations [266], [267], [268], or devise a means of cutting through the EPI volume generating views with multiple centers of projection [269]. A joint image and depth retargeting based on seam carving for light fields has also been proposed [270].

---

1. For works tackling the generation of multiview content from a lower dimensional or subsampled input in the general case please refer to Sec. 4.

9.3 Discussion

If nowadays stereoscopic displays, aiming at displaying an image plus depth, still suffer from a number of well-known limitations, it is not hard to imagine that light field displays, aiming at displaying a full (4D) light field, suffer even more limitations. When we are attempting to display a 4D structure of visual information on a screen, compromises need to be made. Researchers have resorted to multilayer and multiframe architectures to alleviate the problem. There are two main problems with the current trends. First, the large computational cost is required to calculate the contents of each layer and frame. A light field entails a very large amount of information, and solving the optimizations required for the multilayer decomposition and streaming this information in real time is a challenge. Still, real time implementations exist and this problem may find a solution as computational capability increases and algorithms improve. Second, current displays still fail short of recreating a faithful viewing experience. Before this technology is ready for the market, we need to push farther to solve the limitations we have explained in this section. We need to increase more the FOV, to solve the mismatch in visual cues (e.g., lack of proper accommodation cues), improve energy consumption (increase light efficiency), spatio-angular resolution, and overall image quality. For this purpose, Kovács et al. [271] presented a quantitative evaluation for light field displays in terms of spatial and angular resolution. With an industry moving away from stereo displays (in favor of increased spatial resolution or color and luminance fidelity, and with the exception of HMDs), light fields may have the chance of bringing 3D back. However, for this to happen we still need significant improvements in acquisition, compression, and display algorithms for light fields. We have discussed many breakthroughs in these areas in the present survey. Among these, of particular importance for displays are the acquisition or generation of light fields with high enough spatial and angular resolution (Secs. 3 and 4). Both are needed for a faithful viewing experience, but angular resolution is critical to provide accommodation; it is still unclear and an active research topic how high this resolution needs to be in order to provide correct accommodative cues in the viewer. Content processing for certain types of displays also requires depth information (e.g., some multilayer approaches presented in Sec. 9.1, or retargeting techniques in Sec. 9.2); although many are demonstrated with synthetic content featuring perfect depth, techniques such as those in

Sec. 5 are needed for these approaches to be practical and usable with real footage.

As a side note, with the recent rise in popularity of virtual reality and augmented reality, partly driven by the reduced form factor and cost of current head-mounted displays (HMDs), a new application area is opening up for light fields. On the one hand, they are gaining importance for content generation for these type of devices. On the other, some recent works have used light fields to correct spatial calibration of optical see-through HMDs [272], [273]. In these works, the authors propose a light field-based method to compensate for spatial distortion of images and blur caused by optical see-through HMD optics: the distortion is modeled as a 4D to 4D mapping between light fields, and in the case of blur a PSF model is added (the blur in HMDs is view-dependent).

Finally, it has been shown that light fields and holograms can be seen as complementary representations, and transformations between them have been derived [117].

10 Public Datasets and Code

This section introduces several open source datasets and software tools for light field imaging and manipulation. While this list is not exhaustive, it provides a set of useful links for light field research.

Several light field datasets that include synthetic, real-world and microscopy light fields have been made public. The (New) Stanford light field archive [24] contains thirteen high spatial and angular resolution light fields acquired with a Lego gantry, seven microscope light fields, four light fields acquired with a gantry and two light fields acquired with a camera array. The Stanford Lytro Light Field Archive [274] contains abundant light fields acquired with a Lytro light field camera [4], and is divided into nine different categories. Reřábek and Ebrahimi [275] also provided 118 light fields captured with a Lytro camera [4]. The dataset is divided into ten categories, and can be used for light field compression and quality evaluation. The (old) HCI light field dataset [74] contains synthetic light fields and light fields captured using a Raytrix camera [5]; each synthetic light field includes ground truth depth maps for every viewpoint. Recently, Honauer et al. [276] presented a light field benchmark that provides a new HCI light field dataset and an evaluation of depth estimation techniques. Wetzstein [277] provided a number of synthetic light fields containing either 5×5 or 7×7 views. Kim et al. [25] provided high spatial and angular resolution light field datasets captured by mounting a consumer DSLR camera on a motorized linear stage. Jarabo et al. [278] provided the light fields used in [196], including both synthetic light fields and light fields captured with a Lytro camera. Finally, Ramamoorthi [279] shared a dataset for material recognition using light fields [160].

In terms of available open-sourced software, Dansereau [35] provided a set of tools that allow decoding, calibration, rectification, color correction, basic filtering and visualization of light fields. Wanner et al. [66] provided code for light field analysis in terms of both depth estimation and light field super-resolution. Tao et al. [102] provided their code for depth estimation that combined defocus cues and

correspondence cues. Wang et al. [99] and Jeon et al. [98] shared their code for light field depth estimation. Kalantari et al. [85] shared the source code for learning-based light field view synthesis. Finally, Jarabo et al. [196] made public their interface for light field editing.

11 Conclusions

As an alternative way to capture the visual information in a scene, light fields have become increasingly popular due to their capabilities in the representation of visual information, and the subsequent applications. Dimensionality reduction from the seven dimensions of the plenoptic function to a 4D representation promoted light fields beyond a theoretical concept. From a hardware perspective, we have witnessed the development of light field acquisition approaches from expensive and bulky camera array systems to portable consumer cameras, such as Raytrix [5], Lytro [4] and PiCam [22]. When light fields were first introduced to computer graphics [3], [11], their applications were only digital refocusing and rendering of virtual views. Today, however, the ability to expeditely capture abundant datasets enables numerous applications, including image enhancement, segmentation, detection, recognition, depth sensing, 3D reconstruction and 3D display.

Meanwhile, there are still many scientific challenges for researchers to address. The inherent resolution trade-off between spatial, angular and temporal resolution is one of the key issues. Although recent works have mitigated the problem by employing spatial, angular or temporal super-resolution techniques, obtaining a high spatial and angular resolution light field with high efficiency still requires more research efforts. Moreover, compared with active depth acquisition approaches, such as structured light and time-of-flight imaging, light field imaging suffers from several problems, such as low data efficiency and high computing complexity. In addition, current consumer plenoptic cameras can deal with only a very limited DOF due to their narrow baseline. Therefore, it is critical to design algorithms that are capable of accurate and efficient depth recovery from light fields with narrow baselines.

In the near future, light field capture based on compact multi-camera arrays will be more and more popular given its high efficiency in getting high resolution data in the spatial, temporal and angular dimensions. Heterogeneous or hybrid camera arrays with different settings of cameras will enrich the diversity of collected information, reduce the data size, and improve the performance of various computer vision algorithms. The pursuit of high performance imaging capabilities will drive novel light field acquisition approaches, such as the estimation of surface or material properties of the scene by combining light fields with hyperspectral imaging. The developments of light field compression and streaming technologies will increase the popularity of light fields in the multimedia industry. The ability of snapshot axial dimension imaging using light field microscopes opens a new door for biology research. Moreover, light field displays are an inevitable trend in VR/AR display systems. The demand of wider FOV, larger DOF and higher dynamic range photography may push research on light field imaging toward a more sophisticated prospect,

such as Lytro Cinema [280]. Such high performance light field capture at a single position will help to obtain a high quality reconstruction for a dynamic scene [281] (including geometry, texture, illumination and motion reconstruction), and further visual scene understanding. By keeping these trends in mind it is easy to imagine how light fields may become a part of our daily life in the future.

This paper has summarized the development of light field imaging, in terms of both the processing methods involved, and the challenges that lie ahead. We hope it will help on the evolution of this technology, promoting the development of new exciting applications and changing the way in which we are able to "see" and "understand" the world.

ACKNOWLEDGMENTS

REFERENCES

[1] G. Arun, "The light field," *Journal of Mathematics and Physics*, pp. 51–151, 1936.

[2] H. Adelson, Edward and R. Bergen, James, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, 1991.

[3] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH*, 1996, pp. 31–42.

[4] "Lytro." https://www.lytro.com/.

[5] "RayTrix. 3D light field camera technology." http://www.raytrix.de/.

[6] "Pelican Imaging." http://www.pelicanimaging.com.

[7] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.

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

[9] I. Ihrke, J. Restrepo, and L. Mignard-Debise, "Light fields and computational imaging: Briefly revisiting 25 years of research," *Signal Processing Magazine*, vol. 33, no. 5, pp. 59–69, 2016.

[10] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational plenoptic imaging," in *Computer Graphics Forum*, vol. 30, no. 8, 2011, pp. 2397–2426.

[11] C. Michael, G. Steven, J., S. Richard, G. Radek, and S. Rick, "The lumigraph," in *SIGGRAPH*, 1996.

[12] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, 1987.

[13] S. E. Chen, "Quicktime vr-an image-based approach to virtual environment navigation," in *SIGGRAPH*. ACM, 1995, pp. 29–38.

[14] M. Leonard and B. Gary, "Plenoptic modeling: An image-based rendering system," in *SIGGRAPH*, pp. 39–46), year=1995,.

[15] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *SIGGRAPH*, 2000, pp. 307–318.

[16] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, no. 2, pp. 121–138, 2004.

[17] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," in *Siggraph*, 2001, pp. 425–432.

[18] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," in *Eurographics Workshop on Rendering*, 2002.

[19] C. Zhang and T. Chen, "A self-reconfigurable camera array," in *Eurographics Symposium on Rendering*, 2004.

[20] 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," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 765–776.

[21] "ViewPLUS Inc. ProFUSION 25 5x5 camera array system." http://www.viewplus.co.jp/products/profusion25/index-e.html.

[22] 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," in *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6. ACM, 2013, pp. 2504–2507.

[23] X. Lin, J. Wu, G. Zheng, and Q. Dai, "Camera array based light field microscopy," *Biomedical Optics Express*, vol. 6, no. 9, pp. 3179–3189, 2015.

[24] "Stanford (New) Light Field Archive." http://lightfield.stanford.edu/lfs.html.

[25] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 32, no. 4, 2013.

[26] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. Chen, Homer, "Programmable aperture photography: Multiplexed light field acquisition," *ACM Transactions on Graphics*, vol. 27, no. 3, 2008.

[27] Y. Taguchi, A. Agrawal, S. Ramalingam, and A. Veeraraghavan, "Axial light field for curved mirrors: Reflect your perspective, widen your view," in *IEEE CVPR*, 2010.

[28] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Stanford University Computer Science Tech Report*, 2005.

[29] D. Lanman, D. Crispell, M. Wachs, and G. Taubin, "Spherical catadioptric arrays: Construction, multi-view geometry, and calibration," in *International Symposium on 3d Data Processing*, 2006.

[30] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and Chintan, "Spatio-angular resolution tradeoffs in integral photography," in *Eurographics Symposium on Rendering*, 2006.

[31] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H. 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, pp. 47:1–47:14, 2013.

[32] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," in *SIGGRAPH*, 2007.

[33] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 32, no. 4, pp. 1–11, 2013.

[34] N. Antipa, S. Necula, R. Ng, and L. Waller, "Single-shot diffuser-encoded light field imaging," in *IEEE International Conference on Computational Photography*, 2016.

[35] D. Dansereau, "Light-Field Toolbox for Matlab." http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4.

[36] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proc. Computer graphics and interactive techniques*, 2000, pp. 297–306.

[37] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan, "A new reconstruction filter for undersampled light fields," in *Eurographics Workshop on Rendering*, 2003, pp. 150–156.

[38] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion, "A frequency analysis of light transport," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 1115–1126.

[39] C. Soler, K. Subr, F. Durand, N. Holzschuch, and F. Sillion, "Fourier depth of field," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 2, p. 18, 2009.

[40] T. Amalia, *Linear Ray and Wave Optics in Phase Space*. Elsevier, 2005.

[41] Z. Zhang and M. Levoy, "Wigner distributions and how they relate to the light field," in *IEEE ICCP*, 2009.

[42] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-d de-

convolution for the light field microscope," *Opt. Express*, vol. 21, no. 21, pp. 25 418–25 439, 2013.

[43] R. Ng, "Fourier slice photography," in *SIGGRAPH*, 2005.

[44] I. Ihrke, G. Wetzstein, and W. Heidrich, "A theory of plenoptic multiplexing," in *IEEE CVPR*, 2010.

[45] B. Wilburn, M. Smulski, K. Lee, and M. A. Horowitz, "The light field video camera," in *SPIE Electronic Imaging*, 2002.

[46] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High speed video using a dense array of cameras," in *IEEE CVPR*, 2004.

[47] S.-C. Chan, K.-T. Ng, Z.-F. Gan, K.-L. Chan, and H.-Y. Shum, "The plenoptic video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1650–1659, 2005.

[48] Y. Liu, Q. Dai, and W. Xu, "A real time interactive dynamic light field transmission system," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2173–2176.

[49] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec, "Capturing and rendering with incident light fields," in *Eurographics Workshop on Rendering Techniques*, vol. 27, no. 4, 2003, pp. 1293–1301.

[50] J. F. G. W. Donald G. Dansereau, Glenn Schuster, "A wide-field-of-view monocentric light field camera," in *IEEE CVPR*, 2017.

[51] I. Ihrke, T. Stich, H. Gottschlich, M. Magnor, and H.-P. Seidel, "Fast incident light field acquisition and rendering," *Journal of WSCG*, Feb 2008.

[52] G. Lippmann, "La photographie intégrale," *Académie des Sciences*, vol. 146, pp. 446–551, 1908.

[53] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE TPAMI*, vol. 14, no. 14, pp. 99–106, 1992.

[54] L. Wei, C. 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. 152:1–152:11, 2015.

[55] G. Todor and L. Andrew, "Superresolution with plenoptic camera 2.0," in *Adobe Systems Inc., Tech. Rep*, 2009.

[56] R. Horstmeyer, G. Euliss, R. Athale, and M. Levoy, "Flexible multimodal camera using a light field architecture," in *IEEE ICCP*, 2009.

[57] M. Levoy, Z. Zhang, and I. McDowall, "Recording and controlling the 4d light field in a microscope using microlens arrays," *Journal of Microscopy*, vol. 235, pp. 144–162, 2009.

[58] R. Prevedel, Y. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, and A. Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature methods*, vol. 11, no. 7, pp. 727–730, 2014.

[59] T. Nöbauer, O. Skocek, A. J. Pernía-Andrade, L. Weilguny, F. M. Traub, M. I. Molodtsov, and A. Vaziri, "Video rate volumetric ca2+ imaging across cortex using seeded iterative demixing (sid) microscopy," *Nature Methods*, 2017.

[60] L. Mignard-Debise and I. Ihrke, "Light-field microscopy with a consumer light-field camera," in *International Conference on 3D Vision*, 2015, pp. 335–343.

[61] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin, "Shield fields: Modeling and capturing 3D occluders," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, vol. 27, no. 5, pp. 15–19, 2008.

[62] A. Ashok and M. A. Neifeld, "Compressive light field imaging," in *Proc. Spie*, 2010.

[63] N. C. Pégard, H.-Y. Liu, N. Antipa, M. Gerlock, H. Adesnik, and L. Waller, "Compressive light-field microscopy for 3D neural activity recording," *Optica*, vol. 3, no. 5, pp. 517–524, 2016.

[64] N. Cohen, S. Yang, A. Andalman, M. Broxton, L. Grosenick, K. Deisseroth, M. Horowitz, and M. Levoy, "Enhancing the performance of the light field microscope using wavefront coding," *Optics express*, vol. 22, no. 20, pp. 24 817–24 839, 2014.

[65] Y. Wang, L. Wang, D. Kong, and B. Yin, "High-resolution light field capture with coded aperture," *IEEE Trans. on Image Processing*, vol. 24, no. 1, 2015.

[66] S. Wanner and B. Goldlüecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE TPAMI*, vol. 36, no. 3, pp. 606–619, 2014.

[67] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *IEEE ICCP*, 2014.

[68] Y. Wang, Y. Liu, W. Heidrich, and Q. Dai, "The light field attachment: Turning a dslr into a light field camera using a low budget camera ring," *IEEE TVCG*, 2016.

[69] S. Pujades, F. Devernay, and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," in *IEEE CVPR*, 2014, pp. 3906–3913.

[70] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," in *IEEE CVPR*, 2015, pp. 3800–3809.

[71] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *IEEE CVPR*, 2017.

[72] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *IEEE CVPRW*, 2015, pp. 24–32.

[73] T.-C. Wang, J.-Y. Zhu, N. K. Kalantari, A. A. Efros, and R. Ramamoorthi, "Light field video capture using a learning-based hybrid imaging system," *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, vol. 36, no. 4, 2017.

[74] S. Wanner, S. Meister, and B. Goldlücke, "Datasets and benchmarks for densely sampled 4d light fields," in *Vision, Modeling & Visualization*, 2013, pp. 225–226.

[75] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 2, p. 16, 2015.

[76] W.-S. Chan, E. Y. Lam, M. K. Ng, and G. Y. Mak, "Super-resolution reconstruction in a computational compound-eye imaging system," *Multidimensional Systems and Signal Processing*, vol. 18, no. 2, pp. 83–101, 2007.

[77] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE TPAMI*, vol. 34, no. 5, pp. 972–986, 2012.

[78] K. Mitra and V. Ashok, "Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior," in *IEEE CVPRW*, 2012.

[79] R. A. Farrugia, C. Galea, and C. Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[80] P. James, B. Mike, and D. P. Luigi, "Plenoptic layer-based modeling for image based rendering," *IEEE TIP*, 2013.

[81] S. Wanner and B. Goldluecke, "Spatial and angular variational super-resolution of 4d light fields," in *IEEE ECCV*, 2012, pp. 608–621.

[82] G. Chaurasia, S. Duchêne, O. Sorkine-Hornung, and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation," *ACM TOG*, vol. 32, 2013.

[83] F.-L. Zhang, J. Wang, E. Shechtman, Z.-Y. Zhou, J.-X. Shi, and S.-M. Hu, "Plenopatch: Patch-based plenoptic image manipulation," *IEEE TVCG*, 2016.

[84] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deepstereo: Learning to predict new views from the world's imagery," in *IEEE CVPR*, 2015.

[85] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, 2016.

[86] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *IEEE CVPR*, 2010.

[87] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous fourier domain," *ACM TOG*, vol. 34, no. 1, p. 12, 2014.

[88] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand, and W. Matusik, "Joint view expansion and filtering for automultiscopic 3D displays," in *SIGGRAPH*, 2015, pp. 3906–3913.

[89] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," in *IEEE TPAMI*, 2017.

[90] ——, "Accelerated shearlet-domain light field reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[91] D. Mahajan, F.-C. Huang, W. Matusik, R. Ramamoorthi, and P. Belhumeur, "Moving gradients: a path-based method for plausible image interpolation," *ACM Transactions on Graphics*, vol. 28, no. 3, 2009.

[92] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung, "Phase-based frame interpolation for video," in *IEEE CVPR*, 2015, pp. 1410–1418.

[93] J. Wu, B. Xiong, X. Lin, J. He, J. Suo, and Q. Dai, "Snapshot hyperspectral volumetric microscopy," *Scientific Reports*, 2016.

[94] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu, "Snapshot hyperspectral light field imaging," in *IEEE CVPR*, 2017.

[95] H. Rueda, C. Fu, D. L. Lau, and G. R. Arce, "Spectral-tof compressive snapshot camera: Towards hyperspectral+depth imagery," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[96] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in *IEEE CVPR*, 2013.

[97] S. Heber and T. Pock, "Shape from light field meets robust pca," in *IEEE ECCV*, 2014.

[98] H. G. Jeon, J. Park, G. Choe, and J. Park, "Accurate depth map estimation from a lenslet light field camera," in *IEEE CVPR*, 2015, pp. 1547–1555.

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

[100] H. Zhu, Q. Wang, and J. Yu, "Occlusion-model guided anti-occlusion depth estimation in light field," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[101] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *IEEE CVPR*, 2014.

[102] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE CVPR*, 2013.

[103] M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi, "Depth from shading, defocus, and correspondence using light-field angular coherence," in *IEEE CVPR*, 2015, pp. 1940–1948.

[104] J. Tian, Z. Murez, T. Cui, Z. Zhang, D. Kriegman, and R. Ramamoorthi, "Depth and image restoration from light field in a scattering medium," in *IEEE ICCV*, 2017.

[105] W. Williem and I. K. Park, "Robust light field depth estimation for noisy scene with occlusion," in *IEEE CVPR*, 2016.

[106] H. Lin, C. Chen, S. B. Kang, and J. Yu, "Depth recovery from light field using focal stack symmetry," in *IEEE ICCV*, 2015.

[107] J. Y. Lee and R.-H. Park, "Depth estimation from light field by accumulating binary maps based on the foreground and background separation," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[108] M. Tao, J.-C. Su, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras," *IEEE TPAMI*, 2015.

[109] T.-C. Wang, M. Chandraker, A. Efros, and R. Ramamoorthi, "Svbrdf-invariant shape and reflectance estimation from light-field cameras," in *IEEE CVPR*, 2016.

[110] J. Li, M. Lu, and Z.-N. Li, "Continuous depth map reconstruction from light fields," *IEEE TIP*, vol. 24, no. 11, pp. 3257–3265, 2015.

[111] B. Krolla, M. Diebold, B. Goldlücke, and D. Stricker, "Spherical light fields," in *British Machine Vision Conference (BMVC)*, 2014.

[112] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4d light fields," in *IEEE CVPR*, 2012, pp. 41–48.

[113] M. Diebold, B. Jähne, and A. Gatto, "Heterogeneous light fields," in *IEEE CVPR*, 2016, pp. 1745–1753.

[114] S. Heber and T. Pock, "Convolutional networks for shape from light field," in *IEEE CVPR*, 2016, pp. 3746–3754.

[115] A. Criminisi, S. B. Kang, R. Swaminathan, S. Richard, and A. P., "Epipolar-plane image analysis: An approach to determining structure from motion," *Computer Vision and Image Understanding*, 2005.

[116] T. Basha, S. Avidan, A. Sorkine-Hornung, and W. Matusik, "Structure and motion from scene registration," in *IEEE CVPR*, 2012.

[117] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Marcus, and G. Markus, "A bidirectional light field - hologram transform," in *Computer Graphics Forum*, 2007.

[118] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," *Computer Vision and Image Understanding*, vol. 145, pp. 148–159, 2016.

[119] O. Jonhannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," in *IEEE CVPR*, 2016.

[120] S. Heber, R. Ranftl, and T. Pock, "Variational shape from light field," in *International Conference on Energy Minimization Methods in CVPR*, 2013.

[121] M. Diebold, O. Blum, M. Gutsche, S. Wanner, C. Garbe, H. Baker, and B. Jähne, "Light-field camera design for high-accuracy depth estimation," in *SPIE Optical Metrology*. International Society for Optics and Photonics, 2015, pp. 952 803–952 803.

[122] "ISO 12233." Photography-Electronic still picture cameras-Resolution measurements (2000).

[123] Y. Y. B. X. Qiang Fu, Zhiliang Zhou, "Image quality evaluation of light field photography," *Image Quality & System Performance*, vol. 7867, no. 4, pp. 357–366, 2011.

[124] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[125] H. Shidanshidi, F. Safaei, and W. Li, "A quantitative approach for comparison and evaluation of light field rendering techniques," in *IEEE International Conference on Multimedia & Expo*, vol. 50, no. 4, 2011.

[126] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *IEEE CVPR*, 2017.

[127] A. Aggoun, "A 3D DCT compression algorithm for omnidirectional integral images," in *IEEE Inter. Conf. on Acoustics, Speech and Signal Processing*, 2006.

[128] ——, "Compression of 3D integral images using 3D wavelet transform," *Journal of Display Technology*, vol. 7, no. 11, pp. 586–592, 2011.

[129] M. Magnor, A. Endmann, and B. Girod, "Progressive compression and rendering of light fields," in *Vision, Modeling, and Visualization Proceedings*, 2000.

[130] D. Xu, Q. Dai, and W. Xu, "Data compression of light field using wavelet packet," in *IEEE International Conference on Multimedia and Expo*, 2004, pp. 1071–1074.

[131] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod, "Light field compression using disparity-compensated lifting and shape adaptation," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 793–806, 2006.

[132] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338C–343, 2000.

[133] C. Conti, P. Kovacs, T. Balogh, P. Nunes, and L. D. Soares, "Light-field video coding using geometry-based disparity compensation," in *3DTV-Conference*, 2014.

[134] S. Kundu, "Light field compression using homography and 2d warping," in *IEEE International Conference on Acoustics*, 2012, pp. 1349–1352.

[135] X. Jiang, M. L. Pendu, R. Farrugia, and C. Guillemot, "Light field compression with homography-based low rank approximation," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[136] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *IEEE International Conference on Multimedia and Expo Workshops*, 2016.

[137] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 80–91, 2016.

[138] X. Jin, H. Han, and Q. Dai, "Image reshaping for efficient compression of plenoptic content," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[139] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," in *IEEE ICIP*, 2015, pp. 4733–4737.

[140] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, "Data formats for high efficiency coding of lytro-illum light fields," in *International Conference on Image Processing Theory*, 2015, pp. 494–497.

[141] L. Li, Z. Li, B. Li, D. Liu, and H. Li, "Pseudo sequence based 2-d hierarchical coding structure for light-field image compression," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[142] C. Perra, "Lossless plenoptic image compression using adaptive block differential prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[143] P. Helin, P. Astola, B. Rao, and I. Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," in *3DTV-CON*, 2016.

[144] ——, "Minimum description length sparse modeling and region merging for lossless plenoptic image compression," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[145] "JPEG Pleno." https://jpeg.org/jpegpleno/index.html.

[146] "ICME 2016 Grand Challenge on Light-Field Image Compression." http://mmspg.epfl.ch/ICME2016GrandChallenge_1.

[147] C. Liang, "Predictive light field compression," Jul. 21 2016, uS Patent App. 15/085,048. [Online]. Available: https://www.google.com/patents/US20160212443

[148] I. Viola, M. Řeřábek, and T. Ebrahimi, "Evaluation of light field image coding approaches," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[149] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic aperture confocal imaging," *ACM Transactions on Graphics (TOG)*, 2004.

[150] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 15:1–15:20, 2015.

[151] K. James, T., "The rendering equation," *ACM Transactions on Graphics (TOG)*, 1986.

[152] L. Yatziv, G. Sapiro, and M. Levoy, "Lightfield completion," in *IEEE ICIP*, 2004, pp. 1787–1790.

[153] V. Vaish, R. Szeliski, C. L. Zitnick, S. B. Kang, and M. Levoy, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," in *IEEE CVPR*, 2006.

[154] T. Yang, Y. Zhang, J. Yu, J. Li, W. Ma, X. Tong, R. Yu, and L. Ran, "All-in-focus synthetic aperture imaging," in *IEEE ECCV*, 2014.

[155] N. Joshi, S. Avidan, W. Matusik, and K. D. J., "Synthetic aperture tracking: Tracking through occlusions," in *IEEE ICCV*, 2007.

[156] T. Yang, Y. Zhang, and X. Tong, "A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion," *IEEE CSVT*, 2013.

[157] S. Wanner, C. Straehle, and B. Goldluecke, "Globally consistent multi-label assignment on the ray space of 4d light fields," in *IEEE CVPR*, 2013, pp. 1011–1018.

[158] K. Yücer, A. Sorkine-Hornung, O. Wang, and O. Sorkine-Hornung, "Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction," *ACM Transactions on Graphics*, vol. 35, no. 3, 2016.

[159] K. Yücer, C. Kim, A. Sorkine-Hornung, and O. Sorkine-Hornung, "Depth from gradients in dense light fields for object reconstruction," in *International Conference on 3D Vision*, 2016, pp. 249–257.

[160] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4d light-field dataset and cnn architectures for material recognition," in *IEEE ECCV*. Springer, 2016, pp. 121–138.

[161] N. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, "Automatic 3D object segmentation in multiple views using volumetric graph-cuts," *Image & Vision Computing*, vol. 28, no. 1, pp. 14–25, 2010.

[162] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, "Automatic object segmentation from calibrated images," in *Conference for Visual Media Production*, 2011.

[163] M. Hog, N. Sabater, and C. Guillemot, "Light field segmentation using a ray-based graph structure," in *IEEE ECCV*, 2016, pp. 35–50.

[164] J. Berent and D. Luigi, "Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition," in *IEEE Workshop on Multimedia Signal Processing*, 2006, pp. 182–185.

[165] H. Zhu, Q. Zhang, and Q. Wang, "4d light field superpixel and segmentation," in *IEEE CVPR*, 2017.

[166] O. Johannsen, A. Sulc, and B. Goldluecke, "Variational separation of light field layers," in *Vision, Modeling & Visualization, VMV*, 2015, pp. 135–142.

[167] Y. Xu, H. Nagahara, A. Shimada, and R. Taniguchi, "Transcut: Transparent object segmentation from a light-field image," in *IEEE ICCV*, 2015, pp. 3442–3450.

[168] N. Joshi, W. Matusik, and S. Avidan, "Natural video matting using camera arrays," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 779–786, 2006.

[169] D. Cho, S. Kim, and Y.-W. Tai, "Consistent matting for light field images," in *IEEE ECCV*, 2014, pp. 90–104.

[170] J. Fiss, B. Curless, and R. Szeliski, "Light field layer matting," in *IEEE CVPR*, 2015, pp. 623–631.

[171] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *IEEE CVPR*, 2014.

[172] Z. Pei, Y. Zhang, T. Yang, X. Zhang, and Y.-H. Yang, "A novel multi-object detection method in complex scene using synthetic aperture imaging," *Pattern Recognition*, vol. 45, pp. 1637–1658, 2012.

[173] R. Gross, I. Matthews, and S. Baker, "Eigen light-fields and face recognition across pose," in *International Conference on Automatic Face and Gesture Recognition*, vol. 44, no. 41, 2002.

[174] ——, "Appearance-based face recognition and light-fields," *IEEE TPAMI*, vol. 26, no. 4, pp. 449–465, 2004.

[175] R. Raghavendra, B. Yang, K. B. Raja, and C. Busch, "A new perspective - face recognition with light-field camera," in *International Conference on Biometrics*, 2013.

[176] ——, "Robust iris recognition using light-field camera," in *Colour & Visual Computing Symposium*, 2013.

[177] G. Wetzstein, D. Roodnick, R. Raskar, and W. Heidrich, "Refractive Shape from Light Field Distortion," in *IEEE ICCV*, 2011.

[178] G. Wetzstein, R. Raskar, and W. Heidrich, "Hand-Held Schlieren Photography with Light Field Probes," in *IEEE ICCP*, 2011.

[179] K. Maeno, H. Nagahara, A. Shimada, and R. ichiro Taniguichi, "Light field distortion feature for transparent object recognition," in *IEEE ICCV*, 2013.

[180] P. Anglin, S. J. Reeves, and B. S. Thurow, "Characterization of plenoptic imaging systems and efficient volumetric estimation from plenoptic data," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[181] Y. Ding, F. Li, Y. Ji, and J. Yu, "Dynamic fluid surface acquisition using a camera array," in *IEEE ICCV*, 2011.

[182] A. Sulc, A. Alperovich, N. Marniok, and B. Goldluecke, "Reflection separation in light fields based on sparse coding and specular flow," in *Vision, Modeling & Visualization, VMV*, 2016.

[183] F. Lu, L. He, S. You, and Z. Hao, "Identifying surface brdf from a single 4d light field image via deep neural network," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[184] D. G. Dansereau, I. Mahon, O. Pizarro, and S. B. Williams, "Plenoptic flow: Closed-form visual odometry for light field cameras," in *International Conference on Intelligent Robots and Systems*, 2011.

[185] N. Zeller, F. Quint, and U. Stilla, "From the calibration of a light-field camera to direct plenoptic odometry," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[186] I. Schillebeeckx and R. Pless, "Structured light field design for correspondence free rotation estimation," in *IEEE ICCP*, 2015, pp. 1–10.

[187] P. P. Srinivasan, M. W. Tao, R. Ng, and R. Ramamoorthi, "Oriented light-field windows for scene flow," in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, 2015, pp. 3496–3504.

[188] T. Fahringer, K. Lynch, and B. S. Thurow, "Volumetric particle image velocimetry with a single plenoptic camera," *Measurement Science and Technology*, vol. 26, no. 11, pp. 115–201, 2015.

[189] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu, "Enhancing light fields through ray-space stitching," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 7, pp. 1852–1861, 2016.

[190] O. Johannsen, A. Sulc, and B. Goldluecke, "On linear structure from motion for light field cameras," in *IEEE ICCV*, 2015, pp. 720–728.

[191] P. P. Srinivasan, R. Ng, and R. Ramamoorthi, "Light field blind motion deblurring," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[192] B. M. Smith, L. Zhang, H. Jin, and A. Agarwala, "Light field video stabilization," in *IEEE ICCV*, 2009.

[193] S. M. Seitz and K. N. Kutulakos, "Plenoptic image editing," in *IEEE ICCV*, 1998.

[194] ——, "Plenoptic image editing," *International Journal of Computer Vision*, vol. 48, no. 2, 2002.

[195] Z. Zhang, L. Wang, B. Guo, and H.-Y. Shum, "Feature-based light field morphing," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 457–464, 2002.

[196] A. Jarabo, B. Masia, A. Bousseau, F. Pellacini, and D. Gutierrez, "How do people edit light fields?" *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 70–79, 2014.

[197] B. Chen, E. Ofek, H.-Y. Shum, and L. M., "Interactive deformation of light fields," in *Symposium on Interactive 3d Graphics*, vol. 50, 2005, pp. 139–146.

[198] C. Birklbauer, D. C. Schedl, and O. Bimber, "Nonuniform spatial deformation of light fields by locally linear transformations," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 5, p. 156, 2016.

[199] B. Masia, A. Jarabo, and D. Gutierrez, "Favored workflows in light field editing," in *Proceedings of CGVCVIP '14*, 2014.

[200] M. Ortin, A. Jarabo, B. Masia, and D. Gutierrez, "Analyzing interfaces and workflows for light field editing," *IEEE Journal on Selected Topics on Signal Processing*, 2017.

[201] F. Marton, M. Agus, E. Gobbetti, and M. B. Rodriguez, "Natural exploration of 3D massive models on large-scale light field displays using the fox proximal navigation technique," *Computers & Graphics*, vol. 36, no. 8, pp. 893–903, 2012.

[202] J. Tompkin, S. Muff, S. Jakuschevskij, J. McCann, J. Kautz, M. Alexa, and W. Matusik, "Interactive light field painting," *ACM Trans. Graph.*, 2012.

[203] Vertical Horizon, "Lightfield Iris," http://www.verticalhorizon-software.com/LFI/index.html.

[204] R. Horn, Daniel and B. Chen, "Lightshop: interactive light field manipulation and rendering," in *Symposium on Interactive 3d Graphics*, 2007, pp. 121–128.

[205] O. Cossairt, S. K. Nayar, and R. Ramamoorthi, "Light field transfer: global illumination between real and synthetic objects," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 57:1–57:6, 2008.

[206] L. Wang, S. Lin, S. Lee, B. Guo, and Y. Shum, Heung, "Light field morphing using 2d features," *IEEE TVCG*, vol. 11, no. 1, pp. 25–34, 2005.

[207] C. Birklbauer and O. Bimber, "Light-field retargeting," in *Computer Graphics Forum*, vol. 31, 2012, pp. 295–303.

[208] E. Garces, J. I. Echevarria, W. Zhang, H. Wu, K. Zhou, and D. Gutierrez, "Intrinsic light field images," in *Computer Graphics Forum*. Wiley Online Library, 2017.

[209] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez, "Intrinsic video and applications," *ACM Transactions on Graphics (SIGGRAPH 2014)*, vol. 33, no. 4, 2014.

[210] S. Duchêne, C. Riant, G. Chaurasia, J. Lopez-Moreno, P.-Y. Laffont, S. Popov, A. Bousseau, and G. Drettakis, "Multi-view intrinsic images of outdoors scenes with an application to relighting," *ACM Transactions on Graphics*, 2015.

[211] A. Jarabo, B. Masia, and D. Gutierrez, "Efficient propagation of light field edits," in *Proceedings of SIACG*, 2011.

[212] X. An and F. Pellacini, "Appprop: all-pairs appearance-space edit propagation," *ACM Trans. Graph.*, vol. 27, pp. 40:1–40:9, 2008.

[213] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, 2007.

[214] H. Ao, Y. Zhang, A. Jarabo, B. Masia, Y. Liu, D. Gutierrez, and Q. Dai, "Light field editing based on reparameterization," in *Pacific Rim Conference on Multimedia*, 2015.

[215] Williem, K. W. Shon, and I. K. Park, "Spatio-angular consistent editing framework for 4d light field images," *Multimedia Tools and Applications*, vol. 75, no. 23, 2016.

[216] Y. Gryaditskaya, B. Masia, P. Didyk, K. Myszkowski, and H.-P. Seidel, "Gloss editing in light fields," in *Proc. International Workshop on Vision, Modeling and Visualization (VMV)*, 2016.

[217] A. Davis, M. Levoy, and F. Durand, "Unstructured light fields," *Comput. Graph. Forum*, vol. 31, no. 2, pp. 305–314, 2012.

[218] A. Wender, J. Iseringhausen, B. Goldluecke, M. Fuchs, and M. B. Hullin, "Light field imaging through household optics," in *Vision, Modeling & Visualization, VMV*, 2015, pp. 159–166.

[219] K. Xu, Y. Li, T. Ju, S.-M. Hu, and T.-Q. Liu, "Efficient affinity-based edit propagation using k-d tree," *ACM Trans. Graph.*, vol. 28, pp. 118:1–118:6, 2009.

[220] P. Krähenbühl, M. Lang, A. Hornung, and M. Gross, "A system for retargeting of streaming video," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, p. 126, 2009.

[221] A. Meka, M. Zollhöfer, C. Richardt, and C. Theobalt, "Live intrinsic video," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.

[222] S. W. Hasinoff, M. Jozwiak, F. Durand, and W. T. Freeman, "Search-and-replace editing for personal photo collections," in *IEEE ICCP*, 2010.

[223] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph.*, vol. 30, no. 4, 2011.

[224] K. Yücer, A. Jacobson, A. Hornung, and O. Sorkine, "Transfusive image manipulation," *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH ASIA)*, vol. 31, no. 6, pp. 176:1–176:9, 2012.

[225] F. E. Ives, "Parallax stereogram and process of making same." Apr. 14 1903, uS Patent 725,567.

[226] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[227] W. Matusik and H. Pfister, "3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 814–824, 2004.

[228] K. Perlin, S. Paxia, and J. S. Kollin, "An autostereoscopic display," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, New Orleans, LA, USA, July 23-28, 2000*, 2000, pp. 319–326.

[229] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 540–555, 2011.

[230] E. Lueder, *3D Displays*. John Wiley & Sons, 2011, vol. 32.

[231] G. Wetzstein, D. Lanman, D. Gutierrez, and M. Hirsch, "Computational displays," *SIGGRAPH 2012 (course)*, 2012.

[232] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers & Graphics*, vol. 37, no. 8, pp. 1012–1038, 2013.

[233] G. Woodgate, D. Ezra, N. Holliman, B. Omar, R. Moseley, and J. Harrold, "Autostereoscopic display and method of controlling an autostereoscopic display," Sep. 15 1998, uS Patent 5,808,792. [Online]. Available: https://www.google.com/patents/US5808792

[234] T. Peterka, R. L. Kooima, D. J. Sandin, A. Johnson, J. Leigh, and T. A. DeFanti, "Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 3, pp. 487–499, 2008.

[235] S.-K. Kim, K.-H. Yoon, S. K. Yoon, and H. Ju, "Parallax barrier engineering for image quality improvement in an autostereoscopic 3D display," *Optics express*, vol. 23, no. 10, pp. 13 230–13 244, 2015.

[236] H. Gotoda, "A multilayer liquid crystal display for autostereoscopic 3D viewing," in *IS&T/SPIE Electronic Imaging*, 2010, pp. 75 240–75 240.

[237] N. Ranieri, S. Heinzle, Q. Smithwick, D. Reetz, L. S. Smoot, W. Matusik, and M. Gross, "Multi-layered automultiscopic displays," in *Computer Graphics Forum*, vol. 31, no. 7. Wiley Online Library, 2012, pp. 2135–2143.

[238] N. Ranieri, S. Heinzle, P. Barnum, W. Matusik, and M. Gross, "32.3: Light-field approximation using basic display layer primitives," in *SID symposium digest of technical papers*, vol. 44, no. 1. Wiley Online Library, 2013, pp. 408–411.

[239] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien, "Optimal presentation of imagery with focus cues on multi-plane displays," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 59, 2015.

[240] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 80:1–80:11, 2012.

[241] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 6, p. 163, 2010.

[242] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 95.

[243] F. Heide, D. Lanman, D. Reddy, J. Kautz, K. Pulli, and D. Luebke, "Cascaded displays: spatiotemporal superresolution using offset pixel layers," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, 2014.

[244] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs, "Focus 3D: Compressive accommodation display." *ACM Trans. Graph.*, vol. 32, no. 5, pp. 153–1, 2013.

[245] F. Heide, G. Wetzstein, R. Raskar, and W. Heidrich, "Adaptive image synthesis for compressive displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, 2013.

[246] A. Jarabo, R. Buisan, and D. Gutierrez, "Bidirectional clustering for scalable VPL-based global illumination," in *Proceedings of CEIG '15*, 2015.

[247] A. J. Woods, "Crosstalk in stereoscopic displays: a review," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 040 902–040 902, 2012.

[248] I. Tsirlin, L. M. Wilcox, and R. S. Allison, "The effect of crosstalk on the perceived depth from disparity and monocular occlusions," *IEEE Transactions on broadcasting*, vol. 57, no. 2, pp. 445–453, 2011.

[249] ——, "Effect of crosstalk on depth magnitude in thin structures," *J. Electronic Imaging*, vol. 21, no. 1, p. 011003, 2012.

[250] A. Jain and J. Konrad, "Crosstalk in automultiscopic 3-d displays: Blessing in disguise?" in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 649 012–649 012.

[251] X. Wang and C. Hou, "Improved crosstalk reduction on multiview 3D display by using bils algorithm," *Journal of Applied Mathematics*, vol. 2014, 2014.

[252] D. Li, D. Zang, X. Qiao, L. Wang, and M. Zhang, "3D synthesis and crosstalk reduction for lenticular autostereoscopic displays," *Journal of Display Technology*, vol. 11, no. 11, pp. 939–946, 2015.

[253] R. Pei, Z. Geng, Z. X. Zhang, X. Cao, and R. Wang, "A novel optimization method for lenticular 3-d display based on light field decomposion," *Journal of Display Technology*, vol. 12, no. 7, pp. 727–735, July 2016.

[254] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan, "A new reconstruction filter for undersampled light fields," in *ACM International Conference Proceeding Series*, vol. 44, 2003, pp. 150–156.

[255] M. Zwicker, W. Matusik, F. Durand, H. Pfister, and C. Forlines, "Antialiasing for automultiscopic 3D displays," in *ACM SIGGRAPH*. ACM, 2006.

[256] J. Konrad and P. Agniel, "Subsampling models and anti-alias filters for 3-d automultiscopic displays," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 128–140, 2006.

[257] A. Chapiro, S. Heinzle, T. O. Aydın, S. Poulakos, M. Zwicker, A. Smolic, and M. Gross, "Optimizing stereo-to-multiview conversion for autostereoscopic displays," in *Computer graphics forum*, vol. 33, no. 2, 2014, pp. 63–72.

[258] S.-P. Du, P. Didyk, F. Durand, S.-M. Hu, and W. Matusik, "Improving visual quality of view transitions in automultiscopic displays," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, 2014.

[259] S. Xie, X. Sang, P. Wang, N. Guo, Z. Chen, X. Yu, B. Yan, K. Wang, and C. Yu, "Depth-reversal-free three-dimensional display by adjusting the light field alignment," *Optics Communications*, 2016.

[260] J. Liu, T. Malzbender, S. Qin, B. Zhang, C.-A. Wu, and J. Davis, "Dynamic mapping for multiview autostereoscopic displays," in *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 2015.

[261] F. Banterle, A. Artusi, T. O. Aydın, P. Didyk, E. Eisemann, D. Gutierrez, R. Mantiuk, and K. Myszkowski, "Multidimensional image retargeting," in *SIGGRAPH Asia 2011 Courses*, 2011.

[262] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "Adaptive image-space stereo view synthesis." in *VMV*, 2010, pp. 299–306.

[263] N. Plath, S. Knorr, L. Goldmann, and T. Sikora, "Adaptive image warping for hole prevention in 3D view synthesis," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3420–3432, 2013.

[264] M. Kim, S. Lee, C. Choi, G.-M. Um, N. Hur, and J. Kim, "Depth scaling of multiview images for automultiscopic 3D monitors," in *2008 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*. IEEE, 2008, pp. 181–184.

[265] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 4, 2010.

[266] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "A perceptual model for disparity," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011.

[267] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, H.-P. Seidel, and W. Matusik, "A luminance-contrast-aware disparity model and applications," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, 2012.

[268] B. Masia, G. Wetzstein, C. Aliaga, R. Raskar, and D. Gutierrez, "Display adaptive 3D content remapping," *Computers & Graphics*, vol. 37, no. 8, pp. 983–996, 2013.

[269] C. Kim, A. Hornung, S. Heinzle, W. Matusik, and M. Gross, "Multi-perspective stereoscopy from light fields," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, 2011.

[270] V. Ramachandra, M. Zwicker, and T. Q. Nguyen, "Combined image plus depth seam carving for multiview 3D images," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 737–740.

[271] P. T. Kovács, R. Bregović, A. Boev, A. Barsi, and A. Gotchev, "Quantifying spatial and angular resolution of light field 3D displays," *IEEE Journal of Selected Topics in Signal Processing*, 2017.

[272] Y. Itoh and G. Klinker, "Light-field correction for spatial calibration of optical see-through head-mounted displays," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 4, pp. 471–480, 2015.

[273] Y. Itoh, T. Amano, D. Iwai, and G. Klinker, "Gaussian light field: Estimation of viewpoint-dependent blur for optical see-through head-mounted displays," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 11, pp. 2368–2376, 2016.

[274] "Stanford Lytro Light Field Archive." http://lightfields.stanford.edu/.

[275] M. Řeřábek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal*, 2016.

[276] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldlücke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 19–34.

[277] "Synthetic Light Field Archive." http://web.media.mit.edu/ gordonw/SyntheticLightFields/.

[278] "How do people edit light fields?: Dataset," http://giga.cps.unizar.es/ ajarabo/pubs/lfeiSIG14/data/index.html.

[279] "Light Field Datasets." http://cseweb.ucsd.edu/ viscomp/projects/LF/.

[280] "Lytro Cinema." https://lytro.com/cinema.

[281] K. Guo, F. Xu, Y. Wang, Y. Liu, and Q. Dai, "Robust non-rigid motion tracking and surface reconstruction using L_0 regularization," *IEEE Transactions on Visualization & Computer Graphics*, 2017.
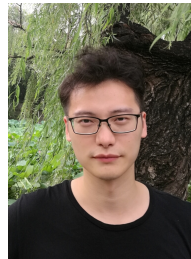
Gaochang Wu received the B.S. and M.S. degrees in mechanical engineering from Northeastern University, Shenyang, China, in 2013, and 2015, respectively. He is currently working toward Ph.D. degree in the State Key Laboratory of Synthetical Automation for Process Industries at Northeastern University, China. His current research interests include image processing, machine learning and computational photography.

Belen Masia received her PhD in Computer Science from Universidad de Zaragoza, Spain, in 2013. She has been a postdoctoral researcher at the Max Planck Institute for Informatics in Saarbruecken, Germany, and a visiting researcher at the Massachusetts Institute of Technology. Belen is currently an Assistant Professor in the Department of Computer Science at Universidad de Zaragoza, and a member of the I3A Research Institute. Her research in computational imaging, displays and applied perception has been published in top venues such as ACM Transactions on Graphics. She is an Associate Editor of ACM Transactions on Applied Perception, and of Computers & Graphics. She has also served in numerous technical papers program committees, including ACM SIGGRAPH Asia, Eurographics, and Pacific Graphics. She has received a number of awards, including the Eurographics Young Researcher Award (2017), a Eurographics PhD Award (2015), she was selected as one of the top ten Innovators Under 35 in Spain by MIT Technology Review (2014), and has also received a NVIDIA Graduate Fellowship (2012). Belen is a Eurographics Young Fellow.

Adrian Jarabo is a postdoctoral researcher at Universidad de Zaragoza, where he got his Ph.D. degree in Computer Graphics in 2015. Before that, he was a visiting researcher in Trinity College Dublin and Microsoft Research Asia. His research interests include topics such as light transport simulation, appearance modeling, or plenoptic manipulation. He received the Eurographics PhD Award in 2017.

Yuchen Zhang received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013, where he is currently pursuing the Ph.D. degree. His main research interests include computational photography and graph signal processing.

Liangyong Wang received the B.E. degree in automation, the M.Sc. degree and Ph.D. degree in control theory and control engineering from Northeastern University, China, respectively in 2001, 2004 and 2010. He is currently with the State Key Laboratory of Synthetical Automation for Process Industries at Northeastern University, China. His research interests are cyber-physical systems, machine learning and big data focused on industrial applications.

Qionghai Dai received the BS degree in mathematics from Shanxi Normal University, China, in 1987, and the ME and PhD degrees in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the faculty of Tsinghua University, Beijing, China, and is currently a professor and the director of the Broadband Networks and Digital Media Laboratory. His research areas include video communication, computer vision, and graphics. He is a senior member of the IEEE.

Tianyou Chai received the Ph.D. degree in control theory and engineering from Northeastern University, Shenyang, China, in 1985. He has been with the Research Center of Automation, Northeastern University, Shenyang, China, since 1985, where he became a Professor in 1988 and a Chair Professor in 2004. His current research interests include adaptive control, intelligent decoupling control, integrated plant control and systems, and the development of control technologies with applications to various industrial processes. Prof. Chai is a member of the Chinese Academy of Engineering, an academician of International Eurasian Academy of Sciences, IEEE Fellow and IFAC Fellow. He is a distinguished visiting fellow of The Royal Academy of Engineering (UK) and an Invitation Fellow of Japan Society for the Promotion of Science (JSPS).

Yebin Liu received the BE degree from Beijing University of Posts and Telecommunications, China, in 2002, and the PhD degree from the Automation Department, Tsinghua University, Beijing, China, in 2009. He has been working as a research fellow at the computer graphics group of the Max Planck Institute for Informatik, Germany, in 2010. He is currently an associate professor in Tsinghua University. His research areas include computer vision and computer graphics.