Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review

Light field imaging for computer vision: a survey*#

Chen JIA[1,2], Fan SHI[†‡1,2], Meng ZHAO[1,2], Shengyong CHEN[1,2]

[1]*Engineering Research Center of Learning-Based Intelligent System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, China*

[2]*Key Laboratory of Computer Vision and System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, China*

[†]E-mail: shifan@email.tjut.edu.cn

Received Apr. 7, 2021; Revision accepted Nov. 10, 2021; Crosschecked Mar. 16, 2022

Abstract: Light field (LF) imaging has attracted attention because of its ability to solve computer vision problems. In this paper we briefly review the research progress in computer vision in recent years. For most factors that affect computer vision development, the richness and accuracy of visual information acquisition are decisive. LF imaging technology has made great contributions to computer vision because it uses cameras or microlens arrays to record the position and direction information of light rays, acquiring complete three-dimensional (3D) scene information. LF imaging technology improves the accuracy of depth estimation, image segmentation, blending, fusion, and 3D reconstruction. LF has also been innovatively applied to iris and face recognition, identification of materials and fake pedestrians, acquisition of epipolar plane images, shape recovery, and LF microscopy. Here, we further summarize the existing problems and the development trends of LF imaging in computer vision, including the establishment and evaluation of the LF dataset, applications under high dynamic range (HDR) conditions, LF image enhancement, virtual reality, 3D display, and 3D movies, military optical camouflage technology, image recognition at micro-scale, image processing method based on HDR, and the optimal relationship between spatial resolution and four-dimensional (4D) LF information acquisition. LF imaging has achieved great success in various studies. Over the past 25 years, more than 180 publications have reported the capability of LF imaging in solving computer vision problems. We summarize these reports to make it easier for researchers to search the detailed methods for specific solutions.

Key words: Light field imaging; Camera array; Microlens array; Epipolar plane image; Computer vision
https://doi.org/10.1631/FITEE.2100180 **CLC number:** TP391.4

1 Introduction

Light field (LF) imaging is a type of computational photographic technology. In contrast to two-dimensional (2D) imaging methods, a light field camera (LFC) can record the intensity and direction of each ray, and multiple views of the scene can be captured in one exposure. This capability brings new possibilities for post-processing of an image. Many studies have proven that LF imaging can greatly improve the capabilities and performances of computer vision technologies, including digital refocusing, image segmentation, shape recovery, and saliency detection.

LF imaging technology was pioneered by Lippmann (1908). Using an early LF imaging method called integral photography (IP), he proposed that true stereo images can be reconstructed based on the reversibility principle of light rays. Later, Gershun (1939) first defined the distribution of light rays as a model. Then, Adelson and Bergen (1991) improved this definition and proposed to name this model "the

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (Nos. 61906133, 62020106004, and 92048301)
# Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2100180) contains supplementary materials, which are available to authorized users
ORCID: Chen JIA, https://orcid.org/0000-0003-3043-7193; Fan SHI, https://orcid.org/0000-0003-2074-0228

plenoptic function" based on the position, angle, wavelength, and time of each light ray. Numerous systems of LF acquisition have been built, most of which are composed of multi-cameras or camera arrays. For these systems, the cameras or camera arrays are distributed on a planar surface, and each camera records the radiance of light rays on a single point. The first handheld LFC, proposed by Ng et al. (2005), was portable, cheap, and useful. In recent years, LFCs such as the Lytro and Raytrix, have been made available commercially. These cameras make it easy to capture the information of the whole LF for common users, and thus have attracted much interest. Because of its outstanding performance, LF imaging technology has great potential in various fields of computer vision research and application, such as biometric detection and robotics vision.

2 Overview of related works

2.1 Summary

In the 120-year history of LF imaging, the rapid application of optical field imaging technology to computer vision has occurred mainly in the last 25 years. Therefore, the research studies summarized in this report were selected primarily from this period. The main developments over that period were as follows: (1) from 1996 to 2010, initial development and application of LF imaging; (2) from 2010 to 2015, rapid development of LF imaging in computer vision research; (3) from 2015 to 2021, peak period of LF imaging. To our knowledge, however, there have been few reviews of research on LF imaging calculations, and very few focusing on the application of LF imaging to computer vision.

Recently, there has been a rapid development of LF imaging technology in computer vision. In particular, since 2015, the commercial promotion, development, and popularization of LFCs have accelerated. Additionally, many innovative computer vision theories and models have been proposed and applied, such as LF depth estimation theory, LF super-resolution theory, and face recognition models. We believe that it is important to summarize the development of LF imaging technology in computer vision, to provide information for researchers.

After LFC was made available commercially in 2005, this topic has attracted many researchers. However, reviews of recent advances are rare. In this paper, we summarize the development of LF imaging technology in computer vision, and discuss the possible research directions and challenges of LF vision. In most cases, we provide a comprehensive list of references to enable researchers to better select and summarize the advanced ideas, along with the key methods. We hope that our work may help relevant workers to quickly understand the theoretical basis and practical development needs of these directions.

2.2 Representatives

LF imaging technology is widely applied in computer vision tasks. The most substantial contributions are hierarchical and can be divided into two categories: LF image based and stereo visual cues based. To clarify the relationship between LF imaging and computer vision tasks, we have used a diagram (Fig. 1) to describe the contributions. In addition, the common tasks and applications, as well as the common problems and future trends, are shown in Figs. 2 and 3.

Studies of LF imaging in computer vision can be divided into four subject areas: (1) LF processing

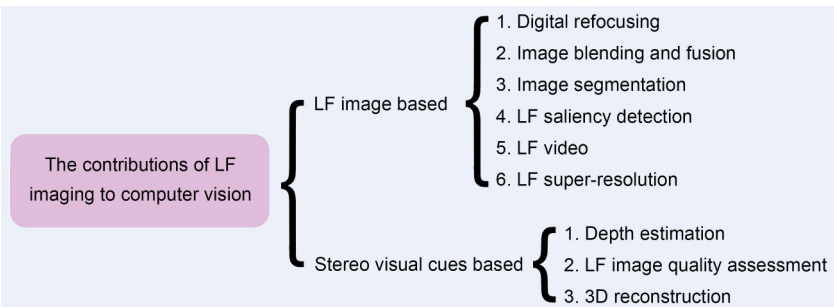

Fig. 1 The contributions of light field (LF) imaging to computer vision

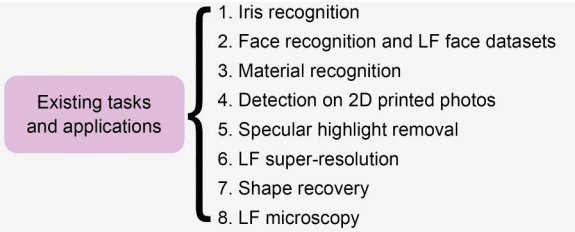

Fig. 2 Eight existing tasks and applications

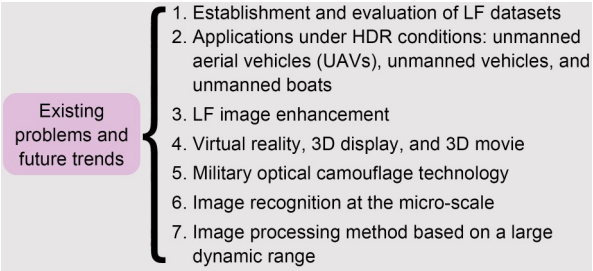

Fig. 3 Seven existing problems and future trends

and image generation; (2) low-level vision; (3) middle-level vision; (4) high-level vision. LF processing and image generation research includes mainly the LF function, optical schematic of LFC, and LF super-resolution. Low-level vision research includes mainly LF depth estimation and 3D reconstruction. Middle-level vision research includes mainly image segmentation and blending. High-level vision research includes mainly face recognition and material recognition or detection.

3 Light field imaging technology

3.1 Light field function

For imaging convenience, traditional 2D imaging systems ignore the analysis of the angle information of the light rays. They directly analyze the direction information of the light rays on which the lens focuses, such as a pixel point (s,t) in a certain exposure time recorded on a charge-coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS). For the basic principle model of LF imaging (Fig. 4), the LF information is determined by recording the intersection points (u, v) and (s, t) information of each light ray intersecting the lens plane and image plane. Based on the above imaging principle, the four-dimensional (4D) LF information can be regarded as a summary of each light ray's information.

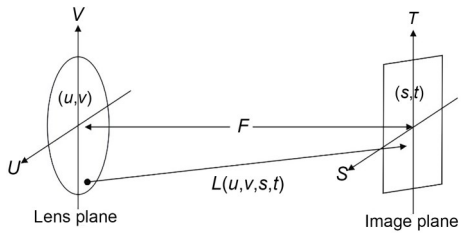

Fig. 4 A conceptual diagram of light field imaging

In addition, the 4D LF can be expressed in a visual way. The *st* plane is regarded as a group of cameras that can obtain light, and the *uv* plane is regarded as the focal plane of the camera. Furthermore, the two-plane LF model $L(u, v, s, t)$ can demonstrate the acquisition of LF in two different ways. First, the camera captures all the light rays passing through the *st* plane and forming the focus with the *uv* plane (a set of rays at a certain viewpoint), and represents the 4D LF with a 2D array of images (Fig. 5a). Each 2D image is called a subaperture image. Second, the number of samples on the *st* plane is positively related to the number of viewpoints, and the number of samples on the *uv* plane is positively related to the resolution of the camera. Therefore, in general, the s and t dimensions are referred to as the angular dimensions, and the u and v dimensions as the spatial dimensions. By acquiring $I_{s^*,t^*}(u,v)$ and $I_{u^*,v^*}(s,t)$, slices $E_{v^*,t^*}(u,s)$ or $E_{u^*,s^*}(v,t)$ can be produced, also known as epipolar plane images (EPIs).

When the camera (fixed s^*t^* plane) is kept stationary, subaperture images $I_{s^*,t^*}(u,v)$ can be collected. Similarly, when the u^*v^* plane is fixed, the LF subviews $I_{u^*,v^*}(s,t)$ composed of light rays from different viewpoints can be collected. The EPI contains spatial information and angular information about an LF. At the same time, the points of different depths of an object are generally regarded as lines with different slopes in EPIs. The slope of a line represents the depth information of a point in the object, and thus the EPI is widely used in object depth estimation research.

3.2 Light field acquisition

In terms of structure, there are two main LF acquisition types: multi-camera array structure and single camera optical element structure. These two LF acquisition methods with different structures can effectively acquire LF scene information.

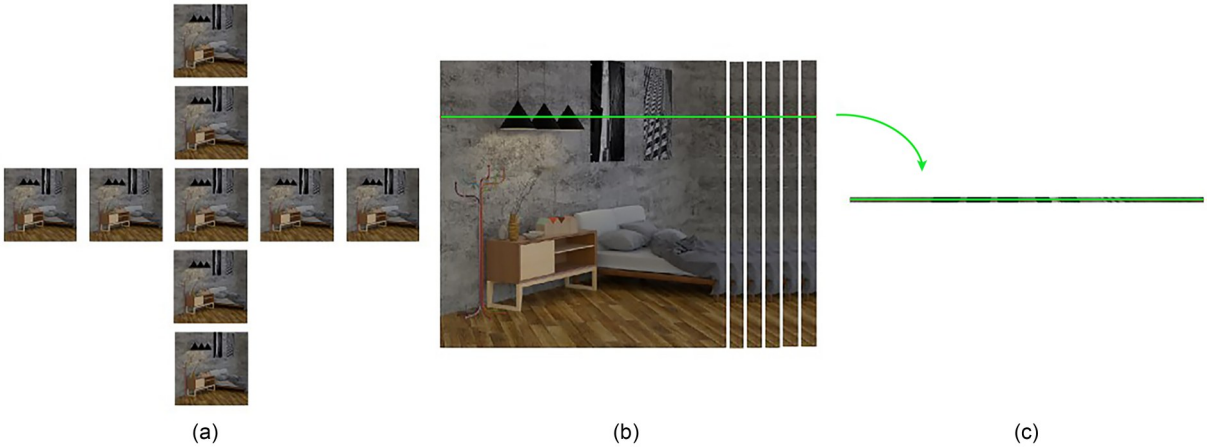

Fig. 5 The 4D LF and EPI: (a) a subaperture image; (b) LF sub-view $I_{u^*,v^*}(s,t)$; (c) EPI

3.2.1 Multi-camera array structure

Multi-camera array structure means that multiple cameras are placed at different viewing angles to capture an image of same object and obtain the LF information of different viewpoints. Levoy and Hanrahan (1996) installed a camera on an LF gantry to acquire complete LF information, including the four degrees of freedom of 2D translation and 2D rotation. Yang (2000) used an 8×11 lens array to conduct multi-perspective imaging of a target, and then used a flat plate scanner to complete a scan of the transversal image plane to finish recording all of the LF information. However, these two LF acquisition devices are suitable only for LF acquisition of static objects. To solve this problem, Zhang C and Chen (2004) proposed a multi-camera array structure that can adjust the attitude independently. Each camera is fixed in a mobile structure unit, which can adjust independently in the horizontal direction and a 2D rotation direction. Wilburn et al. (2005) proposed several camera arrays with different configurations. By controlling the time accuracy and relative position accuracy of each camera, the LF can be processed accurately in time and space to obtain high-quality synthetic images (Fig. 6). At the same time, many studies have proven that increasing the camera array improves the viewing angle range of the imaging system. In addition, synthetic aperture imaging technology has more freedom and flexibility in focus selection and depth of field adjustment. It is applicable to many visual tasks, including recognition and classification.

Fig. 6 The different forms of camera arrays: (a) camera array with independent attitude adjustment; (b) Stanford camera array

Reprinted from Wilburn et al. (2005), Copyright 2005, with permission from ACM

3.2.2 Single camera optical element structure

The LF single camera optical element structure acquisition method refers to adding optical modulation elements to a single camera, and then changing the imaging structure to redistribute the four-bit LF inside the camera to a 2D plane. The optical path model of the first handheld LFC (Plenoptic 1.0) is shown in Fig. 7. The imaging device places a microlens array at the focal plane of the traditional camera,

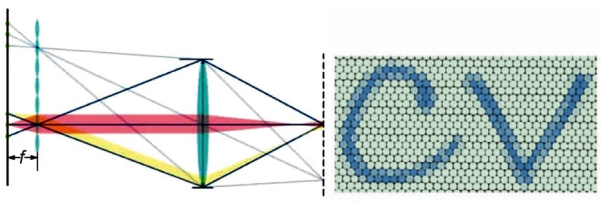

Fig. 7 Optical path model of the first handheld light field camera (Plenoptic 1.0)

Reprinted from Ng et al. (2005), Copyright 2005, with permission from the authors, licensed under CC BY-NC 4.0

and the image sensor is placed at one focal length from the microlens (Ng et al., 2005). Beams from different directions at the same point can be recorded by the image sensor through refraction of the main lens and focusing on the microlens array. The position resolution of the LF recorded by the Plenoptic 1.0 camera is related to the number of microlenses. The angular resolution of the LF is equal to the pixel resolution of the LF. At the same time, the imaging performance shows that each subaperture image is consistent with the aperture shape of the main lens and the regularly arranged sub-images.

Lumsdaine and Georgiev (2009) proposed the Plenoptic 2.0 LFC. The optical path model of the camera is shown in Fig. 8. The largest difference from Plenoptic 1.0 is that the camera places microlens arrays before and after the focal plane position, and the image sensor is placed at a certain distance behind the microlens array for the image sensor to realize early focusing or secondary focusing. The pre-focused sub-image is a positive image, and the secondary focused image is an inverted image. Each sub-image represents the observation at the imaging of the main lens. In addition, because the angular resolution of the LF imaging model is directly related to the relative distance of the main lens, microlens array, and image sensor, there is a compromise between angular resolution and position resolution.

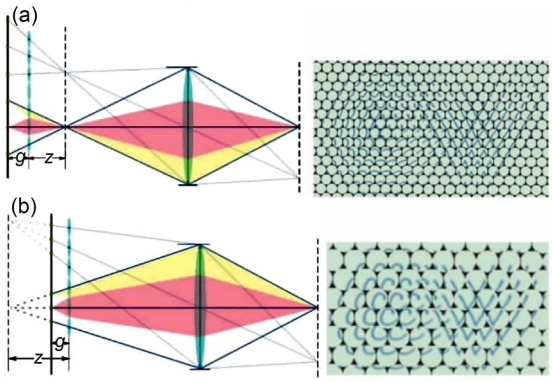

Fig. 8 The optical path model of the Plenoptic 2.0 LFC: (a) secondary focusing; (b) focus in advance
Reprinted from Lumsdaine and Georgiev (2009), Copyright 2009, with permission from IEEE

Liang et al. (2008) proposed a programmable aperture camera. As shown in Fig. 9, the camera samples the subaperture of the main lens through multiple

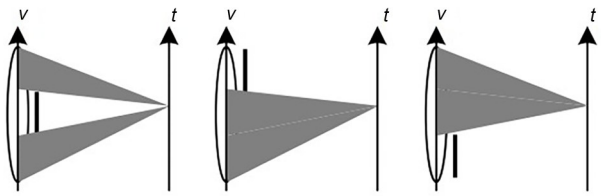

Fig. 9 The programmable aperture
Reprinted from Liang et al. (2008), Copyright 2008, with permission from ACM

exposures, and each exposure allows only the light at the specific subaperture position to be captured on the image sensor. For the selection of subapertures, a specific binary coding form is adopted. The LF collected by the camera has the same spatial resolution as the image sensor. However, this is at the expense of long exposure time and a high image signal-to-noise ratio, and the amount of additional data accrued by multiple exposures is an additional burden.

3.3 Light field representation

The most widely used consumer LFC, the Lytro, is composed mainly of a main lens, a microlens array, and a photosensor. A microlens array is placed between the main lens and a photosensor (Fig. 10) (Zhang J et al., 2020). This microlens array structure can divide the incident light rays in the main lens into many small parts in space. At the same time, each microlens can carry out independent projection transformation, and each part is focused on the focal plane by the corresponding microlens to obtain the corresponding microlens image. Finally, a series of microlens images are represented as LF images.

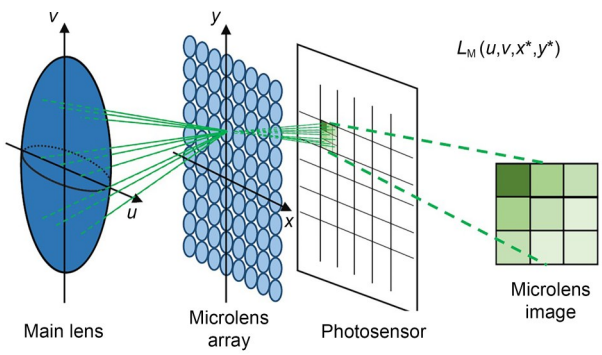

Fig. 10 Optical schematic of an LFC containing a main lens, a microlens array, and a photosensor
Reprinted from Zhang J et al. (2020), Copyright 2020, with permission from IEEE

As shown in Fig. 11, the generation of all LF images results from the integration and summation of all light ray information recorded on the refocus plane:

$$E_F(s,t) = \frac{1}{F^2}\iint L_F(u,v,s,t)\,\mathrm{d}u\mathrm{d}v, \tag{1}$$

where $E_F(s,t)$ represents the radiation intensity of pixel (s,t), $L_F(u,v,s,t)$ denotes the light ray information, and F represents the distance between the lens plane and the sensor plane.

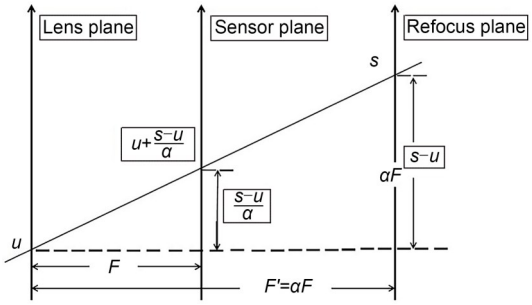

Fig. 11 Schematic of LF refocusing

It is assumed that the intersection of the ray propagating in a certain direction and the refocus plane is (s, t). Based on the geometric relationship, the intersection of the ray and the sensor plane is $\left(u, v, u + \frac{s-u}{\alpha}, v + \frac{t-v}{\alpha}\right)$. Because the radiation energy remains unchanged during propagation under the same light, the following can be obtained:

$$L_F(u,v,s,t) = L_F\left(u,v,u + \frac{s-u}{\alpha}, v + \frac{t-v}{\alpha}\right). \tag{2}$$

Therefore, we have

$$L_F(u,v,s,t) = L_F\left(u,v,u\left(1-\frac{1}{\alpha}\right)+\frac{s}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{t}{\alpha}\right). \tag{3}$$

When Eq. (3) is introduced into Eq. (1), the section image of the refocus plane can be obtained:

$$E_{\alpha F}(s,t) = \frac{1}{(\alpha F)^2}\iint L_F\left(u,v,u\left(1-\frac{1}{\alpha}\right)+\frac{s}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{t}{\alpha}\right)\mathrm{d}u\mathrm{d}v, \tag{4}$$

where u represents the lens plane, $F'=\alpha F$ indicates the distance between the lens plane and the refocus plane, and α refers to the focusing coefficient for the distance between the lens plane and the refocus plane that needs to be adjusted. Different refocused images can be obtained by changing the focusing coefficient.

4 Light field imaging processing

4.1 Digital refocusing

Ng et al. (2005) proposed the first handheld LFC that can record the complete 4D LF information. They suggested that each focus image based on a microlens array is essentially a 2D slice representation of the acquired 4D LF information. Furthermore, they demonstrated digital refocusing and viewpoint manipulation, coupled with custom software (Fig. 12). The experimental examples included human portraits and high-speed action. In addition, Lumsdaine and Georgiev (2009) proposed a new expression of LF. They regarded the microlens array as a separate imaging system. The focal plane of the imaging system was located on the main lens, and the final LF data had a higher spatial resolution. Georgiev and Lumsdaine (2010) proposed a real-time algorithm for synthesizing full screen refocused images from different viewpoints. Fiss et al. (2014) simplified the solution mode of the digital refocusing problem and directly projected the original information captured by the LFC to the focus output plane. Benefiting from the characteristics of LFCs with a large depth of field, Guo XQ et al. (2015) proposed a barcode image capture method that can refocus at different depths and demonstrated the technology in practical applications.

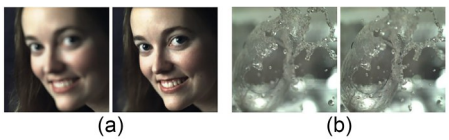

Fig. 12 Reformed photographs: (a) refocused human portraits; (b) refocused high-speed action

Reprinted from Ng et al. (2005), Copyright 2005, with permission from the authors, licensed under CC BY-NC 4.0

4.2 Image blending and fusion

Image blending technology refers to the process of producing a seamless high-resolution image using

multiple overlapping images. Traditional image blending technology includes the following three main parts: image pre-processing, image registration, and image fusion. Traditional 2D image blending technology has been greatly developed, but most techniques are complex and have a blending parallax in the vertical direction.

Afshari et al. (2012) introduced a new linear blending technique based on spherical LFC. The spherical LFC was applied to provide multiangle information for different focal planes to improve the effectiveness of visual reconstruction. Due to the limitation of sensor resolution, the LFC often performs sparse sampling in the spatial or angular domain. Kalantari et al. (2016) introduced a novel convolutional neural network (CNN) based framework to synthesize new views from a sparse set of input views. Raghavendra et al. (2013b) proposed a new weighted image fusion framework, which can adaptively assign a higher weight to the better-focused image to ensure the image fusion effect. Wang YQ et al. (2018) considered that the traditional fusion algorithm would cause color distortion when fusing multifocus images with a small depth of field. Therefore, they applied the wavelet transform method to multifocus image fusion, effectively ensuring visual performance. Nian and Jung (2019) introduced a machine learning based multi-focus image fusion scheme using LF data (Fig. 13). First, they used CNNs to extract the features of each multi-focus image. Second, they trained the network parameters of the multi-focus images together to ensure consistency of features. Then, they put the features to the second CNN to obtain an initial focus map. Finally, they performed morphological operations of opening and closing to process the initial focus map and then obtained the focus map. Based on the processed focus map, a full clear image was generated.

4.3 Image segmentation

Image segmentation divides the image into several regions according to features such as grey, color, spatial texture, and geometry. These features show consistency or similarity in the same region, but obvious differences between regions. Unlike traditional multiple view segmentation methods, the use of LF images helps improve the accuracy of segmentation due to the richness of data and the high correlation between multiple views (Wu et al., 2017).

Berent and Dragotti (2007) proposed an EPI-based energy minimization method for the occlusion segmentation problem. Similarly, Sheng et al. (2016) modified the structure tensor on the EPIs to compute the disparity, and presented a remarkable segmentation model based on superpixel segmentation, as well as a graph-cut algorithm. Using the graph-cut theory and Markov random field (MRF) framework, and analyzing multiple views, Campbell et al. (2010, 2011) proposed two automatic segmentation algorithms in the voxel space and image space. Hog et al. (2016) introduced a new graph representation for interactive LF segmentation using MRF. The proposed method exploited the redundancy in the ray space, decreasing the running time for the MRF process. Moreover, for multilabel segmentation, Wanner et al. (2013b) proposed a variable framework based on analyzing the ray space of 4D LFs. They also proved that LF data can not only be trained, but also provide available data for labelling. Based on the analysis of depth information and the redundancy contained in LF data, Mihara et al. (2016) proposed a supervised 4D LF segmentation method. Considering that the rays are the basic unit of each pixel, Zhu et al. (2017) presented a light field superpixel (LFSP) method, which can effectively solve the ambiguity of the segmentation boundary.

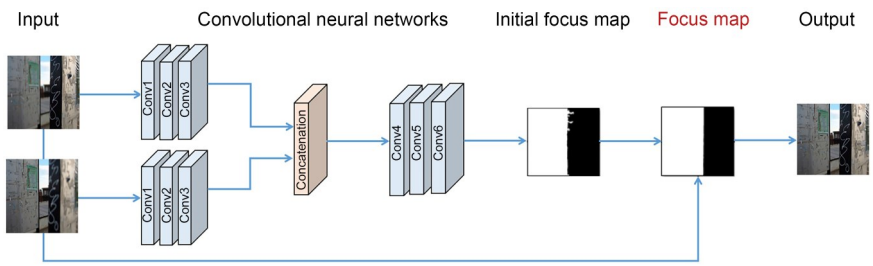

Fig. 13 Machine learning based multi-focus image fusion
Reprinted from Nian and Jung (2019), Copyright 2019, with permission from IEEE

Yücer et al. (2016) used densely sampled images from different angles as input, and presented an available automatic segmentation method for 2D and 3D space. The depth information contained in the LF focus region is easily accessible and is helpful for segmentation. Gao et al. (2017) introduced pulse coupled neural networks (PCNNs) for precise and straightforward segmentation. Furthermore, to address the background sensitivity problem in foreground segmentation, Chen et al. (2015) presented an automatic foreground segmentation algorithm relying on LF all-in-focus images. Lee and Park (2017) introduced an approach separating the foreground and background using gradient information directly at the pixel level, which can be effectively applied to the processing of occlusion problems. To address the problem of inaccurate segmentation of transparent objects, Johannsen et al. (2015) presented a method applying disparity and luminance information from LF data to solve the problem of reflective or transparent surface segmentation. Xu et al. (2015) proposed a segmentation framework for transparent objects, named TransCut. This method uses the LF-linearity and occlusion detector rather than color and texture information to describe transparent objects. Moreover, Xu et al. (2019) updated the LF-linearity algorithm and proposed a new dataset for evaluation to obtain better results. The new algorithm is automatic, requiring no human interaction. Lv et al. (2021) introduced LFSP technology, which improved the accuracy of occlusion boundary area segmentation.

4.4 Light field saliency detection

The main purpose of saliency detection is to detect the most interesting objects in an area. However, in the face of complex scenes such as those with similar background color and depth, the existing detection algorithms for 2D images cannot achieve accurate detection results.

Li NY et al. (2014) first attempted to use LFs as input for the saliency detection problem and developed the first saliency detection algorithm for LFs. The algorithm substantially improves the performance of saliency detection in complex scenes by calculating the image space and the structural similarity between focal stack images, as well as using the prior candidate knowledge of foreground and background. Furthermore, they used the Lytro LFC to construct a light field saliency dataset (LFSD), including the original 4D data and ground truth. The scenes in this dataset were divided into indoor and outdoor parts. Moreover, Li NY et al. (2015) processed LF depth information and focus information as one-dimensional visual features, proposing a single saliency detection framework with strong robustness in multi-dimensional data. Zhang J et al. (2015) proposed a novel saliency detection model based on analyzing all LF cues to improve their accuracy, and this proved to be beneficial for 2D and 3D saliency detection. Taking the accuracy of saliency detection as an evaluation index, Zhang XD et al. (2015) compared LF and 2D saliency detection. Experimental results showed that LF saliency detection outperformed 2D saliency detection in complex occlusions and background clutter. Wang AZ et al. (2017) fused the useful visual cues contained in LF data and proposed an improved Bayesian integration framework for saliency detection. Zhang J et al. (2017) proposed an LF multiple cues based saliency detection framework, with cues including color and depth. Meanwhile, they established a real data LF dataset, Hefei University of Technology-light field saliency detection (HFUT-LFSD). Zhang M et al. (2019) exploited LF combined with long short term memory (LSTM) and constructed the largest LF dataset, Dalian University of Technology-light field saliency detection (DUT-LFSD). Piao et al. (2019a) fully exploited the correlation of the information inherent in LF data and proposed a saliency detection framework based on depth-induced cellular automata (DCA). The framework can effectively solve the main problems of missed and false detection under complex background conditions. The main flow of the framework is shown in Fig. 14. First, the focal stack, depth map, and all-in-focus images are used as the overall input of the framework. Second, the focal stack and depth map are used to guide and obtain object-guided depth and background seeds. Meanwhile, all-in-focus images and object-guided depths are guided for segmentation. Third, the depth-induced saliency is obtained by multiplying the object-guided depth and the contrast saliency. At the same time, a DCA model is used to optimize the parameters, and optimized saliency results are obtained. Finally, the Bayesian framework is used to fuse depth-induced saliency

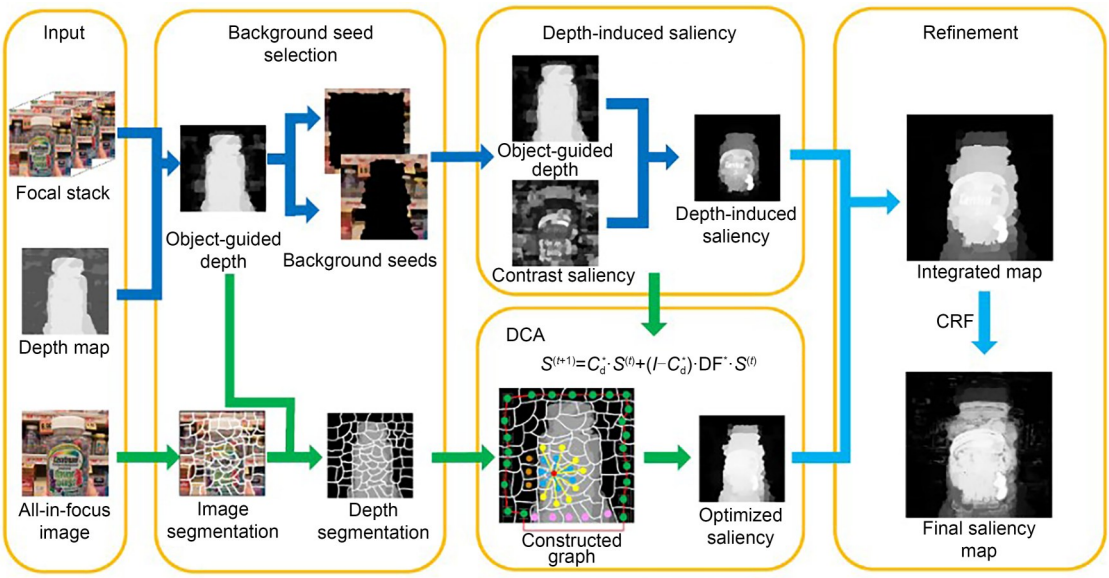

Fig. 14 The framework of depth-induced cellular automata for saliency detection
Reprinted from Piao et al. (2019a), Copyright 2019, with permission from IEEE

and optimized saliency while conditional random field (CRF) is used to obtain the final saliency map. To explore the contribution of each focal slice to saliency detection, Piao et al. (2021) proposed a patch aware network (PANet) with a multisource learning module and achieved optimal results at the same time.

Although some LF saliency detection datasets have been established, most have limitations. Therefore, deep learning methods based on big data cannot perform well; thus, the advantage of the rich LF image information cannot be demonstrated. To solve this problem, Wang TT et al. (2019) introduced a large-scale dataset for 4D LF saliency detection. The dataset consisted of 1465 all-in-focus images. Each image had a well-labelled ground truth and focal stacks. In addition, they introduced a fusion framework based on an attentive recurrent CNN, which improves the performance for 4D LF saliency detection. Furthermore, Piao et al. (2019b) used the Lytro LFC to construct a new 4D LF dataset for saliency detection with 1580 LF images. Each LF image consisted of multiple views and a pixelwise ground truth of the central view. Moreover, they presented an available end-to-end CNN scheme based on LF synthesis and LF-driven saliency detection to ensure the improvement in the final experimental results. In addition, Zhang J et al. (2020) first proposed to analyze the angular features with a CNN from an LF image for saliency detection. Finally, to compare the performance of the algorithms comprehensively using three datasets, we adopted four popular metrics, including the S_α-measure, E_s-measure, F_β-measure, and mean absolute error (MAE). The results of quantitative comparisons are shown in Table 1.

The S_α-measure can evaluate the region-aware and object-aware structural similarity between the saliency map and ground truth. The definition of S_α is given as

$$S_\alpha = (1-\alpha)S_{\mathrm{o}}\left(S_{\mathrm{p}},G\right) + \alpha S_{\mathrm{r}}\left(S_{\mathrm{p}},G\right), \tag{5}$$

where S_{o} is the final object-aware structural similarity measure, S_{r} is the final region-aware structural similarity measure, $\alpha\in[0,1]$ is a balance factor between S_{o} and S_{r}, G is the ground truth map, and S_{p} is the prediction map.

The E_s-measure can capture local pixel matching information and image-level statistics. The definition of E_s is given as

$$E_s = \frac{1}{w\times h}\sum_{x}^{w}\sum_{y}^{h} s\left(S_{\mathrm{p}}(x,y),G(x,y)\right), \tag{6}$$

where w and h are the width and height of G respectively, symbol s is the enhanced alignment matrix, and (x,y) are the coordinates of each pixel.

Table 1 Quantitative comparisons of the E_s-measure, S_α-measure, F_β-measure, and MAE score on the 4D LF dataset

Algorithm	DUT-LFSD				HFUT-LFSD				LFSD			
	E_s ↑	S_α ↑	F_β ↑	MAE ↓	E_s ↑	S_α ↑	F_β ↑	MAE ↓	E_s ↑	S_α ↑	F_β ↑	MAE ↓
Zhang J et al. (2017)	–	–	–	–	–	–	–	–	0.841	0.749	0.815	0.150
Wang YQ et al. (2018)	0.913	0.878	0.833	0.055	0.770	0.736	0.620	0.096	0.882	0.820	0.805	0.092
Zhang M et al. (2019)	0.923	0.887	0.843	0.053	0.785	0.752	0.627	0.095	0.886	0.830	0.819	0.089
Wang TT et al. (2019)	0.905	0.852	–	0.070	–	–	–	–	0.877	0.826	–	0.093
Piao et al. (2019a)	0.891	0.841	0.801	0.076	0.783	0.741	0.615	0.098	0.806	0.737	0.715	0.147
Piao et al. (2021)	0.941	0.897	0.892	0.042	0.843	0.802	0.704	0.073	0.882	0.842	0.853	0.080

↑ The higher the index, the better the performance. ↓ The lower the index, the better the performance

The F_β-measure is the harmonic mean of the average precision and average recall. It is defined as

$$F_\beta = \frac{\left(1+\beta^2\right)\text{precision}\cdot\text{recall}}{\beta^2\cdot\text{precision}+\text{recall}}, \tag{7}$$

where β is a parameter to achieve a trade-off between recall and precision.

The MAE is the average difference between S_p and G. The definition of MAE is given as

$$\text{MAE} = \frac{1}{w\times h}\sum_{x}^{w}\sum_{y}^{h}\left|S_\mathrm{p}\left(x,y\right)-G\left(x,y\right)\right|. \tag{8}$$

4.5 Light field video

Traditional 2D video equipment cannot express a scene's real stereo effect because of the limitations of imaging sensors. LF imaging technology can provide complete LF video information, including the required six degrees of freedom for natural scenes.

In early research, Wilburn et al. (2002) proposed a flexible and modular LF video camera. This LF video camera is composed mainly of more than 100 controllable 2D cameras, and visual control is realized mainly by one computer at the same time. To solve the problem of a shaky output video of a handheld LF video camera, Smith et al. (2009) used the space-time optimization method to optimize the objects while enabling the important image features in the input video to move smoothly in the output video. Meanwhile, they used LF video to provide the characteristics of multiple views and proposed an LF video optimization method without path reconstruction. Furthermore, to improve LF data capture, processing, and display, Balogh and Kovács (2010) proposed an end-to-end LF video rendering processing software. The software includes multiple cameras, a high-performance personal computer, and a high-speed network. The LF rendering video processed by this system is real-time, and has a high frame number and high resolution. In particular, Tambe et al. (2013) broke the trade-off between spatial resolution and angular resolution due to the prior knowledge of LFCs, and introduced an LF video camera for dynamic high-resolution LF video reconstruction using redundant scene information. Furthermore, Sabater et al. (2017) proposed a real-time, accurate pipeline for LF video capture and processing, including depth estimation algorithms and color homogenisation algorithms.

Wang TC et al. (2017) proposed a hybrid imaging system using a Lytro camera and a digital video camera. A digital camera was used to capture the temporal information, solving the problem of missing a large amount of information between adjacent frames. The Lytro camera was first combined with a 30-frame/s digital video camera (Fig. 15a). The inputs in the system included a standard 2D video and a 3-frame/s LF sequence (Fig. 15b). Finally, combining the angular information and temporal information obtained from the 3-frame/s LF sequence and 30-frame/s 2D video, a complete LF video with all angular views could be obtained (Fig. 15c, left). This system implements digital focus and parallax generation during video playback. Mehajabin et al. (2020) proposed a novel pseudo-sequence-based coding order, which efficiently improves compression for LF videos.

4.6 Light field super-resolution

As a piece of new imaging equipment, the LFC has a trade-off between spatial resolution and angular

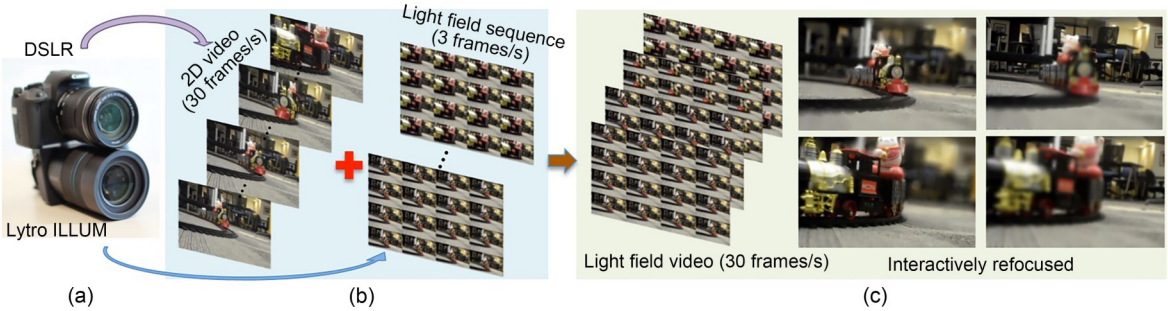

Fig. 15 The LF video hybrid imaging system: (a) system setup; (b) system input; (c) system output
Reprinted from Wang TC et al. (2017), Copyright 2017, with permission from ACM

resolution. The image resolution generated by the LFC is low, restricting the application of LF imaging technology (Cheng et al., 2019). Recently, to solve this problem, many super-resolution methods have been proposed, including three main categories: LF data structure based methods, learning-based methods, and multisensor-based methods (see supplementary materials, Sections 1.1.1, 1.1.2, and 1.1.3).

4.7 Depth estimation

The rich 4D LF data make LF depth estimation research possible. The LFC is a single-sensor imaging device. Data acquisition is not affected by other sensors operating in cooperation, which gives it stability and convenience. In our summary, we divided the methods of LF depth estimation into two categories: data-based methods and learning-based methods. The data-based methods can be divided further into two parts: EPI-based methods and LF-image-based methods. LF-image-based methods include the measurement analysis of the subaperture image and the refocus image. Furthermore, they include some processing methods based on LF data for some unique problems, such as occlusion and scattering medium (see supplementary materials, Sections 1.2.1 and 1.2.2).

4.8 Light field image quality assessment

Many LF image processing methods and applications have been proposed. LF images inevitably struggle with various wide distortions, leading to a reduction in image quality. Therefore, the quality assessment of LF images is critical, and can better guide the collection, processing, and application of LF images. In recent decades, various image quality assessment (IQA) algorithm models have been proposed, but these are used mostly to evaluate natural images and screen images. Because different image types have different characteristics, they are not suitable for direct assessment of LF images. Therefore, accurately evaluating LF image quality and efficiency according to visual characteristics becomes an urgent need.

Fang et al. (2018) presented a full reference LF image quality assessment (FRLFIQA) algorithm. This method predicts the LF image quality by measuring the gradient similarity between the EPI of the original LF image and the EPI of the distorted image. Huang et al. (2018) proposed another FRLFIQA algorithm. This algorithm estimates the LF image quality by fitting the distribution of the average difference between each perspective of the original image and the distorted image. Furthermore, Tian et al. (2021) proposed a symmetry and depth feature based model for LFIQA. The main idea was to use symmetry and depth features to fully search the color and geometric information. In addition, Paudyal et al. (2019) proposed a reduced-reference LFIQA (RR LFIQA) algorithm. This algorithm is based on the correlation between the depth image quality of the LF image and the overall quality of the LF image. It measures the structural similarity between the depth image of the original image and that of the distorted image to predict the overall quality of the LF image. Furthermore, considering that existing LF image quality evaluation is closely related to the reliability and correlation of LF datasets, Paudyal et al. (2017) built an LF image quality dataset composed mainly of the original data, compressed images, and annotation information. Shi et al. (2018) built an LF image dataset containing five degrees of freedom that can provide depth cues and the most realistic LF display. At the same time, they used the dataset as experimental

samples to verify the effect of the inherent LF properties on the quality of the LF image.

The methods by Fang et al. (2018), Huang et al. (2018), Tian et al. (2021), and Paudyal et al. (2019) need to use the original image information, which requires additional bits to transmit the information of the original LF image. Therefore, in the past two years, no-reference LF quality evaluation methods that fully consider the influence factors of LF image quality have been widely studied. Based on tensor theory, Shi et al. (2019) explored the LF 4D structure characteristics and proposed the first blind quality evaluator of the LF image. Considering the influence of angle consistency, chrominance, and luminance on LF image quality, Zhou W et al. (2020) proposed a novel LF image quality evaluation model, called the tensor-oriented no-reference LF image quality evaluator. To reduce the impact of LF image quality degradation, Shi et al. (2020) measured EPI in the form of local binary pattern (LBP) features to obtain angular consistency features, which were used as the basis to address the problem of LF image quality degradation. Shan et al. (2019) constructed an LF image quality assignment dataset with a total of 240 samples by combining 4D decoding technology and human subjective feelings. At the same time, they used a support vector regression (SVR) model to represent the 2D and 3D characteristics of the LF image and proposed a no-reference IQA metric. Meng et al. (2019) proposed an accurate, time-efficient LFIQA framework based on the fact that the degree of distortion of LF refocusing images can accurately reflect the whole LF image quality. Cui et al. (2021) proposed a macro-pixel differential operation method based on spatial-angular characteristics to quantify the LF image quality.

4.9 Three-dimensional reconstruction

LF information containing multiple data types has great potential in 3D reconstruction. Levoy et al. (2006) applied the theory of LF rendering to microscopic imaging and developed an LF microscope (LFM), which can obtain object images of different depths in an exposure and carry out 3D reconstruction of the object. Broxton et al. (2013) proposed a 3D deconvolution method for 3D reconstruction by decoding dense spatial-angular sampling. At the same time, they demonstrated by experiment the high-resolution characteristics of this method for object reconstruction. Murgia et al. (2015) proposed a novel algorithm that can reconstruct a 3D point cloud from a single LF image through a sequential combination of image fusion, feature extraction, and other technologies. Sun et al. (2017) fully optimized the calibration of focused LFCs based on the Levenberg-Marquardt algorithm and proposed a 3D reconstruction method for flame temperature based on the least-square QR-factorization algorithm. Marquez et al. (2020) proposed a tensor-based LF reconstruction algorithm, which shows better performance and lower computational complexity than matrix-based methods. Through the fusion of LF data and regression prediction, Cui et al. (2021) proposed a 3D reconstruction algorithm for bubble flow.

5 Tasks and applications

5.1 Iris recognition

Iris features are widely used in biometrics because of their uniqueness and non-replicability. However, most iris image collectors are affected by the limitation of a small depth of field, which leads to poor quality of the collected iris images, and inevitably affects recognition.

Recently, to address the above problems, the use of LFCs was explored for iris recognition. Zhang C et al. (2013) presented a new iris sensor based on LF photography and constructed the first LF iris image database using the sensor. Raghavendra et al. (2013a) leveraged the LFC and provided useful information in terms of multiple depths (Fig. 16). They collected a new iris and periocular biometric dataset, and then proposed a new scheme for iris and periocular recognition based on the LBP feature extraction algorithm and the sparse reconstruction classifier. In addition, to prevent spoof attacks on the biometric system, Raghavendra and Busch (2014) presented a novel way of addressing spoof detection by fully exploiting the depth and focus information of the LFC for visible spectrum iris biometric systems. Furthermore, because LFCs can hold additional information that is quite useful for biometric applications, Raghavendra et al. (2016) carried out an empirical

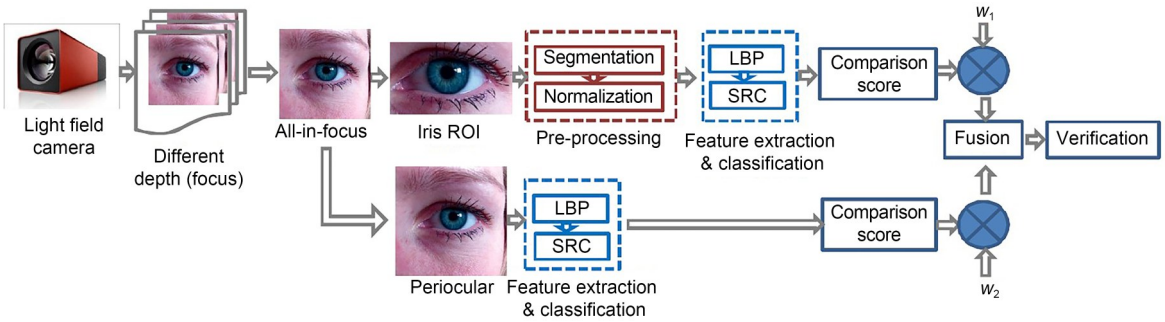

Fig. 16 The iris and periocular recognition scheme of Raghavendra et al. (2013a)
Reprinted from Raghavendra et al. (2013a), Copyright 2013, with permission from IEEE

study for iris recognition using an LFC. They first collected a new iris dataset in an unconstrained environment by simulating a real-life scenario, and then explored the supplementary information available from different depth images. These were rendered by the LFC by either choosing the best focus from multiple depth images, or exploiting the supplementary information by combining all of the depth images using super-resolution. Finally, iris recognition results with more than 90% reliability were reported from an extensive set of experiments.

5.2 Face recognition, detection, and a light field face dataset

As the most extensive biometric scheme, face recognition is vulnerable to biological attacks (Sepas-Moghaddam et al., 2018). There are two types of LF-based methods for face recognition: texture-based methods and focus and depth based methods (see supplementary materials, Sections 2.1.1 and 2.1.2).

5.3 Material recognition

Although there have been few studies of material recognition based on LF data, the approach has provided a wide range of ideas for related research. The direct measurement method has been widely used in traditional image material recognition research. The principle is to analyze multiple views of a point simultaneously. Wang TC et al. (2016a) proposed a 4D LF material recognition system to prove whether the multiple views obtained in LF data have a better effect than traditional 2D images in material recognition (Fig. 17). First, they used a Lytro LFC to collect a material recognition dataset containing 1200 images in 12 categories. Second, the images in the dataset were used as experimental samples to extract information from a total of 30 000 patches. Furthermore, this patch information was input to a specially designed CNN that can take the form of an LF as input and train the LF information to obtain a patch model. Finally, the patch model was fine-tuned in the full scene to obtain

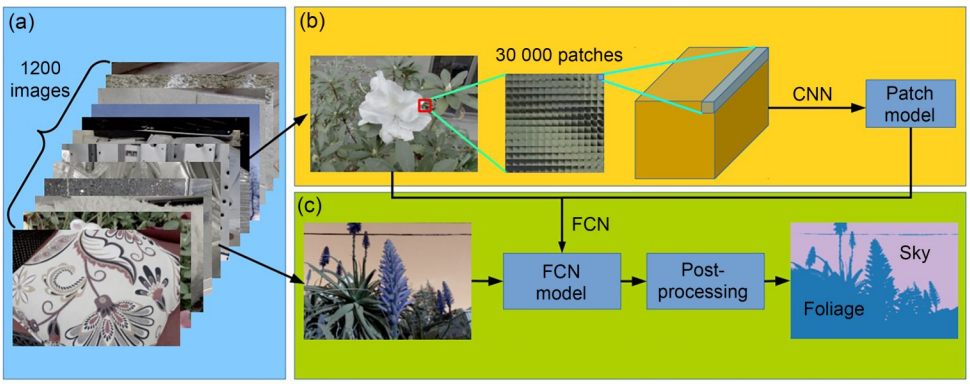

Fig. 17 The material recognition system based on 4D LF information: (a) light field dataset; (b) CNN training procedure for 4D light fields; (c) full scene material recognition
Reprinted from Wang TC et al. (2016a), Copyright 2016, with permission from Springer Nature

the fully convolutional network (FCN) model, and post-processing was carried out to obtain the final recognition results. Guo et al. (2020) proposed correspondences in the angular domain to achieve the decoupling of spatial-angular features for LF material recognition.

5.4 Detection on 2D printed photos

Many biological detection methods are based on conventional 2D imaging devices, and are at the mercy of imaging principles. Traditional 2D cameras will lose stereo information during scene information acquisition, thereby leading to a large number of 2D printed photos attacking the authentication program. The most substantial differences between the printed photo and the original scene photo are the gradients of epipolar lines and depth information. Compared with the original scene photo, the gradients of the epipolar lines of the printed photo remain unchanged, and the printed photo does not contain depth information.

Based on this, Ghasemi and Vetterli (2014) extracted an energy feature vector from epipolar lines, and then distinguished the printed photo from the original photo by distinguishing the feature vector. Similarly, through the analysis of viewpoint and light information of the subadjacent aperture image, Kim et al. (2013) proposed a spoof attack detection algorithm based on LBP and support vector machine (SVM). This algorithm not only can detect printed photos correctly, but also has good robustness for the detection of gradient objects. Considering that the system of pedestrian detection is easy to disturb by many 2D printed fake photos, Jia et al. (2018) constructed an LF pedestrian dataset including more than 1000 images. They proposed a 2D fake pedestrian detection framework based on LF imaging technology and an efficient SVM classifier. In addition, by analyzing the variation in the focus of depth images generated by LFC, Raghavendra et al. (2016) proposed a parallel spoof attack detection framework (Fig. 18). First, parallel face detection and pre-processing operations were carried out for the acquired images of different depths. The pre-processing operation was designed mainly to filter out image noise interference. Second, the focus measure operators were obtained by applying different focusing measurement methods. Third, the relative and absolute values of focus transformations were estimated. Finally, the SVM classification method was used to distinguish the relative value and absolute value.

5.5 Specular highlight removal

The removal of specular highlights from an image can effectively ensure image quality. In recent years, LFCs have been widely used to remove specular highlights.

Through the decomposition of the EPI in LF, Criminisi et al. (2005) found that EPI has a high regularity. Therefore, they proposed an EPI framework based on math characteristics for specular highlight removal. Tao et al. (2015a) took advantage of this ability to later modify the focus point in LF data, and proposed an iterative method for specular highlight removal. Furthermore, they first proposed an algorithm to determine the light source color by analyzing the angle pixels in LF data. Meanwhile, the specular highlight can be removed by analyzing the light source color. Wang HQ et al. (2016) proposed a combined specular highlight removal algorithm based on depth estimation and specularity detection. First, the depth map corresponding to the LF image with specularity was obtained by analyzing 4D EPI data. Then, each pixel of the image was classified as a saturated or unsaturated pixel using the threshold-based specularity detection method. Then the k-means

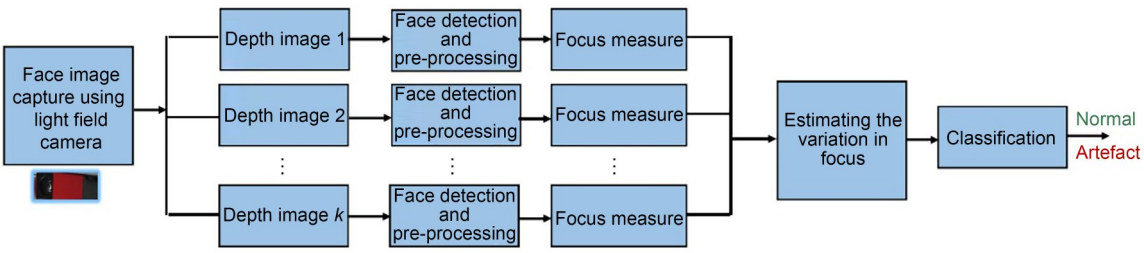

Fig. 18 A 2D printed face detection framework based on different focus images
Reprinted from Raghavendra et al. (2016), Copyright 2016, with permission from IEEE

algorithm was used for pixel clustering, and the depth map obtained before was used for refocusing and filtering unsaturated pixels. Finally, the local color refinement method was used to complete the color correction of some saturated pixels. In addition, based on the analysis of 4D LF data, Alperovich and Goldluecke (2017) proposed a variational model based on a review of additional data available in LF. The proposed model can remove specular highlights by separating the shadow and the albedo in the LF. Recent works (Gryaditskaya et al., 2016; Sulc et al., 2016) also analyzed the LF data structure or later edited the composition to better complete specular highlight removal.

5.6 Shape recovery

Effective and accurate acquisition of multiangle images of objects in a scene is the key to shape recovery. Completing an accurate shape recovery of crystals, ceramics, and other objects with glossy surfaces is difficult. In recent years, LF imaging technology has been widely used in shape recovery.

Tao et al. (2015a) presented a new photo consistency metric, line-consistency, for shape recovery, revealing how viewpoint changes affect specular points. Furthermore, they made full use of the defocus, correspondence cues, and shading in LF data. They used defocus and correspondence cues for local shape recovery and shading to further determine the depth of the object and improve the accuracy of the final shape recovery. Wang TC et al. (2016b) proposed a bidirectional reflectance distribution function (BRDF) invariant theory and invariant equation suitable for LF data analysis, which can effectively complete the shape recovery of different objects. The experimental results for the methods report the excellent effect on shape recovery. However, Li ZQ et al. (2017) found that Tao et al. (2015a) failed to consider that texture gradient information may play a complementary role in shape restoration, and may distort the shape recovery of objects for non-Lambertian surfaces. They proposed a general energy minimization formulation with autobalance ability, which can effectively recover the shape of complex BRDFs using only one LF image. Experimental results from a large number of studies show that this formulation is more accurate and robust than BRDF invariant reconstruction methods. In addition, Zhou MY et al. (2020) proposed a fixed-structure concentric multispectra LF (CMSLF) acquisition system. In this system, each concentric circle has a fixed number of cameras to acquire the information of multiple views. The proposed system was highly effective for shape recovery under non-Lambertian conditions based on the dichromatic Phong reflectance model. A comparison of the experimental results with state-of-the-art methods is shown in Fig. 19.

5.7 Light field microscopy

LF microscopy (LFM) is a new 4D LF imaging technology that uses microlenses to achieve a balance between spatial resolution and angular resolution. The application of this technology can make weakly scattering medium or fluorescence specimens achieve the effect of high-speed imaging without scanning. Levoy et al. (2006) demonstrated a prototype LFM (Fig. 20), composed of a high resolution

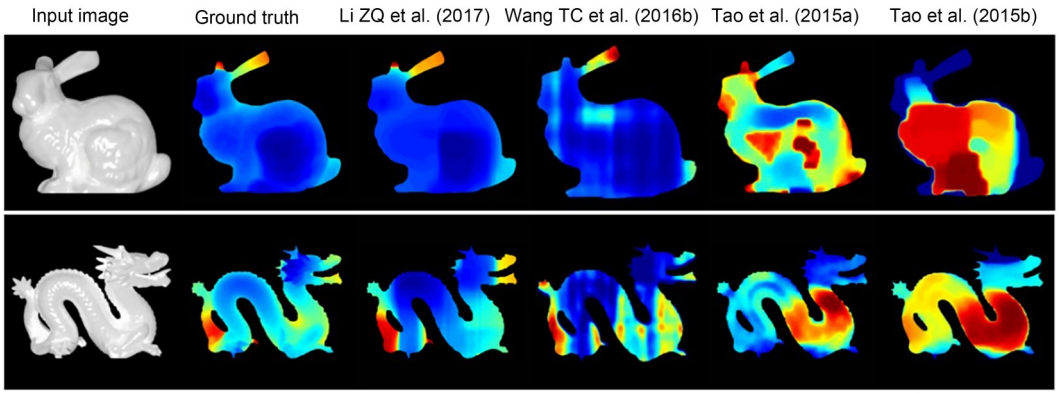

Fig. 19 Comparison of the experimental results from state-of-the-art methods
Reprinted from Zhou MY et al. (2020), Copyright 2020, with permission from IEEE

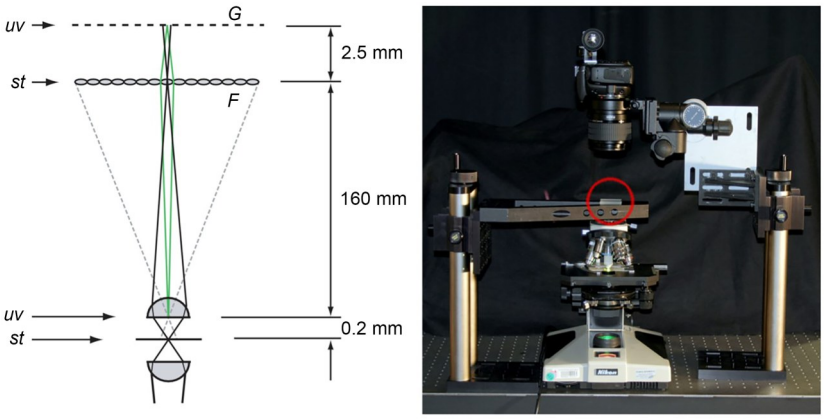

Fig. 20 The first prototype light field microscopy
Reprinted from Levoy et al. (2006), Copyright 2006, with permission from ACM SIGGRAPH

Nikon 2D camera and a custom microlens array. The camera was placed at *G*, and the microlens array at *F*. Each microlens array recorded the relevant, focused image. They analyzed the optical performance of this method and showed its advantages for capturing the 3D structure of micro-objects.

In addition, an optical model for LFM and a 3D deconvolution method for LFM were proposed by Broxton et al. (2013). Experimental results showed that this model could help obtain a higher spatial resolution with better optical sectioning in reconstructing volumes. However, the spatial resolution reconstructed by the LFM was nonuniform in depth information. The resolution at most imaging centers was low, while the resolution at the *z* plane was high. Wavefront coding techniques were applied by Cohen et al. (2014) to address this nonuniform resolution limitation by including phase masks in the optical path of the microscope. The performance of the LFM was improved. Furthermore, to complete the observation of in vivo specimens, a new 3D LFM with a high spatial resolution was proposed by Zhang M et al. (2017) by adding a microlens array and relevant focusing optical elements to a traditional microscope. The experimental results showed that the 3D LFM can acquire seven subaperture images from different perspectives. To further expand the working range of the LFM, Hsieh et al. (2018) proposed a system based on a multifocal high-resistance liquid crystal microlens, which has advantages such as short response time and low driving voltage. Vizcaíno et al. (2021) applied fully CNN architecture to reconstruct configurable microscopic stacks from single LF images, which improved the time efficiency and reconstruction accuracy.

6 Existing problems and future trends

Although LF imaging technology has been widely studied and applied in computer vision, specific research is still lacking. In combination with current trending topics, we believe that researchers should make efforts in the following areas.

1. Establishment and evaluation of the LF dataset

LF dataset construction is essential for the application of LF imaging technology to computer vision tasks. Kim et al. (2013) constructed an LF dataset for scene reconstruction from a high spatial-angular resolution. Wanner et al. (2013a) constructed an LF benchmark dataset by densely sampling 4D LF data, Heidelberg Collaboratory for image processing1 (HCI1). Li NY et al. (2014) created an LFSD. Raghavendra et al. (2016) constructed an LF dataset for face and iris recognition. Tao et al. (2015a) built an LF dataset for depth estimation. Wang TC et al. (2015, 2016a) constructed an LF dataset for material recognition and depth estimation under occlusion. Other kinds of LF datasets have been proposed by Paudyal et al. (2016) and Rerabek and Ebrahimi (2016), but there was a notable lack of large LF datasets for computer vision in previous studies. The datasets of Kim et al. (2013), Tao et al. (2015a), and Wang TC et al. (2015) each had fewer than 10 images. Moreover, part of the LF dataset was not representative and proved only

the rationality. Therefore, the establishment of rich data and a wide range of LF datasets is a priority for future research.

2. Applications under high dynamic range (HDR) conditions: unmanned aerial vehicles (UAVs), unmanned vehicles, unmanned boats

The key to the application of UAVs, unmanned vehicles, and unmanned boats is that the imaging equipment must accurately obtain the scene information under highly dynamic conditions for summary analysis. When LFC is applied, it can accurately obtain the complete 4D LF information of the scene under conditions of high-speed movement. Simultaneously, through the analysis and processing of 4D LF information by software, clear images of different focal planes can be obtained. However, because of the high dimensionality of 4D LF and technical limitations, the 4D LF information collected by LFC cannot be analyzed in real time, and the imaging effect is considerably reduced in HDR conditions. The HDR conditions refer to areas with high brightness under strong light sources (e.g., sunlight, lamps, or reflection) and areas with low brightness such as shadows and backlight. Therefore, leveraging LFC imaging, solving the real-time problems of software and hardware, and limiting the use of LFC under complex conditions are the key issues.

3. Light field image enhancement

LFC will be polluted by various kinds of noise in image acquisition, transmission, and storage, which will affect the quality of the LF image. Reducing the influence of noise in the acquisition of LF images and enhancing the known LF image to improve the visual expression effect are important aims for future research.

4. Virtual reality, 3D displays, and 3D movies

Accurate and efficient data acquisition in dynamic environments is the core content of virtual reality, 3D displays, and 3D movies. However, the LFC has poor applicability in harsh conditions, such as overexposure, dim light, or occlusion. Widely used LFCs, such as the consumer-Lytro or industrial-Raytrix, which cannot meet the shooting effect higher than 30 frames/s, limit the development of 3D computer vision tasks. Therefore, the development of strategies to compensate for the shortcomings of LFCs while making full use of 4D LF information is a future trend.

5. Military optical camouflage technology

In military optical camouflage research, the goal is to reduce the optical differences between the target and the surrounding background as much as possible, making it difficult for reconnaissance instruments to detect and distinguish. Therefore, the analysis and application of 4D LF information to improve military optical camouflage technology is a future trend.

6. Image recognition at the micro-scale

LFC can obtain the complete 4D LF information of an object at the expense of losing part of the spatial resolution, restricting the development of image recognition at the micro-scale. Ensuring the integrity of 4D information acquisition without losing the image resolution at the micro-scale is still an urgent problem to resolve.

7. Image processing method based on HDR

Because the luminance contrast of the real visual environment is far beyond the limits of the dynamic range of image sensors, the performance of LFCs is weak when there is a limited dynamic range, such as with overexposure, strong light, or dim light. Therefore, using LFC to capture 4D LF information with an HDR condition is an urgent problem.

8. Optimal relationship between spatial resolution and 4D light field information acquisition

The image acquired by an LFC usually has a low spatial resolution. If the spatial image resolution is improved considering the axial resolution, higher requirements are needed for the performance of the photodetection device. Obtaining complete 4D LF information without losing spatial image resolution is one of the most vital aims of LF imaging research. In addition, increased storage capacity and processor speed are required because a huge amount of data will be obtained in one exposure. Based on the above analysis, the application of LF imaging technology in computer vision research still needs to resolve the following issues: the specific implementation, the balance between the software process ability and hardware shooting effect, commercialization costs, and convenience.

7 Conclusions

Light field imaging technology aims to establish the relationship among light information in spatial

domain, visual angle, spectrum and time domain, and realize coupling sensing, decoupling reconstruction, and intelligent processing.

In this paper, we have summarized the importance of light field imaging technology for computer vision tasks and listed representative contributions from all interested researchers. The representative studies were focused mainly on depth estimation, image segmentation, saliency detection on light field, light field image quality assessment, image blending, fusion, face recognition, and light field super-resolution. Applying spatial information, angular information, and epipolar plane image information in the light field to computer vision tasks was also investigated. Benefiting from the progress of software and hardware, light field cameras are gradually applied to industrial detection, unmanned systems, and virtual reality fields. It has significant academic value and industrial application prospects.

Contributors

Chen JIA and Fan SHI investigated and summarized the literature. Chen JIA and Meng ZHAO drafted the paper. Chen JIA and Shengyong CHEN revised and finalized the paper.

Compliance with ethics guidelines

Chen JIA, Fan SHI, Meng ZHAO, and Shengyong CHEN declare that they have no conflict of interest.

References

Adelson EH, Bergen JR, 1991. The plenoptic function and the elements of early vision. In: Landy MS, Movshon JA (Eds.), Computational Models of Visual Processing. MIT Press, Cambridge, USA, p.3-20.

Afshari H, Akin A, Popovic V, et al., 2012. Real-time FPGA implementation of linear blending vision reconstruction algorithm using a spherical light field camera. Proc IEEE Workshop on Signal Processing Systems, p.49-54. https://doi.org/10.1109/SiPS.2012.49

Alperovich A, Goldluecke B, 2017. A variational model for intrinsic light field decomposition. Proc 13th Asian Conf on Computer Vision, p.66-82. https://doi.org/10.1007/978-3-319-54187-7_5

Balogh T, Kovács PT, 2010. Real-time 3D light field transmission. Proc SPIE 7724, Real-Time Image and Video Processing, Article 772406. https://doi.org/10.1117/12.854571

Berent J, Dragotti PL, 2007. Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition. Proc IEEE Workshop on Multimedia Signal Processing, p.182-185. https://doi.org/10.1109/MMSP.2006.285293

Broxton M, Grosenick L, Yang S, et al., 2013. Wave optics theory and 3-D deconvolution for the light field microscope. *Opt Expr*, 21(21):25418-25439. https://doi.org/10.1364/OE.21.025418

Campbell NDF, Vogiatzis G, Hernández C, et al., 2010. Automatic 3D object segmentation in multiple views using volumetric graph-cuts. *Image Vis Comput*, 28(1):14-25. https://doi.org/10.1016/j.imavis.2008.09.005

Campbell NDF, Vogiatzis G, Hernandez C, et al., 2011. Automatic object segmentation from calibrated images. Proc Conf for Visual Media Production, p.126-137. https://doi.org/10.1109/CVMP.2011.21

Chen XY, Dai F, Ma YK, et al., 2015. Automatic foreground segmentation using light field images. Proc Visual Communications and Image Processing, p.1-4. https://doi.org/10.1109/VCIP.2015.7457895

Cheng Z, Xiong ZW, Chen C, et al., 2019. Light field super-resolution: a benchmark. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops, p.1804-1813. https://doi.org/10.1109/CVPRW.2019.00231

Cohen N, Yang S, Andalman A, et al., 2014. Enhancing the performance of the light field microscope using wavefront coding. *Opt Expr*, 22(20):24817-24839. https://doi.org/10.1364/oe.22.024817

Criminisi A, Kang SB, Swaminathan R, et al., 2005. Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. *Comput Vis Image Underst*, 97(1):51-85. https://doi.org/10.1016/j.cviu.2004.06.001

Cui YL, Yu M, Jiang ZD, et al., 2021. Blind light field image quality assessment by analyzing angular-spatial characteristics. *Dig Signal Process*, 117:103138. https://doi.org/10.1016/j.dsp.2021.103138

Fang YM, Wei KK, Hou JH, et al., 2018. Light filed image quality assessment by local and global features of epipolar plane image. Proc IEEE 4th Int Conf on Multimedia Big Data, p.1-6. https://doi.org/10.1109/BigMM.2018.8499086

Fiss J, Curless B, Szeliski R, 2014. Refocusing plenoptic images using depth-adaptive splatting. Proc IEEE Int Conf on Computational Photography, p.1-9. https://doi.org/10.1109/ICCPHOT.2014.6831809

Gao Q, Han L, Shen J, et al., 2017. Focused-region segmentation for light field images based on PCNN. Proc Int Smart Cities Conf, p.1-6. https://doi.org/10.1109/ISC2.2017.8090851

Georgiev TG, Lumsdaine A, 2010. Focused plenoptic camera and rendering. *J Electron Imag*, 19(2):021106. https://doi.org/10.1117/1.3442712

Gershun A, 1939. The light field. *J Math Phys*, 18(1-4):51-151. https://doi.org/10.1002/sapm193918151

Ghasemi A, Vetterli M, 2014. Detecting planar surface using a light-field camera with application to distinguishing real scenes from printed photos. Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, p.4588-4592. https://doi.org/10.1109/ICASSP.2014.6854471

Gryaditskaya Y, Masia B, Didyk P, et al., 2016. Gloss editing in light fields. Proc Conf on Vision, Modeling and Visualization, p.127-135. https://doi.org/10.5555/3056901.3056923

Guo BC, Wen JT, Han YX, 2020. Deep material recognition in light-fields via disentanglement of spatial and angular information. Proc 16th European Conf on Computer Vision, p.664-679. https://doi.org/10.1007/978-3-030-58586-0_39

Guo XQ, Lin HT, Yu Z, et al., 2015. Barcode imaging using a light field camera. Proc European Conf on Computer Vision, p.519-532. https://doi.org/10.1007/978-3-319-16181-5_40

Hog M, Sabater N, Guillemot C, 2016. Light field segmentation using a ray-based graph structure. Proc 14th European Conf on Computer Vision, p.35-50. https://doi.org/10.1007/978-3-319-46478-7_3

Hsieh PY, Chou PY, Lin HA, et al., 2018. Long working range light field microscope with fast scanning multifocal liquid crystal microlens array. *Opt Expr*, 26(8):10981-10996. https://doi.org/10.1364/oe.26.010981

Huang ZJ, Yu M, Xu HY, et al., 2018. New quality assessment method for dense light fields. Proc SPIE 10817, Optoelectronic Imaging and Multimedia Technology V, Article 1081717. https://doi.org/10.1117/12.2502277

Jia C, Shi F, Zhao YF, et al., 2018. Identification of pedestrians from confused planar objects using light field imaging. *IEEE Access*, 6:39375-39384. https://doi.org/10.1109/ACCESS.2018.2855723

Johannsen O, Sulc A, Goldluecke B, 2015. Variational separation of light field layers. Proc 20th Int Symp on Vision, Modeling, and Visualization, p.135-142. https://doi.org/10.2312/vmv.20151268

Kalantari NK, Wang TC, Ramamoorthi R, 2016. Learning-based view synthesis for light field cameras. *ACM Trans Graph*, 35(6):193. https://doi.org/10.1145/2980179.2980251

Kim C, Zimmer H, Pritch Y, et al., 2013. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans Graph*, 32(4):73. https://doi.org/10.1145/2461912.2461926

Lee JY, Park RH, 2017. Separation of foreground and background from light field using gradient information. *Appl Opt*, 56(4):1069-1078. https://doi.org/10.1364/AO.56.001069

Levoy M, Hanrahan P, 1996. Light field rendering. Proc 23rd Annual Conf on Computer Graphics and Interactive Techniques, p.31-42. https://doi.org/10.1145/237170.237199

Levoy M, Ng R, Adams A, et al., 2006. Light field microscopy. Proc ACM SIGGRAPH, p.924-934. https://doi.org/10.1145/1179352.1141976

Li NY, Ye JW, Ji Y, et al., 2014. Saliency detection on light field. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.2806-2813. https://doi.org/10.1109/CVPR.2014.359

Li NY, Sun BL, Yu JY, 2015. A weighted sparse coding framework for saliency detection. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.5216-5223. https://doi.org/10.1109/CVPR.2015.7299158

Li ZQ, Xu ZX, Ramamoorthi R, et al., 2017. Robust energy minimization for BRDF-invariant shape from light fields. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.578-586. https://doi.org/10.1109/CVPR.2017.69

Liang CK, Lin TH, Wong BY, et al., 2008. Programmable aperture photography: multiplexed light field acquisition. *ACM Trans Graph*, 27(3):1-10. https://doi.org/10.1145/1360612.1360654

Lippmann G, 1908. Epreuves reversibles, photographies integrales. *Comput R Acad Sci*, 444:446-451.

Lumsdaine A, Georgiev T, 2009. The focused plenoptic camera. Proc IEEE Int Conf on Computational Photography, p.1-8. https://doi.org/10.1109/ICCPHOT.2009.5559008

Lv XQ, Wang X, Wang Q, et al., 2021. 4D light field segmentation from light field super-pixel hypergraph representation. *IEEE Trans Vis Comput Graph*, 27(9):3597-3610. https://doi.org/10.1109/TVCG.2020.2982158

Marquez M, Rueda-Chacon H, Arguello H, 2020. Compressive spectral light field image reconstruction via online tensor representation. *IEEE Trans Image Process*, 29:3558-3568. https://doi.org/10.1109/TIP.2019.2963376

Mehajabin N, Pourazad M, Nasiopoulos P, 2020. SSIM assisted pseudo-sequence-based prediction structure for light field video compression. Proc IEEE Int Conf on Consumer Electronics, p.1-2. https://doi.org/10.1109/ICCE46568.2020.9042968

Meng CL, An P, Huang XP, et al., 2019. Objective quality assessment for light field based on refocus characteristic. Proc 10th Int Conf on Image and Graphics, p.193-204. https://doi.org/10.1007/978-3-030-34113-8_17

Mihara H, Funatomi T, Tanaka K, et al., 2016. 4D light field segmentation with spatial and angular consistencies. Proc IEEE Int Conf on Computational Photography, p.1-8. https://doi.org/10.1109/ICCPHOT.2016.7492872

Murgia F, Giusto D, Perra C, et al., 2015. 3D reconstruction from plenoptic image. Proc 23rd Telecommunications Forum Telfor, p.448-451. https://doi.org/10.1109/TELFOR.2015.7377504

Ng R, Levoy M, Brédif M, et al., 2005. Light field photography with a hand-held plenoptic camera. Stanford Tech Report CTSR 2005-02.

Nian ZC, Jung C, 2019. CNN-based multi-focus image fusion with light field data. Proc IEEE Int Conf on Image Processing, p.1044-1048. https://doi.org/10.1109/ICIP.2019.8803065

Paudyal P, Olsson R, Sjöström M, et al., 2016. SMART: a light field image quality dataset. Proc 7th Int Conf on Multimedia Systems, Article 49. https://doi.org/10.1145/2910017.2910623

Paudyal P, Battisti F, Sjöström M, et al., 2017. Towards the perceptual quality evaluation of compressed light field images. *IEEE Trans Broadcast*, 63(3):507-522. https://doi.org/10.1109/TBC.2017.2704430

Paudyal P, Battisti F, Carli M, 2019. Reduced reference quality assessment of light field images. *IEEE Trans Broadcast*, 65(1):152-165. https://doi.org/10.1109/TBC.2019.2892092

Piao YR, Li X, Zhang M, et al., 2019a. Saliency detection via depth-induced cellular automata on light field. *IEEE Trans Image Process*, 29:1879-1889. https://doi.org/10.1109/TIP.2019.2942434

Piao YR, Rong ZK, Zhang M, et al., 2019b. Deep light-field-driven saliency detection from a single view. Proc 28th Int Joint Conf on Artificial Intelligence, p.904-911. https://doi.org/10.24963/ijcai.2019/127

Piao YR, Jiang YY, Zhang M, et al., 2021. PANet: patch-aware network for light field salient object detection. *IEEE Trans Cybern*, early access. https://doi.org/10.1109/TCYB.2021.3095512

Raghavendra R, Busch C, 2014. Presentation attack detection on visible spectrum iris recognition by exploring inherent

characteristics of light field camera. Proc IEEE Int Joint Conf on Biometrics, p.1-8. https://doi.org/10.1109/BTAS.2014.6996226

Raghavendra R, Raja KB, Yang B, et al., 2013a. Combining iris and periocular recognition using light field camera. Proc 2nd IAPR Asian Conf on Pattern Recognition, p.155-159. https://doi.org/10.1109/ACPR.2013.22

Raghavendra R, Raja KB, Yang B, et al., 2013b. A novel image fusion scheme for robust multiple face recognition with light-field camera. Proc 16th Int Conf on Information Fusion, p.722-729.

Raghavendra R, Raja KB, Busch C, 2016. Exploring the usefulness of light field cameras for biometrics: an empirical study on face and iris recognition. *IEEE Trans Inform Forens Secur*, 11(5):922-936. https://doi.org/10.1109/TIFS.2015.2512559

Rerabek M, Ebrahimi T, 2016. New light field image dataset. Proc 8th Int Conf on Quality of Multimedia Experience.

Sabater N, Boisson G, Vandame B, et al., 2017. Dataset and pipeline for multi-view light-field video. Proc IEEE Conf on Computer Vision and Pattern Recognition Workshops, p.1743-1753. https://doi.org/10.1109/CVPRW.2017.221

Sepas-Moghaddam A, Pereira F, Correia PL, 2018. Light field-based face presentation attack detection: reviewing, benchmarking and one step further. *IEEE Trans Inform Forens Secur*, 13(7):1696-1709. https://doi.org/10.1109/TIFS.2018.2799427

Shan L, An P, Meng CL, et al., 2019. A no-reference image quality assessment metric by multiple characteristics of light field images. *IEEE Access*, 7:127217-127229. https://doi.org/10.1109/ACCESS.2019.2940093

Sheng H, Deng SY, Zhang S, et al., 2016. Segmentation of light field image with the structure tensor. Proc IEEE Int Conf on Image Processing, p.1469-1473. https://doi.org/10.1109/ICIP.2016.7532602

Shi LK, Zhao SY, Zhou W, et al., 2018. Perceptual evaluation of light field image. Proc 25th IEEE Int Conf on Image Processing, p.41-45. https://doi.org/10.1109/ICIP.2018.8451077

Shi LK, Zhao SY, Chen ZB, 2019. Belif: blind quality evaluator of light field image with tensor structure variation index. Proc IEEE Int Conf on Image Processing, p.3781-3785. https://doi.org/10.1109/ICIP.2019.8803559

Shi LK, Zhou W, Chen ZB, et al., 2020. No-reference light field image quality assessment based on spatial-angular measurement. *IEEE Trans Circ Syst Video Technol*, 30(11): 4114-4128. https://doi.org/10.1109/TCSVT.2019.2955011

Smith BM, Zhang L, Jin HL, et al., 2009. Light field video stabilization. Proc IEEE 12th Int Conf on Computer Vision, p.341-348. https://doi.org/10.1109/ICCV.2009.5459270

Sulc A, Alperovich A, Marniok N, et al., 2016. Reflection separation in light fields based on sparse coding and specular flow. Proc Conf on Vision, Modeling and Visualization, p.137-144. https://doi.org/10.5555/3056901.3056924

Sun J, Hossain M, Xu CL, et al., 2017. A novel calibration method of focused light field camera for 3-D reconstruction of flame temperature. *Opt Commun*, 390:7-15. https://doi.org/10.1016/j.optcom.2016.12.056

Tambe S, Veeraraghavan A, Agrawal A, 2013. Towards motion aware light field video for dynamic scenes. Proc IEEE Int Conf on Computer Vision, p.1009-1016. https://doi.org/10.1109/ICCV.2013.129

Tao MW, Srinivasan PP, Malik J, et al., 2015a. Depth from shading, defocus, and correspondence using light-field angular coherence. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1940-1948. https://doi.org/10.1109/CVPR.2015.7298804

Tao MW, Su JC, Wang TC, et al., 2015b. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE Trans Patt Anal Mach Intell*, 38(6):1155-1169. https://doi.org/10.1109/TPAMI.2015.2477811

Tian Y, Zeng HQ, Hou JH, et al., 2021. A light field image quality assessment model based on symmetry and depth features. *IEEE Trans Circ Syst Video Technol*, 31(5):2046-2050. https://doi.org/10.1109/TCSVT.2020.2971256

Vizcaíno JP, Saltarin F, Belyaev Y, et al., 2021. Learning to reconstruct confocal microscopy stacks from single light field images. *IEEE Trans Comput Imag*, 7:775-788. https://doi.org/10.1109/TCI.2021.3097611

Wang AZ, Wang MH, Li XY, et al., 2017. A two-stage Bayesian integration framework for salient object detection on light field. *Neur Process Lett*, 46(3):1083-1094. https://doi.org/10.1007/s11063-017-9610-x

Wang HQ, Xu CX, Wang XZ, et al., 2016. Light field imaging based accurate image specular highlight removal. *PLoS ONE*, 11(6):e0156173. https://doi.org/10.1371/journal.pone.0156173

Wang TC, Efros AA, Ramamoorthi R, 2015. Occlusion-aware depth estimation using light-field cameras. Proc IEEE Int Conf on Computer Vision, p.3487-3495. https://doi.org/10.1109/ICCV.2015.398

Wang TC, Zhu JY, Hiroaki E, et al., 2016a. A 4D light-field dataset and CNN architectures for material recognition. Proc 14th European Conf on Computer Vision, p.121-138. https://doi.org/10.1007/978-3-319-46487-9_8

Wang TC, Chandraker M, Efros AA, et al., 2016b. SVBRDF-invariant shape and reflectance estimation from light-field cameras. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.5451-5459. https://doi.org/10.1109/CVPR.2016.588

Wang TC, Zhu JY, Kalantari NK, et al., 2017. Light field video capture using a learning-based hybrid imaging system. *ACM Trans Graph*, 36(4):133. https://doi.org/10.1145/3072959.3073614

Wang TT, Piao YR, Li XC, et al., 2019. Deep learning for light field saliency detection. Proc IEEE/CVF Int Conf on Computer Vision, p.8837-8847. https://doi.org/10.1109/ICCV.2019.00893

Wang YQ, Yang JG, Xiao C, et al., 2018. An efficient method for the fusion of light field refocused images. Proc SPIE 9th Int Conf on Graphic and Image Processing, Article 1061536. https://doi.org/10.1117/12.2302687

Wanner S, Meister S, Goldluecke B, 2013a. Datasets and benchmarks for densely sampled 4D light fields. Proc 18th Int Workshop on Vision, Modeling, and Visualization,

p.225-226. https://doi.org/10.2312/PE.VMV.VMV13.225-226
Wanner S, Straehle C, Goldluecke B, 2013b. Globally consistent multi-label assignment on the ray space of 4D light fields. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1011-1018. https://doi.org/10.1109/CVPR.2013.135
Wilburn B, Smulski M, Lee HHK, et al., 2002. Light field video camera. Proc SPIE 6474, Media Processors, p.29-36. https://doi.org/10.1117/12.451074
Wilburn B, Joshi N, Vaish V, et al., 2005. High performance imaging using large camera arrays. *ACM Trans Graph*, 24(3):765-776. https://doi.org/10.1145/1073204.1073259
Wu GC, Masia B, Jarabo A, et al., 2017. Light field image processing: an overview. *IEEE J Sel Top Signal Process*, 11(7): 926-954. https://doi.org/10.1109/JSTSP.2017.2747126
Xu YC, Nagahara H, Shimada A, et al., 2015. TransCut: transparent object segmentation from a light-field image. Proc IEEE Int Conf on Computer Vision, p.3442-3450. https://doi.org/10.1109/ICCV.2015.393
Xu YC, Nagahara H, Shimada A, et al., 2019. TransCut2: transparent object segmentation from a light-field image. *IEEE Trans Comput Imag*, 5(3):465-477. https://doi.org/10.1109/TCI.2019.2893820
Yang JC, 2000. A Light Field Camera for Image Based Rendering. MS Thesis, Massachusetts Institute of Technology, Cambridge, USA.
Yücer K, Sorkine-Hornung A, Wang O, et al., 2016. Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction. *ACM Trans Graph*, 35(3):22. https://doi.org/10.1145/2876504
Zhang C, Chen T, 2004. A self-reconfigurable camera array. Proc ACM SIGGRAPH Sketches, p.151. https://doi.org/10.1145/1186223.1186412
Zhang C, Hou GQ, Sun ZA, et al., 2013. Light field photography for iris image acquisition. Proc 8th Chinese Conf on Biometric Recognition, p.345-352. https://doi.org/10.1007/978-3-319-02961-0_43
Zhang J, Wang M, Gao J, et al., 2015. Saliency detection with a deeper investigation of light field. Proc 24th Int Joint Conf on Artificial Intelligence, p.2212-2218.
Zhang J, Wang M, Lin L, et al., 2017. Saliency detection on light field: a multi-cue approach. *ACM Trans Multim Comput Commun Appl*, 13(3):32. https://doi.org/10.1145/3107956
Zhang J, Liu YM, Zhang SP, et al., 2020. Light field saliency detection with deep convolutional networks. *IEEE Trans Image Process*, 29:4421-4434. https://doi.org/10.1109/TIP.2020.2970529
Zhang M, Geng Z, Pei RJ, et al., 2017. Three-dimensional light field microscope based on a lenslet array. *Opt Commun*, 403:133-142. https://doi.org/10.1016/j.optcom.2017.07.026
Zhang M, Li JJ, Wei J, et al., 2019. Memory-oriented decoder for light field salient object detection. Proc Advances in Neural Information Processing Systems 32, p.2898-2909.
Zhang XD, Wang Y, Zhang J, et al., 2015. Light field saliency vs. 2D saliency: a comparative study. *Neurocomputing*, 166:389-396. https://doi.org/10.1016/j.neucom.2015.03.042
Zhou MY, Ding YQ, Ji Y et al., 2020. Shape and reflectance reconstruction using concentric multi-spectral light field. *IEEE Trans Patt Anal Mach Intell*, 42(7):1594-1605. https://doi.org/10.1109/TPAMI.2020.2986764
Zhou W, Shi LK, Chen ZB, et al., 2020. Tensor oriented no-reference light field image quality assessment. *IEEE Trans Image Process*, 29:4070-4084. https://doi.org/10.1109/TIP.2020.2969777
Zhu H, Zhang Q, Wang Q, 2017. 4D light field superpixel and segmentation. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.6709-6717. https://doi.org/10.1109/CVPR.2017.710

List of supplementary materials

1 Light field imaging processing
2 Tasks and applications
Table S1 The peak signal-to-noise ratio comparision on three LF datasets
Table S2 The structural similarity comparision on three LF datasets
Table S3 MSE performance evaluation on the synthetic images
Table S4 BadPix performance evaluation on the synthetic images
Table S5 Overview of publicly avalable LF-based face artefact datasets
Fig. S1 The framework of depth estimation based on LF information
Fig. S2 Face recognition framework based on subaperture image analysis
Fig. S3 The face recognition framework based on different depth images