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:*

# Multi-aperture optical imaging systems and their mathematical light field acquisition models*

Qiming QI†, Ruigang FU, Zhengzheng SHAO, Ping WANG, Hongqi FAN†‡

*National Key Laboratory of Science and Technology on ATR, College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China*

†E-mail: qiqiming19@163.com; fanhongqi@nudt.edu.cn

Received Jan. 31, 2021; Revision accepted May 18, 2021; Crosschecked Mar. 24, 2022

**Abstract:** Inspired by the compound eyes of insects, many multi-aperture optical imaging systems have been proposed to improve the imaging quality, e.g., to yield a high-resolution image or an image with a large field-of-view. Previous research has reviewed existing multi-aperture optical imaging systems, but few papers emphasize the light field acquisition model which is essential to bridge the gap between configuration design and application. In this paper, we review typical multi-aperture optical imaging systems (i.e., artificial compound eye, light field camera, and camera array), and then summarize general mathematical light field acquisition models for different configurations. These mathematical models provide methods for calculating the key indexes of a specific multi-aperture optical imaging system, such as the field-of-view and sub-image overlap ratio. The mathematical tools simplify the quantitative design and evaluation of imaging systems for researchers.

**Key words:** Multi-aperture optical imaging system; Artificial compound eye; Light field camera; Camera array; Light field acquisition model

https://doi.org/10.1631/FITEE.2100058 **CLC number:** O438

## 1 Introduction

Over the past few decades, the imaging performance of single-aperture optical imaging devices, such as the single-lens reflex (SLR) camera, has been significantly improved. However, because a space point can be recorded by only one pixel in an imaging sensor, the ability to obtain more information about the scene (e.g., the depth of imaged objects) is limited. Generally, digital imaging sensors, e.g., complementary metal oxide semiconductor (CMOS) and charge coupled device (CCD), are flat, which accounts for the difficulty of adapting the optical device with wide field-of-view (FOV) imaging (i.e., obvious chromatic aberration and distortion). In addition, one imaging sensor cannot meet both wide FOV and high-resolution requirements.

To reach some goals such as high resolution, wide FOV, depth information acquisition, and multi-target detection, many multi-aperture optical imaging systems have been proposed since the begining of the 20th century. Multi-aperture optical imaging systems are based on the compound eye of insects (Land, 1989; Wen et al., 2019), which has the advantages of small volume, wide FOV, and high sensitivity to moving targets. Researchers refer to multi-aperture optical imaging systems as the artificial compound eye (ACE) (Gong et al., 2013; Hao and Li, 2015; Wu SD et al., 2017; Cheng et al., 2019) and the light field camera and camera array (Wu G et al., 2017; Zhu H et al., 2017).

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 62001482) and the Hunan Provincial Natural Science Foundation of China (No. 2021JJ40676)
ORCID: Qiming QI, https://orcid.org/0000-0001-9141-4767; Hongqi FAN, https://orcid.org/0000-0002-9990-9163
© Zhejiang University Press 2022

In ACE research (Gong et al., 2013; Hao and Li, 2015; Wu SD et al., 2017; Cheng et al., 2019), existing ACEs were classified and the preparation and application prospects of ACE were introduced, but the light field acquisition model was not included. Also, the use of the light field camera and planar camera array (Wu G et al., 2017; Zhu H et al., 2017) was introduced to obtain discrete light field information. Then, concerning the light field technique (Levoy, 2006), which is a key area of computational photography (Suo et al., 2012), researchers have analyzed the light field representation and calculation method.

An ACE consists of multiple imaging apertures, which are called sub-eyes. These sub-eyes can be an individual lens module or a combination of a microlens array and common imaging sensors. Because a single camera has all the elements of a lens module, a camera array is actually an ACE with a planar structure. Different from ACEs, a light field camera inserts a microlens array behind the main lens of an ordinary camera. However, after a decoding operation, as in planar ACEs, sub-images from different viewpoints can be obtained. Thus, a light field camera is a special planar ACE. Summarizing current research, the ACE, light field camera, and camera array have common characteristics: multiple imaging apertures are integrated, and the relative position of each imaging aperture conforms to the symmetrical arrangement rule. Therefore, one can refer to these systems collectively as multi-aperture optical imaging systems.

Multi-aperture optical imaging systems integrate preparation technology, optical design, and machine vision algorithms, and they have great value in applications such as reconnaissance, image navigation, computational photography, and medical endoscopy. The mathematical light field acquisition models play an essential role in closing the gap between configuration design and application during research. However, there is little research that summarizes light field acquisition for multi-aperture optical imaging systems.

In this paper, some typical multi-aperture optical imaging systems are enumerated and categorized. In contrast to other research, general mathematical light field acquisition models are summarized for different kinds of multi-aperture optical imaging systems. Based on the models, it will be easier to quantitatively analyze the configuration design and carry out information processing research. In addition, the basic applications of multi-aperture optical imaging systems are analyzed referring to the light field acquisition models.

The rest of this paper focuses on three aspects: (1) typical multi-aperture optical imaging systems, (2) general light field acquisition models, and (3) the application analysis of multi-aperture optical imaging systems.

## 2 Typical multi-aperture optical imaging systems

In previous research (Gong et al., 2013; Hao and Li, 2015; Wu G et al., 2017; Wu SD et al., 2017; Zhu H et al., 2017; Cheng et al., 2019), some typical multi-aperture optical imaging systems have been enumerated and the performance has been compared from different perspectives. The development trend of multi-aperture optical imaging systems is shown in Fig. 1.

The aim of multi-aperture optical imaging systems is to obtain sub-images from different viewpoints. Each sub-image corresponds to an imaging aperture. For vision application, multi-aperture optical imaging systems can be divided into two main structures based on the position of each sub-image's optical center, i.e., planar structure (Tanida et al., 2000; Yang JC et al., 2002; Duparré et al., 2005; Ng and Hanrahan, 2005; Wilburn et al., 2005; Levoy et al., 2006; Lumsdaine and Georgiev, 2009; Li and Yi, 2012; Venkataraman et al., 2013; Cao et al., 2018) and convex structure (Duparré et al., 2007; Zhang YK et al., 2010; Brady et al., 2012; Guo et al., 2012; Afshari et al., 2013; Song et al., 2013; Leitel et al., 2014; Cao et al., 2015; Luo et al., 2015; Deng et al., 2016; Pang et al., 2017; Shi et al., 2017; Yu et al., 2019; Zhang JM et al., 2020; Zhou et al., 2020). In addition, three possible arrangements exist in the convex structure: spherical multi-loop (Brady et al., 2012; Guo et al., 2012; Afshari et al., 2013; Cao et al., 2015; Luo et al., 2015; Pang et al., 2017; Shi et al., 2017; Yu et al., 2019; Zhang JM et al., 2020; Zhou et al., 2020), spherical multi-row (Zhang YK et al., 2010; Song et al., 2013; Deng et al., 2016), and cylinder (Leitel et al., 2014).

There are basically two types of ACEs in terms of size, the microlens array (Tanida et al., 2000;

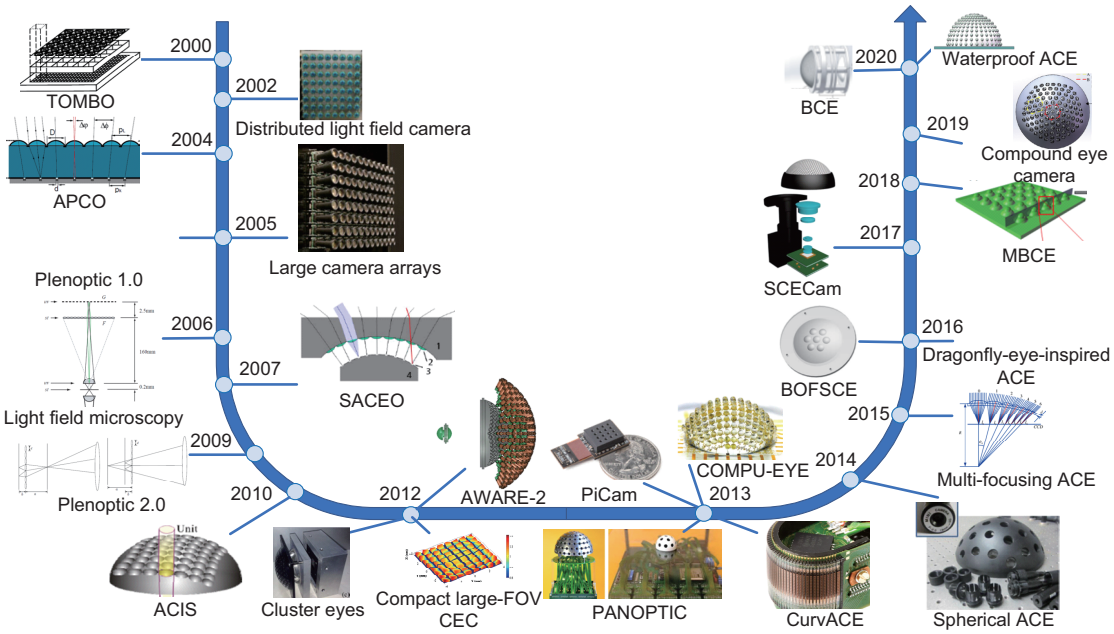


**Fig. 1 The development trend of multi-aperture optical imaging systems**

Duparré et al., 2005; Duparré et al., 2007; Zhang YK et al., 2010; Li and Yi, 2012; Song et al., 2013; Leitel et al., 2014; Luo et al., 2015; Deng et al., 2016; Pang et al., 2017; Cao et al., 2018; Zhang JM et al., 2020; Zhou et al., 2020) and the lens module array (Brady et al., 2012; Guo et al., 2012; Afshari et al., 2013; Cao et al., 2015; Shi et al., 2017; Yu et al., 2019). As for light field acquisition devices, when a microlens array is placed behind the main lens of an ordinary camera, it becomes a light field camera (Ng and Hanrahan, 2005; Levoy et al., 2006; Lumsdaine and Georgiev, 2009). Except for non-traditional light field acquisition approaches (e.g., time-sequential capture, conventional camera with a coded aperture), both the camera array (Yang JC et al., 2002; Wilburn et al., 2005) and the lens module array (Venkataraman et al., 2013) can be called planar ACEs.

By setting the layout of optical lenses and imaging sensors, the optical center and optical axis of each sub-image can be determined. In general, the distribution of optical lens often determines the distribution of the sub-images' optical centers. The best result is that each lens has an imaging sensor and the imaging sensors are distributed in an array according to the positions of the optical lenses. However, because they are limited by size, some multi-aperture optical imaging systems (Tanida et al., 2000; Ng and Hanrahan, 2005; Levoy et al., 2006; Duparré et al., 2007; Lumsdaine and Georgiev, 2009; Guo et al., 2012; Luo et al., 2015; Pang et al., 2017; Shi et al., 2017; Yu et al., 2019; Zhang JM et al., 2020) use only one imaging sensor to obtain all sub-images. To improve the utilization of the imaging sensor, sub-images are focused on a common imaging sensor by an optical transferring system in some convex microlens arrays (Duparré et al., 2007; Zhang JM et al., 2020) and convex lens module arrays (Guo et al., 2012; Shi et al., 2017; Yu et al., 2019).

For optical lenses, there are some special cases, such as free-form surface optics (Li and Yi, 2012; Pang et al., 2017) and multi-focal optics (Cao et al., 2018). Li and Yi (2012) used free-form surface design to make the optical axis of each sub-image point differently. Pang et al. (2017) reduced aberrations using free-form surface design. Cao et al. (2018) designed a microlens array with two focal lengths to demonstrate excellent two-order focusing abilities.

The classification is shown in Fig. 2. Table 1 shows the features of the multi-aperture optical imaging systems above. Briefly, multi-aperture optical imaging systems can be classified in three ways. First, for vision applications, there are planar structures and convex structures, which are based on the position of each sub-image's optical center. Second, there are two styles in terms of size: microlens array

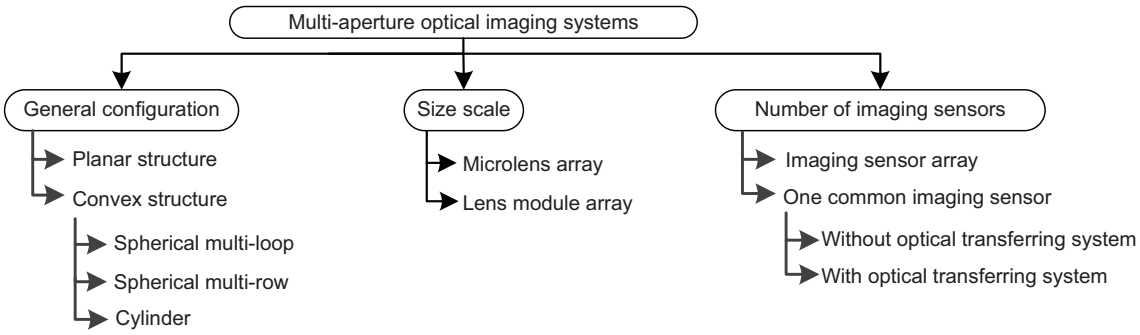


**Fig. 2 The classification of multi-aperture optical imaging systems**

**Table 1 The features of typical multi-aperture optical imaging systems**

| Name | General configuration | Size scale | Number of imaging sensors |
|---|---|---|---|
| TOMBO (Tanida et al., 2000) | 1 | 1 | 2(1) |
| Distributed light field camera (Yang JC et al., 2002) | 1 | 2 | 1 |
| APCO (Duparré et al., 2005) | 1 | 1 | 1 |
| Plenoptic 1.0 (Ng and Hanrahan, 2005) | 1 | 1 | 2(1) |
| Large camera arrays (Wilburn et al., 2005) | 1 | 2 | 1 |
| Light field microscopy (Levoy et al., 2006) | 1 | 1 | 2(1) |
| SACEO (Duparré et al., 2007) | 2(–) | 1 | 2(2) |
| Plenoptic 2.0 (Lumsdaine and Georgiev, 2009) | 1 | 1 | 2(1) |
| ACIS (Zhang YK et al., 2010) | 2(2) | 1 | 1 |
| Cluster eyes (Guo et al., 2012) | 2(1) | 2 | 2(2) |
| AWARE-2 (Brady et al., 2012) | 2(1) | 2 | 1 |
| PANOPTIC (Afshari et al., 2013) | 2(1) | 2 | 1 |
| Compact large-FOV CEC (Li and Yi, 2012) | 1 | 1 | 2(1) |
| PiCam (Venkataraman et al., 2013) | 1 | 2 | 1 |
| COMPU-EYE (Song et al., 2013) | 2(2) | 1 | 1 |
| CurvACE (Leitel et al., 2014) | 2(3) | 1 | 1 |
| Spherical ACE (Cao et al., 2015) | 2(1) | 2 | 1 |
| Multi-focusing ACE (Luo et al., 2015) | 2(1) | 1 | 2(1) |
| Dragonfly-eye-inspired ACE (Deng et al., 2016) | 2(2) | 1 | – |
| BOFSCE (Pang et al., 2017) | 2(1) | 1 | 2(1) |
| SCECam (Shi et al., 2017) | 2(1) | 1 | 2(2) |
| MBCE (Cao et al., 2018) | 1 | 1 | – |
| Compound eye camera (Yu et al., 2019) | 2(1) | 2 | 2(2) |
| BCE (Zhang JM et al., 2020) | 2(1) | 1 | 2(2) |
| Waterproof ACE (Zhou et al., 2020) | 2(1) | 1 | – |

In the 2nd column: 1, planar structure; 2(1/2/3), convex structure (sphere multi-loop/sphere multi-row/cylinder). In the 3rd column: 1, microlens array; 2, lens module array. In the 4th column: 1, imaging sensor array; 2(1/2), one common imaging sensor (without/with optical transferring system). –: not clear

and lens module array. Third, some multi-aperture optical imaging systems use an imaging sensor array to obtain sub-images, while others use only one common imaging sensor.

Multi-aperture optical imaging systems face mainly two problems: one is the preparation and optimization of optical components, and the other is the combination of optical components and an imaging sensor. There are two considerations in the preparation and optimization of optical components: the microlens array manufacturing process (Yuan et al., 2018; Zhu L et al., 2019) and image quality improvement. To improve the image quality of a traditional lens, the trend is to apply free-form surface optics in multi-aperture optical imaging systems. Concerning the combination of optical components and an imaging sensor, the first problem is how to design a curved imaging sensor to fit a curved microlens array, and the second is how to optimize the optical transferring system for a lens module array using a common imaging sensor to achieve both high pixel efficiency and high-quality imaging.

## 3 Basic theory: single-aperture optical imaging model

To analyze the light field acquisition of multi-aperture optical imaging systems, the first step is to use the general single-aperture optical imaging model. In this section, the optical path model of single-aperture imaging, which is on the physical level, is first briefly introduced. Then the equivalent mathematical model is introduced as the basis of this paper.

### 3.1 Physical level: optical path model

A multi-aperture optical imaging system consists of multiple imaging apertures. The design of a single imaging aperture usually follows geometric optics principles (O'Shea and Zajac, 1986; Born et al., 1999; Lindlein and Leuchs, 2012). As an imaging optical system, the basic principle of single aperture is that a lens forms a real inverted picture of a scene on the surface of an imaging sensor. Also, each imaging aperture has a diaphragm near the lens (Lindlein and Leuchs, 2012). When the object depth is much greater than the focal length of the lenses, single aperture can be simplified as a thin lens model (Barsky et al., 2003; Liang et al., 2011), as shown in Fig. 3.

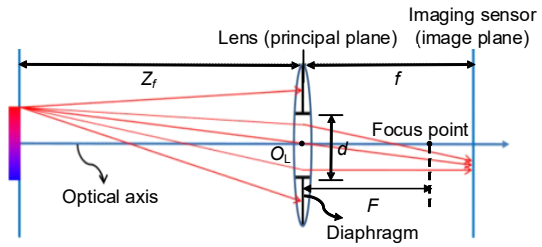


**Fig. 3 Single-aperture optical path model. An imaging aperture can be simplified as a thin lens model: a thin lens, a diaphragm, and an imaging sensor**

The center of lens $O_L$ is called the optical center, and the axis that passes through $O_L$ is the optical axis. The plane that is normal to the optical axis at $O_L$ is called the principal plane, and the image plane is located at the surface of the imaging sensor. The focus point on the optical axis has the property that any ray emanating from it or proceeding toward it travels parallel to the optical axis after refraction by the lens. Focal length $F$ is the distance between the principal plane and the focus point. The distance between the principal plane and the image plane is the principal distance $f$. The relationship among $f$, $F$, and the focused object depth $Z_f$ follows $1/Z_f + 1/f = 1/F$. In practice, the principal distance $f$ is approximately equal to the focal length $F$. The size, $d$, of the diaphragm, controls the light flux of the lens.

Because optical design cannot be perfect, any single imaging aperture has depth of field (DOF), chromatic aberration, distortion, and other errors. Although these errors may be needed in photography, for vision application, it is necessary to optimize the optical design to eliminate the errors due to the requirement for high measurement accuracy. For multi-aperture optical imaging systems, attention should be paid to the main imaging indexes, such as the principal distance $f$ or focal length $F$ and the FOV.

### 3.2 Mathematical level: pinhole model

The single-aperture mathematical model plays a key role in extracting information from the obtained sub-images. In practice, two models are used to link configuration design and application. The first is the pinhole model (Hartley and Zisserman, 2004), and the other is the unified omnidirectional camera model (Geyer and Daniilidis, 2000). The former is a linear image projection that limits the FOV below 180°. The latter is suitable for a wide FOV, but the image projection is nonlinear. With the benefit of enough imaging apertures, the FOV of each aperture is generally small. Linear image projection is essential for multi-aperture optical imaging systems, and to this end, the pinhole model is common as the general mathematical model.

According to projective geometry (Hartley and Zisserman, 2004), after calibration, the position of a pixel in a sub-image can represent the direction of a ray passing through the optical center. As shown in Fig. 4, for each aperture, the sub-image is composed of $N_u \times N_v$ pixels and associated with two coordinate systems: the pixel coordinate system $O_p - uv$ and the local coordinate system $O_L - X_L Y_L Z_L$. For the multi-aperture optical imaging system, global coordinate system $O_G - X_G Y_G Z_G$ is established with its geometric center.

The origin of local coordinate system $O_L$ is located at the optical center, and the $Z_L$ axis follows the direction of the optical axis. The optical axis intersects the image plane at $(c_u, c_v)$. The distance

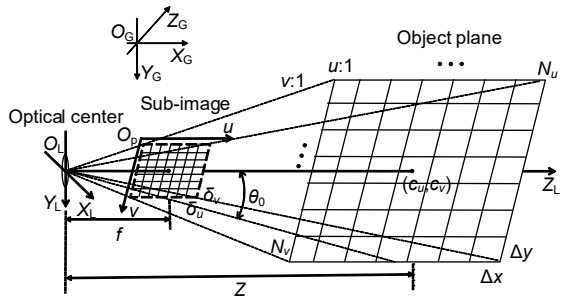


**Fig. 4 Pinhole camera model. There are two coordinate systems associated with a single aperture: pixel coordinate system $\boldsymbol{O_P - uv}$ and local coordinate system $\boldsymbol{O_L - X_L Y_L Z_L}$. The global coordinate system $\boldsymbol{O_G - X_G Y_G Z_G}$ is established with the geometric center of the multi-aperture optical imaging system**

between $O_\mathrm{L}$ and the image plane is the principal distance $f$. According to Eq. (1), if the global coordinates of an object $\boldsymbol{P}_\mathrm{G} = \left(X_\mathrm{G}, Y_\mathrm{G}, Z_\mathrm{G}\right)^\mathrm{T}$ are offered, the local coordinates $\boldsymbol{P}_\mathrm{L} = \left(X_\mathrm{L}, Y_\mathrm{L}, Z_\mathrm{L}\right)^\mathrm{T}$ are obtained. In this paper, $\hat{\cdot}$ is used to denote the homogeneous vector by adding 1 as the last element: $\hat{\boldsymbol{P}} = \left(X, Y, Z, 1\right)^\mathrm{T}$. The rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and translation vector $\boldsymbol{t} \in \mathbb{R}^3$ are called extrinsic parameters. We have

$$\boldsymbol{P}_\mathrm{L} = \left(\boldsymbol{R} \quad \boldsymbol{t}\right) \hat{\boldsymbol{P}}_\mathrm{G}. \tag{1}$$

In this paper, every sub-image is assumed to have no distortion (Weng et al., 1992; Heikkila and Silven, 1997). $\hat{\boldsymbol{p}} = \left(u, v, 1\right)^\mathrm{T}$, which is the position of a pixel in the image plane, represents the homogeneous pixel coordinates. $\delta_u$ and $\delta_v$, measured in microns, are the width and height of a pixel on the imaging sensor, respectively. The pixel focal lengths $f_u$ and $f_v$ are obtained using $f_u = f/\delta_u$ and $f_v = f/\delta_v$. Using Eq. (2), the pixel coordinates are obtained from the corresponding local coordinates. The matrix $\boldsymbol{K} \in \mathbb{R}^{3\times3}$, called the calibration matrix, consists of the intrinsic parameter $\{f_u, f_v, c_u, c_v\}$.

$$\begin{aligned}\hat{\boldsymbol{p}} &= \frac{1}{Z_\mathrm{L}} \begin{pmatrix} \frac{1}{\delta_u} & 0 & c_u \\ 0 & \frac{1}{\delta_v} & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{P}_\mathrm{L} \\ &= \frac{1}{Z_\mathrm{L}} \begin{pmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{P}_\mathrm{L} = \frac{1}{Z_\mathrm{L}} \boldsymbol{K} \boldsymbol{P}_\mathrm{L}.\end{aligned} \tag{2}$$

As a result, $\hat{\boldsymbol{p}}$ is calculated from the corresponding homogeneous global coordinates $\hat{\boldsymbol{P}}_\mathrm{G}$ using Eq. (3). Using calibration (Weng et al., 1992; Heikkila and Silven, 1997; Zhang ZY, 2000; Kannala and Brandt, 2006; Tardif et al., 2006), the extrinsic parameter, intrinsic parameter, and distortion coefficient will be obtained.

$$\hat{\boldsymbol{p}} = \frac{1}{Z_\mathrm{L}} \boldsymbol{K} \left(\boldsymbol{R} \quad \boldsymbol{t}\right) \hat{\boldsymbol{P}}_\mathrm{G}. \tag{3}$$

As shown in Fig. 4, the horizontal FOV and vertical FOV of an imaging aperture are $2\phi_0$ and $2\theta_0$, respectively, where the half-horizontal FOV $\phi_0$ and half-vertical FOV $\theta_0$ are obtained using Eq. (4). When the object depth is $Z$, the pixel resolution is $\Delta x = Z/f_u$ and $\Delta y = Z/f_v$.

$$\begin{cases} \phi_0 = \arctan\dfrac{N_u \delta_u}{2f} = \arctan\dfrac{N_u}{2f_u}, \\ \theta_0 = \arctan\dfrac{N_v \delta_v}{2f} = \arctan\dfrac{N_v}{2f_v}. \end{cases} \tag{4}$$

## 4 Light field acquisition models for the planar structure

In this section, light field acquisition models are derived from the planar structure. Note that the inner optical path of light field cameras is different from that of an ordinary planar ACE. In this section, we will introduce the light field acquisition models.

As a general inductive derivation, this section and the following are based on these assumptions:

1. All single apertures are identical.
2. The structure of the multi-aperture optical imaging system is symmetrical.
3. Sub-images are planar and rectangular.
4. Distortion can be ignored in each sub-image.

### 4.1 Planar ACE

For a planar ACE, the optical centers of all sub-images are coplanar. Except for a few planar ACEs, such as artificial apposition compound eye objective (APCO) (Duparré et al., 2005) and compact large-FOV compound-eye camera (CEC) (Li and Yi, 2012), all the optical axes are parallel. As shown in Fig. 5, all imaging apertures are in a grid pattern on a plane. Supposing that the number of apertures is $N_X \times N_Y$, the intervals between the adjacent apertures are denoted as $\Delta s$ and $\Delta t$ respectively, and $(n_X, n_Y)$ represents the label of each aperture. In this way, the global coordinates of aperture $(n_X, n_Y)$ are $\boldsymbol{P}_\mathrm{G} = \left(s, t, 0\right)^\mathrm{T}$, in which $s = n_X \Delta s$ and $t = n_Y \Delta t$.

According to the mathematical model of single-aperture optical imaging in Section 3.2, for the aperture $(n_X, n_Y)$, the rotation matrix $\boldsymbol{R}$ is obviously the identity matrix, $\boldsymbol{R} = \boldsymbol{E}$, and the translation vector is $\boldsymbol{t} = (-s, -t, 0)^{\mathrm{T}}$. Then, if the global coordinates of an object $\boldsymbol{P}_{\mathrm{G}} = (x, y, Z)^{\mathrm{T}}$ are given, the pixel coordinates $\boldsymbol{p}$ can be computed using Eq. (3). As shown in Fig. 6, assuming that the object plane is $Z$ away from the planar ACE, every aperture will have a corresponding sub-image region on the object plane. In a common overlapping region, an object point is recorded by all apertures.

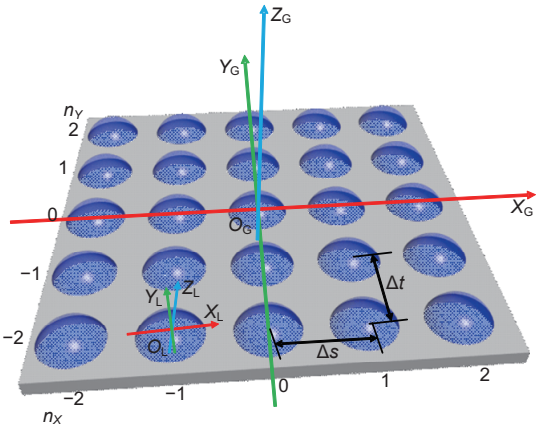


**Fig. 5 Schematic of a general planar ACE, where all apertures are in a grid pattern on a plane**

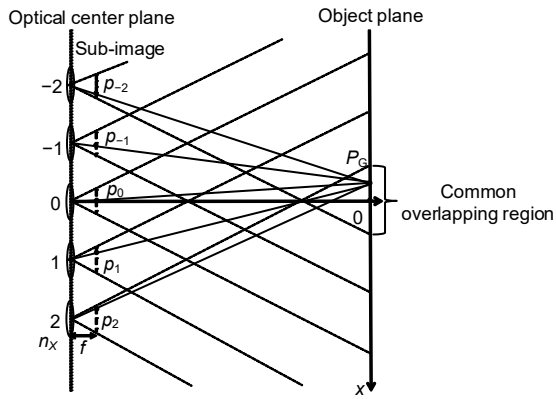


**Fig. 6 Light field acquisition of planar ACE. Every aperture will have a sub-image region on the object plane. In addition, an object point in a common overlapping region is recorded by all apertures**

The whole FOV of planar ACE is the same as that of a single aperture. See Appendix A for details. For adjacent apertures, when object depth $Z$ satisfies $Z \geq f_u \Delta s / N_u$, the sub-images overlap. The sub-image overlap ratio, which is the ratio of the overlapping image region size to the sub-image region size, can be obtained using Eq. (5):

$$\eta_X = 1 - \frac{f_u \Delta s}{N_u Z}. \tag{5}$$

According to the similarity theory, the whole object region that a planar ACE can record is

$$\begin{cases} x_{\min} = n_{X(\min)} \Delta s + \dfrac{Z}{f_u}(1 - c_u), \\ x_{\max} = n_{X(\max)} \Delta s + \dfrac{Z}{f_u}(N_u - c_u). \end{cases} \tag{6}$$

Thus, the width of the whole object region is $(N_u - 1)Z/f_u + (N_X - 1)\Delta s$, and the height is $(N_v - 1)Z/f_v + (N_Y - 1)\Delta t$. $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are used to denote rounding down and up, respectively, and the equivalent resolution $N_x \times N_y$ of the whole object region is

$$\begin{cases} N_x = \left\lfloor (N_X - 1)\Delta s \dfrac{f_u}{Z} \right\rfloor + N_u - 1, \\ N_y = \left\lfloor (N_Y - 1)\Delta t \dfrac{f_v}{Z} \right\rfloor + N_v - 1. \end{cases} \tag{7}$$

Note that when object depth $Z$ satisfies a specific condition as shown in inequality (8), all apertures would have a common overlapping region referring to Eq. (9):

$$Z \geq \max\left( \frac{f_u \Delta s\,(N_X - 1)}{N_u}, \frac{f_v \Delta t\,(N_Y - 1)}{N_v} \right), \tag{8}$$

$$\begin{cases} x'_{\min} = n_{X(\max)} \Delta s + \dfrac{Z}{f_u}(1 - c_u), \\ x'_{\max} = n_{X(\min)} \Delta s + \dfrac{Z}{f_u}(N_u - c_u). \end{cases} \tag{9}$$

In this case, a point on the object plane is recorded by all sub-images. The size of the common overlapping region is $(N_u - 1)Z/f_u - (N_X - 1)\Delta s$. The adjacent aperture will record this point with pixel shift, which is called parallax. The parallax denoted as $\Delta p$ is computed by referring to $\Delta p = \lfloor f_u \Delta s / Z \rfloor$. The larger $\Delta s$ is, the more obvious $\Delta p$ will be. In vision applications, $\Delta s$ is termed the baseline.

Ignoring subpixel shift, the equivalent resolution $N_{x'} \times N_{y'}$ of the common overlapping region is shown in Eq. (10):

$$\begin{cases} N_{x'} = N_u - 1 - \left\lceil (N_X - 1)\Delta s \dfrac{f_u}{Z} \right\rceil, \\ N_{y'} = N_v - 1 - \left\lceil (N_Y - 1)\Delta t \dfrac{f_v}{Z} \right\rceil. \end{cases} \tag{10}$$

The corresponding values are set for quantitative analysis using the above equations. Assume that the number of apertures is $N_X = 5$, every sub-image resolution is $N_u = 640$ without nonlinear distortion, pixel focal length is $f_u = 800$, and the pixel coordinate of the optical center is $c_u = 320$. If $Z$ increases from 0.3 m to 10 m and $\Delta s$ increases from 0.03 m to 0.2 m, the sub-image overlap ratio, the size of the whole object region, and the size of the common overlapping region will vary (Fig. 7).

If $\Delta s = 0.1$ m, when $Z$ increases from 1 m to 10 m, the sizes of the whole object region and the common overlapping region will vary linearly from 1.199 m to 8.387 m and 0.3987 m to 7.587 m, respectively. However, for $\eta_X$, when $Z$ is 0.5, 3, 5.5, and 8 m, the corresponding $\eta_X$ is 0.75, 0.9583, 0.9773, and 0.9844, respectively. Therefore, as object depth $Z$ becomes larger, the sub-image overlap ratio approaches 1.

Let $Z = 5$ m. As $\Delta s$ increases from 0.03 m to 0.2 m, the whole object region is going to increase linearly from 4.114 m to 4.794 m, but $\eta_X$ and the common overlapping region size will decrease linearly from 0.9925 to 0.95 and 3.874 m to 3.194 m, respectively. In addition, when $\Delta s = 0.1$ m and $Z = 5$ m, the common overlapping region is 3.594 m, whose equivalent resolution is $N_{x'} = 575$ according to Eq. (10).

### 4.2 Light field camera

As Fig. 6 shows, an object point $\boldsymbol{P}_{\mathrm{G}}$ in the common overlapping region is recorded by all imaging apertures (i.e., $\boldsymbol{p}_{-2}$–$\boldsymbol{p}_2$). This makes it possible to capture the angular light field information so that the three-dimensional (3D) scene can be recovered from two-dimensional (2D) sub-images. Based on this idea, light field imaging (Wu G et al., 2017; Zhu H et al., 2017) emerged.

Moon and Spencer (1953) defined a light field as a complete collection of rays in space. Later, Adelson and Bergen (1991) proposed a seven-dimensional (7D) function $L(\boldsymbol{P}, \boldsymbol{\omega}, \lambda, t)$, called the plenoptic function, to represent a light field. The plenoptic function models a ray with eight parameters: position $\boldsymbol{P} = (x, y, z)^{\mathrm{T}}$, direction $\boldsymbol{\omega} = (\theta, \phi)^{\mathrm{T}}$, wavelength $\lambda$, time $t$, and brightness $|L|$. The 7D function is seldom used due to its complicated calculation. Fortunately, because the dynamic light field can be captured in continuous frames, the time $t$ of the plenoptic function can be ignored. In addition, most imaging sensors are CCD or CMOS, having red, green, and blue channels, so the wavelength $\lambda$ can be ignored as well. Without wavelength $\lambda$ and time $t$, the 5D function $L(\boldsymbol{P}, \boldsymbol{\omega})$, as shown in Fig. 8a, is used to represent the light field (McMillan and Bishop, 1995). However, the dimension of the 5D function can be further reduced. Levoy and Hanrahan (1996) and Gortler et al. (1996) proposed a two-parallel-plane (2PP) model to represent the light field (Fig. 8b). In their proposed model, a ray can be represented by two intersections (e.g., $(s, t)^{\mathrm{T}}$ on plane $st$ and $(x, y)^{\mathrm{T}}$ on plane $xy$) when the interval between two parallel planes is $Z$.

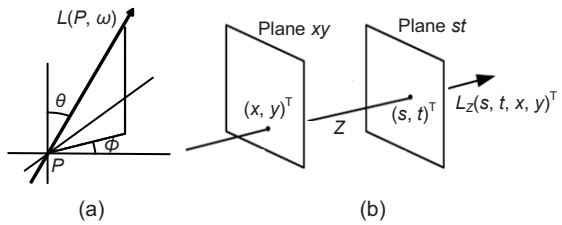


**Fig. 8 Light field representation: (a) a ray is recorded by position $\boldsymbol{P}$ and direction $\boldsymbol{\omega}$ in the 5D function; (b) a ray is recorded by two intersections $(\boldsymbol{s}, \boldsymbol{t})^{\mathrm{T}}$ and $(\boldsymbol{x}, \boldsymbol{y})^{\mathrm{T}}$ in the two-parallel-plane (2PP) model**

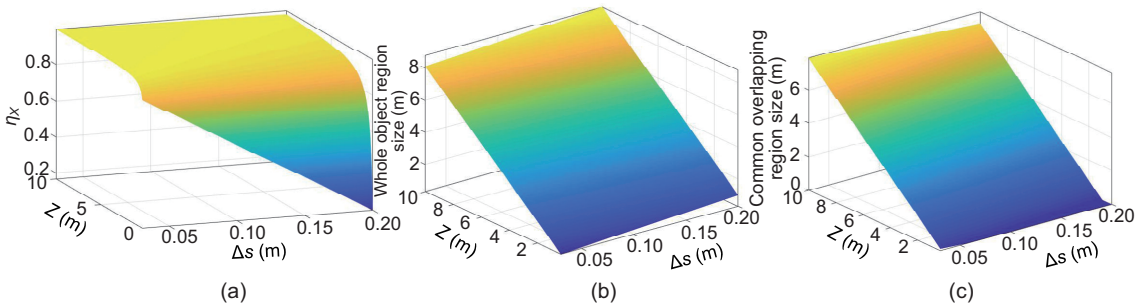


**Fig. 7 Quantitative analysis: (a) variation trend of $\eta_X$; (b) variation trend of the whole object region size; (c) variation trend of the common overlapping region size**

Compared with other light field representations, it is convenient to calculate the 4D function $L_Z(s,t,x,y)$ because it has the least dimension. More importantly, the 4D function allows the light field information to be used flexibly in the space domain (Bolles et al., 1987; Isaksen et al., 2000; Lin ZC and Shum, 2004) and frequency domain (Chai et al., 2000; Zhang C and Chen, 2003; Durand et al., 2005; Ng, 2005; Soler et al., 2009), like the digital signal. Hence, the 2PP model is widely used in light field imaging. Taking a planar ACE as an example, if the object depth is $Z$, the recorded space resolution of the light field $L_Z(s,t,x,y)$ is determined by the equivalent resolution of the common overlapping region $N_{x'} \times N_{y'}$, and the angular resolution is determined by the number of apertures $N_X \times N_Y$.

Light field cameras based on microlens arrays can obtain light field information in a single photographic exposure (Ng and Hanrahan, 2005). Compared to a planar ACE, light field cameras have the advantage of being small and easy to carry, and thus are popular among researchers and photographers. There are mainly two types of light field cameras based on a microlens array: Plenoptic 1.0 (Ng and Hanrahan, 2005; Levoy et al., 2006) and Plenoptic 2.0 (Lumsdaine and Georgiev, 2009). In the rest of this subsection, the light field acquisition models of these two kinds of light field cameras are summarized.

### 4.2.1 Plenoptic 1.0

Without loss of generality, as shown in Fig. 9, it is assumed that there are $N_x \times N_y$ microlenses in a light field camera, and each microlens label is denoted as $(n_x, n_y)$. On the imaging sensor, every microlens has a corresponding pixel region with $N_u \times N_v$ pixels, and the coordinates of a pixel in the corresponding pixel region are denoted by $\boldsymbol{p}_{(n_x,n_y)} = (u,v)^{\mathrm{T}}$. The diaphragm diameter of the main lens is $d_{\mathrm{main}}$ and that of the microlens is $d_{\mathrm{mic}}$. In terms of the main lens, the principal distance $f_{\mathrm{main}}$ is a little larger than the focal length $F_{\mathrm{main}}$.

Now that light field cameras are designed to obtain light field information, the 2PP model is used to represent the light field acquisition for Plenoptic 1.0 (Fig. 10). The viewpoint plane puts the focal length $F_{\mathrm{main}}$ in front of the principal plane of the main lens (Hahne et al., 2014a, 2014b). The microlens array is located at the image plane of the main lens, so that rays focused by the main lens are separated by the microlens array.

An important characteristic of Plenoptic 1.0 is that the main lens and all microlenses have the same $f$-number, which is defined as the ratio of the principal length to the size of the diaphragm (Ng and Hanrahan, 2005). In this case, the principal length of the microlens is $f_{\mathrm{mic}} = d_{\mathrm{mic}} f_{\mathrm{main}} / d_{\mathrm{main}}$. Notice that adjacent pixel regions will not overlap if the $f$-number is reasonable. Specifically, when $d_{\mathrm{main}}$ is the same as the size of the main lens, adjacent pixel regions are tangent to each other, and thus pixels are used effectively.

If the coordinates of a pixel $\boldsymbol{p}_{(n_x,n_y)} = (u,v)^{\mathrm{T}}$ are offered, a ray $L(s,t,x,y)$ in the 2PP model will be obtained referring to Eqs. (11) and (12), which is called decoding. Through decoding, we will obtain sub-images as planar ACEs do. The resolution of every sub-image is $N_x \times N_y$, and the number of viewpoints is $N_u \times N_v$.

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \left(u - \frac{N_u+1}{2}\right) \dfrac{d_{\mathrm{main}}}{N_u} \\ \left(v - \frac{N_v+1}{2}\right) \dfrac{d_{\mathrm{main}}}{N_v} \end{pmatrix}, \tag{11}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s - \left(n_x - \frac{N_x+1}{2}\right) \dfrac{d_{\mathrm{mic}}(Z - F_{\mathrm{main}})}{F_{\mathrm{main}}} \\ t - \left(n_y - \frac{N_y+1}{2}\right) \dfrac{d_{\mathrm{mic}}(Z - F_{\mathrm{main}})}{F_{\mathrm{main}}} \end{pmatrix}. \tag{12}$$

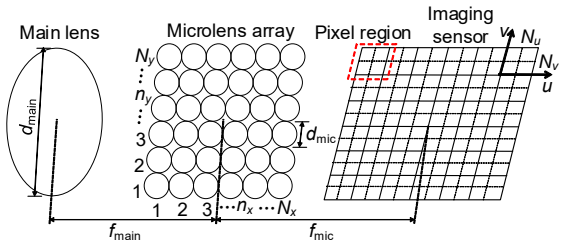


**Fig. 9 Internal structure of light field cameras**

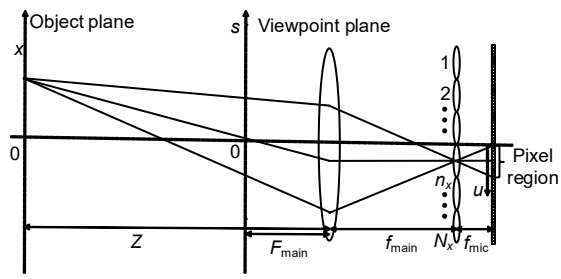


**Fig. 10 Light field acquisition of Plenoptic 1.0. The viewpoint plane puts the focal length $F_{\mathrm{main}}$ in front of the principal plane of the main lens. The microlens array is located at the image plane of the main lens**

For every sub-image, the pixel resolution on the object plane is $\Delta x = d_{\mathrm{mic}}(Z - F_{\mathrm{main}})/F_{\mathrm{main}}$, while the viewpoint interval is $\Delta s = d_{\mathrm{main}}/N_u$. In addition, when $n_x$ and $u$ are taken to be the maximum and minimum respectively, when the object depth is $Z$, the range of the object region is shown in Eq. (13):

$$\begin{cases} x_{\min} = \\ -\dfrac{(N_u-1)d_{\mathrm{main}}}{2N_u} - \dfrac{(N_x-1)(Z-F_{\mathrm{main}})d_{\mathrm{mic}}}{2F_{\mathrm{main}}}, \\ x_{\max} = \\ \dfrac{(N_u-1)d_{\mathrm{main}}}{2N_u} + \dfrac{(N_x-1)(Z-F_{\mathrm{main}})d_{\mathrm{mic}}}{2F_{\mathrm{main}}}. \end{cases} \tag{13}$$

Therefore, the size of the whole object region is $d_{\mathrm{main}} + N_x d_{\mathrm{mic}}(Z - F_{\mathrm{main}})/F_{\mathrm{main}}$. The sub-image overlap ratio $\eta_X$ can be obtained using Eq. (14):

$$\eta_X = 1 - \frac{d_{\mathrm{main}}F_{\mathrm{main}}}{N_x N_u d_{\mathrm{mic}}(Z - F_{\mathrm{main}})}. \tag{14}$$

Because $F_{\mathrm{main}}$ is almost negligible compared to $Z$ and $N_x d_{\mathrm{mic}}$ is approximately $d_{\mathrm{main}}$, $\eta_X$ can be considered as 1 in practice. As a result, the number of microlenses $N_x \times N_y$ determines the space resolution of the light field, and the angular resolution of the light field is determined by the number of pixels in a pixel region, which is $N_u \times N_v$.

It seems that Plenoptic 1.0 is equivalent to an $N_u \times N_v$ planar ACE, whose resolution of each imaging sensor is $N_x \times N_y$ and the pixel focal length is $f_u = f_v = ZF_{\mathrm{main}}/(d_{\mathrm{mic}}(Z - F_{\mathrm{main}}))$. The baselines are $\Delta s = d_{\mathrm{main}}/N_u$ and $\Delta t = d_{\mathrm{main}}/N_v$.

According to Ng and Hanrahan (2005), the proposed Plenoptic 1.0 has $296 \times 296$ microlenses, whose pixel region is $14 \times 14$. The focal length of a microlens is 500 μm, and the diaphragm diameters of a microlens and main lens are $d_{\mathrm{mic}} = 125$ μm and $d_{\mathrm{main}} = 35$ mm, respectively. After calculation, the parameter of an equivalent planar ACE is as follows: the number of apertures is $14 \times 14$, the resolution of each imaging sensor is $296 \times 296$, the pixel focal length is $f_u = f_v = 1152$ when $Z = 5$ m, and the baselines are $\Delta s = \Delta t = 2.5$ mm.

### 4.2.2 Plenoptic 2.0

Better than Plenoptic 1.0, Plenoptic 2.0 can provide a variable tradeoff between space and angular resolution (Lumsdaine and Georgiev, 2009). Fig. 11 shows the 2PP model of light field acquisition for Plenoptic 2.0. The focused image plane of the main lens is in front of the microlens array with distance $a$. The interval between the microlens array and the imaging sensor is $b$. In this way, the viewpoint plane coincides with the principal plane of the main lens.

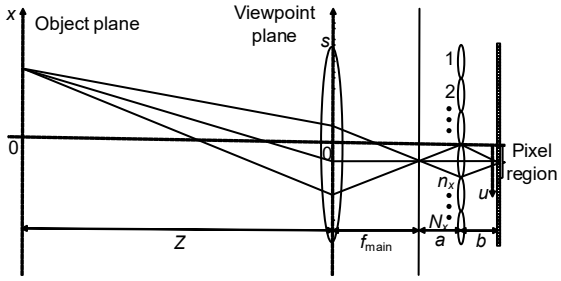


**Fig. 11 Light field acquisition of Plenoptic 2.0. The focused image plane of the main lens is in front of the microlens array with distance *a***

In this case, according to Eqs. (15) and (16), we will decode the light field information from all pixels on the imaging sensor.

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \left(-\left\lceil \dfrac{au}{bN_u} \right\rceil + \dfrac{1}{2}\left\lfloor \dfrac{a}{b} \right\rfloor + \dfrac{1}{2}\right)\dfrac{d_{\mathrm{mic}}Z}{f_{\mathrm{main}}} \\ \left(-\left\lceil \dfrac{av}{bN_v} \right\rceil + \dfrac{1}{2}\left\lfloor \dfrac{a}{b} \right\rfloor + \dfrac{1}{2}\right)\dfrac{d_{\mathrm{mic}}Z}{f_{\mathrm{main}}} \end{pmatrix}, \tag{15}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \left(n_x - \dfrac{N_x+1}{2} - \left(u - \dfrac{N_u+1}{2}\right)\dfrac{a}{bN_u}\right)\dfrac{d_{\mathrm{mic}}Z}{f_{\mathrm{main}}} \\ \left(n_y - \dfrac{N_y+1}{2} - \left(v - \dfrac{N_v+1}{2}\right)\dfrac{a}{bN_v}\right)\dfrac{d_{\mathrm{mic}}Z}{f_{\mathrm{main}}} \end{pmatrix}. \tag{16}$$

On the object plane, the pixel resolution of a sub-image is $\Delta x = ad_{\mathrm{mic}}Z/(bN_u f_{\mathrm{main}})$. In addition, the viewpoint interval is $\Delta s = d_{\mathrm{mic}}Z/f_{\mathrm{main}}$. Taking $n_x$ and $u$ to be the maximum and minimum respectively, the range of the object region will be obtained as follows:

$$\begin{cases} x_{\min} = -\left(N_x - 1 + (N_u - 1)\dfrac{a}{bN_u}\right)\dfrac{d_{\mathrm{mic}}Z}{2f_{\mathrm{main}}}, \\ x_{\max} = \left(N_x - 1 + (N_u - 1)\dfrac{a}{bN_u}\right)\dfrac{d_{\mathrm{mic}}Z}{2f_{\mathrm{main}}}, \end{cases} \tag{17}$$

and the size of the whole object region is $(N_x + a/b)\, d_{\mathrm{mic}}Z/f_{\mathrm{main}}$.

The sub-image overlap ratio $\eta_X$ can be computed using Eq. (18):

$$\eta_X = 1 - \frac{1}{N_x}. \tag{18}$$

In practical applications, $\eta_X$ could be considered as 1, similar to the setting in Plenoptic 1.0. However, in contrast to Plenoptic 1.0, in Plenoptic 2.0 the resolution of every sub-image is $(N_x N_u b/a) \times (N_y N_v b/a)$ and the number of viewpoints is $(a/b) \times (a/b)$. The former represents the space resolution of the light field, while the latter is the angular resolution. Obviously, the ratio of $a$ to $b$ determines the tradeoff between the space and angular resolution of light field acquisition.

Plenoptic 2.0 can also be equivalent to a planar ACE. This planar ACE has $(a/b) \times (a/b)$ apertures; the resolution of each imaging sensor is $(N_x N_u b/a) \times (N_y N_v b/a)$ and the pixel focal length is $f_u = bN_u f_{\text{main}}/(ad_{\text{mic}})$. In addition, the baselines are $\Delta s = \Delta t = d_{\text{mic}} Z/f_{\text{main}}$.

Lumsdaine and Georgiev (2009) proposed Plenoptic 2.0, which has $130 \times 122$ microlenses whose pixel region is $32 \times 32$. The focal lengths of the microlenses and the main lens are 750 μm and $f_{\text{main}} = 140$ mm, respectively. The diaphragm diameter of a microlens is $d_{\text{mic}} = 250$ μm. Assuming that $a/b = 8$, after calculation, the equivalent planar ACE is as follows: the number of apertures is $8 \times 8$, the resolution of each imaging sensor is $520 \times 488$, and the pixel focal length is $f_u = f_v = 2240$. If $Z = 5$ m, the baselines will be $\Delta s = \Delta t = 8.93$ mm.

## 5 Light field acquisition models for the convex structure

Compared with the planar structure, the optical centers of all sub-images are not coplanar for the convex structure. In this section, the light field acquisition models of three representative arrangements for the convex structure, spherical multi-loop, spherical multi-row, and cylinder, are summarized.

### 5.1 Spherical multi-loop arrangement

Most convex ACEs (Brady et al., 2012; Guo et al., 2012; Afshari et al., 2013; Cao et al., 2015; Luo et al., 2015; Pang et al., 2017; Shi et al., 2017; Yu et al., 2019; Zhang JM et al., 2020; Zhou et al., 2020) are in a spherical multi-loop arrangement. In this arrangement, imaging apertures are divided into $N_\theta$ latitude loops, and all optical centers located in the same loop have the same latitude angle, as shown in Fig. 12.

Suppose that the radius of the sphere is $R$ and that each loop is evenly spaced with the latitude interval $\Delta\theta$. At loop $n_\theta$, when the number of apertures uniformly arranged is $N_{\phi(n_\theta)}$, the longitude interval between two adjacent sub-eyes is $\Delta\phi(n_\theta) = 360°/N_{\phi(n_\theta)}$. The label of each aperture is denoted as $(n_\theta, n_{\phi(n_\theta)})$, in which $n_{\phi(n_\theta)}$ is the counterclockwise number for loop $n_\theta$ in the top view. Thus, the global coordinates of aperture $(n_\theta, n_{\phi(n_\theta)})$ could be computed using spherical coordinates $(\theta_{\text{G}}, \phi_{\text{G}})$. The zenith angle and azimuth angle are $\theta_{\text{G}} = n_\theta \Delta\theta$ and $\phi_{\text{G}} = n_{\phi(n_\theta)} \Delta\phi(n_\theta) + \phi_0(n_\theta)$, respectively, in which $\phi_0(n_\theta)$ stands for azimuth deviation.

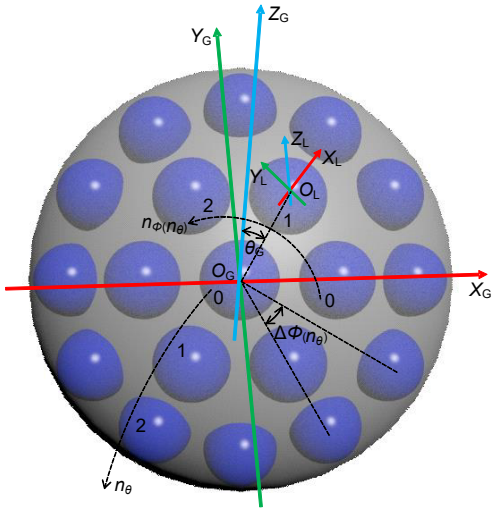


**Fig. 12 Spherical multi-loop arrangement**

The local coordinate system, $O_{\text{L}} - X_{\text{L}} Y_{\text{L}} Z_{\text{L}}$, is established for each aperture. The $Z_{\text{L}}$ axis coincides with the optical axis, whose extension line passes through $O_{\text{G}}$. In this case, the rotation matrix $\boldsymbol{R}$ of aperture $(n_\theta, n_{\phi(n_\theta)})$ is shown in Eq. (19), and translation vector $\boldsymbol{t}$ is $(0, 0, -R)^{\text{T}}$:

$$\boldsymbol{R} = \begin{pmatrix} \cos\theta_{\text{G}}\cos\phi_{\text{G}} & \cos\theta_{\text{G}}\sin\phi_{\text{G}} & -\sin\theta_{\text{G}} \\ -\sin\phi_{\text{G}} & \cos\phi_{\text{G}} & 0 \\ \sin\theta_{\text{G}}\cos\phi_{\text{G}} & \sin\theta_{\text{G}}\sin\phi_{\text{G}} & \cos\theta_{\text{G}} \end{pmatrix}. \quad (19)$$

Fig. 13 shows the side view of this arrangement. In this arrangement, the maximum half-FOV of light field acquisition is $(N_\theta - 1)\Delta\theta + \arctan((Z - R)\tan\phi_0/Z)$; see Appendix B for details. If the adjacent apertures overlap, $\Delta\theta < 2\phi_0$ should be satisfied. The red shaded region is the overlapping sub-image region of adjacent apertures. The minimum depth of the spherical object surface $Z_{\min}$ is computed using Eq. (20), and the sub-image overlap ratio is

computed using Eq. (21):

$$Z_{\min} = R\frac{\sin\phi_0\cos\frac{\Delta\theta}{2}}{\sin(\phi_0 - \frac{\Delta\theta}{2})}, \tag{20}$$

$$\eta_\theta = 1 - \frac{\Delta\theta}{2\phi_0 - 2\arcsin\frac{R\sin\phi_0}{Z}}. \tag{21}$$

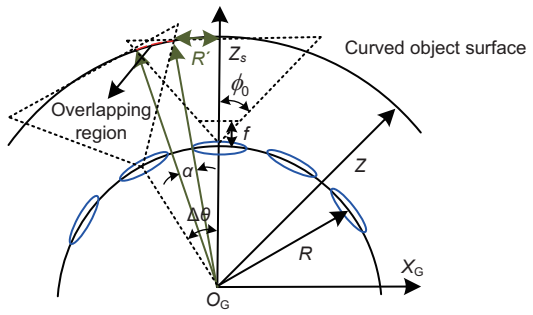


**Fig. 13 Vertical view of a convex ACE**

One of the key issues in convex ACEs is how to obtain a panoramic image. Two determinants need to be considered: distribution density of apertures and full-view coverage distance (FCD) (Afshari et al., 2013). A well-established geometry concept known as the Voronoi diagram (Aurenhammer, 1991; de Berg et al., 2001), which is applied on a spherical surface, is used to determine the parameters of this arrangement, so that each direction of the curved object surface is observable by at least one aperture. Afshari et al. (2013) and Wang YW et al. (2017) analyzed how to obtain a panoramic image. In their opinion, as shown in Fig. 14, when every loop is evenly spaced with the latitude interval $\Delta\theta$, for the apertures at the same loop, the edge of each aperture's object plane in the non-overlapping part should be seamlessly connected to obtain a panoramic image.

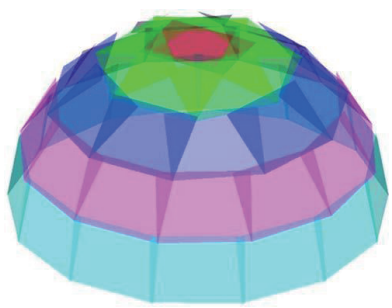

**Fig. 14 The object planes of a convex ACE in a spherical multi-loop arrangement. The edge of each aperture's object plane in the non-overlapping part is seamlessly connected. Reprinted from Wang YW et al. (2017), Copyright 2017, with permission from John Wiley and Sons**

In this case, the number of apertures located at loop $n_\theta$ $(0 < n_\theta < N_\theta - 1)$ usually satisfies inequality (22). Also, the full-view coverage distance is $Z_{\mathrm{FCD}} = Z_{\min} = R\sin\phi_0\cos\frac{\Delta\theta}{2}/\sin(\phi_0 - \frac{\Delta\theta}{2})$.

$$N_{\phi(n_\theta)} \geq \left\lceil \frac{180^\circ}{\arctan\frac{(Z-R)\tan\theta_0}{Z\sin(\theta_{\mathrm{G}}+\Delta\theta-\phi_0+\arcsin\frac{R\sin\phi_0}{Z})}} \right\rceil. \tag{22}$$

Specifically, as for the last loop, $n_\theta = N_\theta - 1$, $N_{\phi(n_\theta)}$ ought to satisfy inequality (23), in which $\varphi = \arctan\frac{(Z-R)\tan\phi_0}{Z}$.

$$N_{\phi(N_\theta-1)} \geq \left\lceil \frac{180^\circ}{\arctan\frac{\tan\theta_0\sin\varphi}{\tan\phi_0\sin(\varphi+\theta_{\mathrm{G}})}} \right\rceil. \tag{23}$$

Assuming that for every sub-image, the resolution is $N_u = 640$ without nonlinear distortion and the pixel focal length is $f_u = 800$, according to Eqs. (20) and (21) and inequality (22), the corresponding values are set for quantitative analysis. When the radius of the sphere is $R = 0.1$ m, if $\Delta\theta$ increases from 10° to 30° and $Z$ increases from 0.5 m to 10 m, the FCD, sub-image overlap ratio, and minimum number of apertures located at loop $n_\theta = 1$ will vary (Fig. 15).

When $\Delta\theta$ is 10°, 15°, 20°, 25°, and 30°, the corresponding $Z_{\mathrm{FCD}}$ is 0.128, 0.1491, 0.1788, 0.2243, and 0.3029 m, respectively. As $\Delta\theta$ increases, the FCD is going to increase faster and faster. If $Z = 5$ m, when $\Delta\theta$ increases from 10° to 30°, $\eta_\theta$ will decrease linearly from 0.7661 to 0.2983. In addition, if $\Delta\theta = 30^\circ$, when $Z$ is 0.5, 3, 5.5, 8, and 10 m, the corresponding $\eta_\theta$ is 0.1449, 0.2892, 0.2995, 0.3035, and 0.3051, respectively. With an increase in the spherical object surface's depth, the sub-image overlap ratio will approach $1 - \Delta\theta/(2\phi_0) = 0.312$. Let $\Delta\theta = 15^\circ$ and $Z = 5$ m; there should be at least three apertures located at loop $n_\theta = 1$ to ensure full-view coverage.

### 5.2 Spherical multi-row arrangement

Part of the convex ACE (Zhang YK et al., 2010; Song et al., 2013; Deng et al., 2016) is its spherical multi-row arrangement. As in the spherical multi-loop arrangement, all apertures are located on the

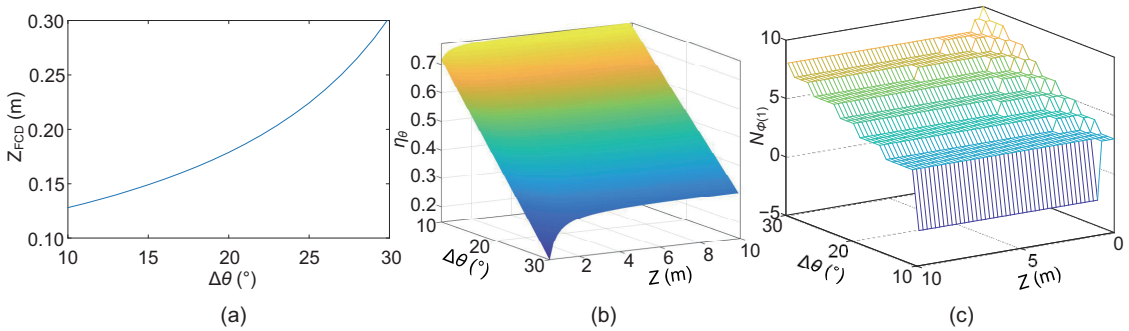


**Fig. 15 Quantitative analysis: (a) variation trend of $Z_{\mathrm{FCD}}$ when $\Delta\theta$ varies; (b) variation trend of $\eta_\theta$ when $\Delta\theta$ and $Z$ vary; (c) variation trend of $N_{\phi(1)}$ when $\Delta\theta$ and $Z$ vary**

sphere with radius $R$ and all optical axis extension lines intersect at spherical center $O_{\mathrm{G}}$ (Fig. 16).

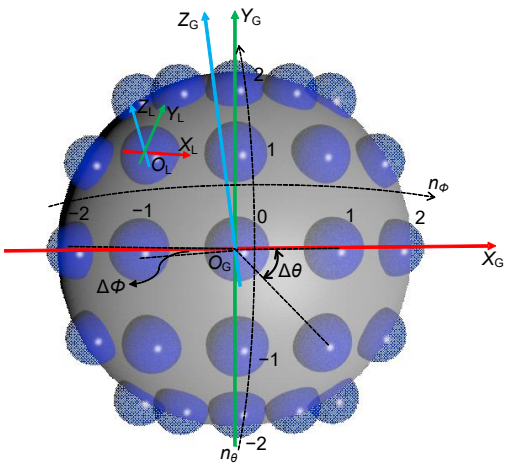


**Fig. 16 Spherical multi-row arrangement**

Suppose that the number of apertures is $N_\phi \times N_\theta$ and that the label of each aperture is denoted as $(n_\phi, n_\theta)$. Then the global coordinates of aperture $(n_\phi, n_\theta)$ could be computed using polar coordinates $(\phi_{\mathrm{G}}, \theta_{\mathrm{G}})$, in which $\phi_G$ and $\theta_{\mathrm{G}}$ are the azimuth angle and elevation angle, respectively. If the elevation angle between adjacent apertures is $\Delta\theta$ and the azimuth angle is $\Delta\phi$, the azimuth and elevation angles of each aperture will be $\phi_{\mathrm{G}} = n_\phi \Delta\phi$ and $\theta_{\mathrm{G}} = n_\theta \Delta\theta$, respectively.

Taking the sub-image's optical center of each aperture as the origin to establish a local coordinate system, the $Z_{\mathrm{L}}$ axis is along the optical axis, while the $X_{\mathrm{L}}$ axis is parallel to plane $Z_{\mathrm{G}} O_{\mathrm{G}} X_{\mathrm{G}}$. For the extrinsic aperture parameter $(n_\phi, n_\theta)$, the rotation matrix $\boldsymbol{R}$ is shown in Eq. (24) and translation vector $\boldsymbol{t}$ is $\left(0, 0, -R\right)^{\mathrm{T}}$.

$$\boldsymbol{R} = \begin{pmatrix} \cos\phi_{\mathrm{G}} & 0 & -\sin\phi_{\mathrm{G}} \\ -\sin\theta_{\mathrm{G}}\sin\phi_{\mathrm{G}} & \cos\theta_{\mathrm{G}} & -\sin\theta_{\mathrm{G}}\cos\phi_{\mathrm{G}} \\ \cos\theta_{\mathrm{G}}\sin\phi_{\mathrm{G}} & \sin\theta_{\mathrm{G}} & \cos\theta_{\mathrm{G}}\cos\phi_{\mathrm{G}} \end{pmatrix}. \tag{24}$$

If the adjacent sub-images overlap, $\Delta\phi < 2\phi_0$ and $\Delta\theta < 2\theta_0$ should be satisfied. When the spherical object surface has a depth of $Z$, the vertical sub-image overlap ratio of adjacent apertures can be computed using Eq. (25):

$$\eta_\theta = 1 - \frac{\Delta\theta}{2\theta_0 - 2\arcsin\dfrac{R\sin\theta_0}{Z}}. \tag{25}$$

When $\theta_{\mathrm{G}} = 0$, the minimum horizontal sub-image overlap ratio is computed using Eq. (26):

$$\eta_\phi = 1 - \frac{\Delta\phi}{2\phi_0 - 2\arcsin\dfrac{R\sin\phi_0}{Z}}. \tag{26}$$

To obtain a panoramic image, the depth of the spherical object surface should satisfy inequality (27):

$$Z \geq \max\left( R\frac{\sin\phi_0\cos\dfrac{\Delta\phi}{2}}{\sin(\phi_0 - \dfrac{\Delta\phi}{2})}, R\frac{\sin\theta_0\cos\dfrac{\Delta\theta}{2}}{\sin(\theta_0 - \dfrac{\Delta\theta}{2})} \right). \tag{27}$$

In this case, the vertical FOV of light field acquisition is $(N_\theta - 1)\Delta\theta + 2\arctan((Z - R)\tan\theta_0/Z)$. In addition, the horizontal FOV is $2\left((N_\phi - 1)\arcsin(\cos\theta_{\mathrm{G}}\sin\frac{\Delta\phi}{2}) + \arctan\frac{(Z-R)\tan\phi_0}{Z\cos\theta_{\mathrm{G}}}\right)$.

Many qualitative conclusions are consistent with the convex ACE with a spherical multi-loop arrangement. For example, with the increase of $\Delta\phi$

or $\Delta\theta$, the horizontal or vertical FOV will increase linearly, but the sub-image overlap ratio will decrease linearly.

### 5.3 Cylinder arrangement

Leitel et al. (2014) proposed a typical convex ACE in a cylinder arrangement. To summarize the general light field acquisition model of the cylinder arrangement, it is assumed that all vertical apertures are parallel. As shown in Fig. 17, similar to the spherical multi-row arrangement, $N_\phi \times N_Y$ apertures are arranged on the surface of a cylinder whose radius is $R$. The vertical interval between adjacent apertures is $\Delta Y$, and $\Delta\phi$ is used to represent the azimuth angle between adjacent horizontal apertures. Likewise, if $(n_\phi, n_Y)$ is denoted as the label of each aperture, then the global coordinates of aperture $(n_\phi, n_Y)$ can be computed using columnar coordinates $(\phi_\mathrm{G}, Y_\mathrm{G})$, in which the azimuth angle is $\phi_\mathrm{G} = n_\phi \Delta\phi$ and the height is $Y_\mathrm{G} = n_Y \Delta Y$.

Any aperture takes a corresponding sub-image's optical center as the origin to establish a local coordinate system $O_\mathrm{L} - X_\mathrm{L}Y_\mathrm{L}Z_\mathrm{L}$, in which the $Z_\mathrm{L}$ axis is along the optical axis and the $Y_\mathrm{L}$ axis is parallel to the $Y_\mathrm{G}$ axis. The rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ for the aperture $(n_\phi, n_Y)$ are shown in Eqs. (28) and (29), respectively:

$$\boldsymbol{R} = \begin{pmatrix} \cos\phi_\mathrm{G} & 0 & -\sin\phi_\mathrm{G} \\ 0 & 1 & 0 \\ \sin\phi_\mathrm{G} & 0 & \cos\phi_\mathrm{G} \end{pmatrix}, \tag{28}$$

$$\boldsymbol{t} = (0, -Y_\mathrm{G}, -R)^\mathrm{T}. \tag{29}$$

When the depth of the cylindrical object surface is $Z$, the sub-image overlap ratio of adjacent apertures in the horizontal and vertical directions can be computed using Eq. (30):

$$\begin{cases} \eta_\phi = 1 - \dfrac{\Delta\phi}{2\phi_0 - 2\arcsin\dfrac{R\sin\phi_0}{Z}}, \\ \eta_Y = 1 - \dfrac{f_v \Delta Y}{N_v(Z-R)}. \end{cases} \tag{30}$$

To synthesize a panoramic image, the following conditions should be satisfied: $\Delta\phi < 2\phi_0$ and $Z \geq Z_\mathrm{FCD}$, in which $Z_\mathrm{FCD}$ is computed using Eq. (31):

$$Z_\mathrm{FCD} = \max\left(R\frac{\sin\phi_0\cos\frac{\Delta\phi}{2}}{\sin(\phi_0 - \frac{\Delta\phi}{2})}, R + \frac{\Delta Y f_v}{N_v}\right). \tag{31}$$

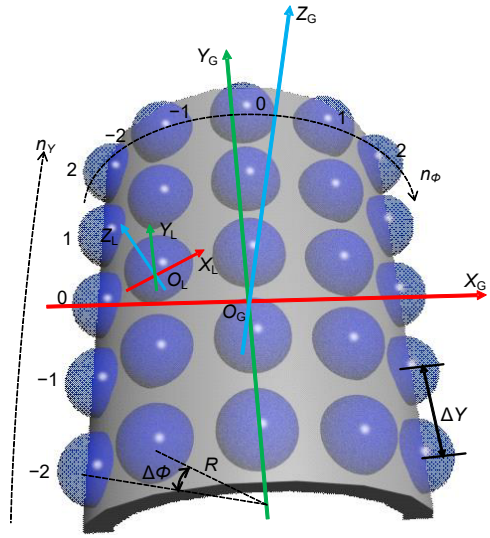


**Fig. 17 Cylinder arrangement**

In this case, the vertical FOV is $2\theta_0$, and the horizontal FOV of light field acquisition is $(N_\phi - 1)\Delta\phi + 2\arctan((Z-R)\tan\phi_0/Z)$. Similar to a planar ACE, the vertical whole object region is $(N_v - 1)(Z - R)/f_v + (N_Y - 1)\Delta Y$, and the size of the vertical common overlapping region is $(N_v - 1)(Z - R)/f_v - (N_Y - 1)\Delta Y$.

Consistent with some qualitative conclusions concerning planar ACEs and convex ACEs in a spherical multi-loop arrangement, if $\Delta\phi$ or $\Delta Y$ increases, the horizontal FOV or vertical whole object region will become larger, but the sub-image overlap ratio will decrease. When the depth of the cylindrical object surface increases, both the size of the vertical whole object region and the sub-image overlap ratio will increase.

## 6 Application analysis of multi-aperture optical imaging systems

Researchers have discussed the applications of multi-aperture optical imaging systems based on ACEs (Gong et al., 2013; Wu SD et al., 2017; Cheng et al., 2019), light field cameras, and camera arrays (Wu G et al., 2017; Zhu H et al., 2017). Apparently, the sub-image overlap ratio for the planar structure is greater than that for the convex structure, and all sub-images have a large common overlapping region. Compared to the planar structure, the whole FOV of the convex structure is much larger.

It is therefore easy to infer that these two kinds

of multi-aperture optical imaging systems have different applications. Planar ACEs and light field cameras based on a microlens array are suitable for light field imaging. Convex ACEs focus on wide FOV imaging, associated with high-definition surveillance and multi-target detection. In this section, based on the summarized light field acquisition models, the application of different structures is analyzed.

In the rest of this section we assume that for every aperture, the imaging sensor resolution is $N_u = N_v = 640$ without nonlinear distortion, the pixel focal length is $f_u = f_v = 800$, and the pixel coordinates of the optical center are $c_u = c_v = 320$. In this case, the half FOV of the sub-image is $\phi_0 = \theta_0 = 21.8014°$.

### 6.1 Application analysis of the planar structure

Planar multi-aperture optical imaging systems are used mainly in light field imaging; basic algorithms include superresolution reconstruction (Bishop et al., 2009; Lim et al., 2009; Georgiev et al., 2011; Wanner and Goldluecke, 2012b; Carles et al., 2014), depth estimation (Wanner and Goldluecke, 2012a, 2014; Kim et al., 2013; Lin HT et al., 2015; Johannsen et al., 2016; Williem and Park, 2016; Wang TC et al., 2018; Wu SD et al., 2018a, 2018b), and refocusing (Vaish et al., 2004; Yang T et al., 2014; Wang YQ et al., 2019). In this subsection, these three types of applications are analyzed.

Assuming that the number of apertures is $N_X = 5$ and that the interval of adjacent apertures is $\Delta s = 0.05$ m, if object depth $Z$ increases from 0.3 m to 50 m, the sub-image overlap ratio $\eta_X$ and the common overlap ratio $\eta$, which is the ratio of the common overlapping region size to the whole object region size, are shown in Fig. 18a. The parallax $\Delta p$ will vary as shown in Fig. 18b.

According to Fig. 18a, when $Z > 10$ m, $\eta_X$ is greater than 0.9511, which means that the pixel utilization of each aperture is high enough to obtain the common overlapping region. Because the common overlapping region can be recorded by all apertures, there are $N_X$ sub-pixels in a pixel region, and a shift occurs within each sub-pixel. Therefore, superresolution reconstruction has been studied for planar ACEs (Tanida et al., 2000; Duparré et al., 2005) and light field cameras (Ng and Hanrahan, 2005; Lumsdaine and Georgiev, 2009) to obtain high-definition images.

According to Fig. 18b, when $Z < 40$ m, the parallax between adjacent sub-images is greater than 1 pixel. In this case, the parallax can be used to determine the depth of an object point. Depth estimation, which is the basis of 3D reconstruction, can be applied to planar multi-aperture optical imaging systems as a result. The depth interval is shown in Fig. 19. When $Z > 10$ m, only three depth intervals can be identified.

The accuracy of depth estimation can be represented by the depth change $\Delta Z$ of a pixel (i.e., $f_u \Delta s/(Z - \Delta Z) - f_u \Delta s/Z = 1$), which is computed using Eq. (32), as shown in Fig. 20. As $Z$ increases, the depth estimation accuracy deteriorates rapidly. However, if $\Delta s$ is set as 0.1, 0.2, 0.3, or 0.4 m when $Z = 25$ m, the corresponding $\Delta Z$ is 5.952, 3.378, 2.358, or 1.812 m, respectively.

$$\Delta Z = \frac{Z^2}{f_u \Delta s + Z}. \tag{32}$$

Hence, depth estimation is used mainly in small

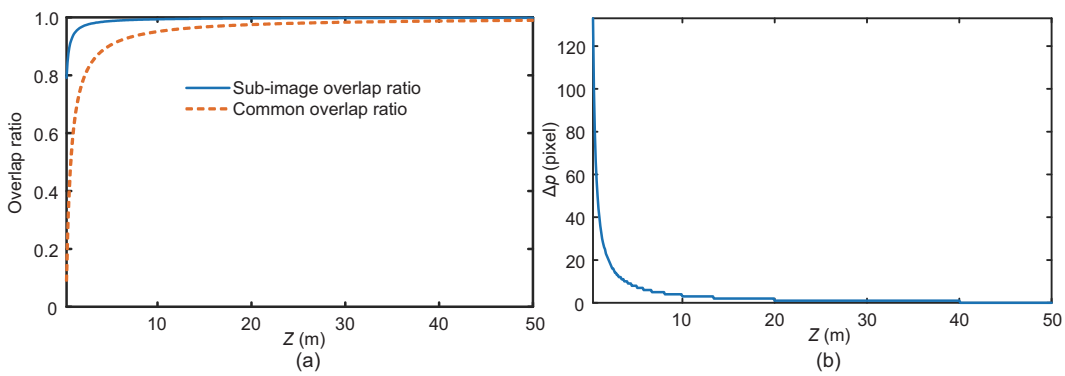


**Fig. 18 The variation trend of the sub-image overlap ratio and common overlap ratio (a) and $\Delta p$ (b)**

scenes with limited distances, such as indoors. The smaller the object depth, the higher the accuracy, but the larger the common overlapping region. At the same time, as the baseline $\Delta s$ increases, the depth estimation accuracy will improve. Some planar multi-aperture optical imaging systems (Tanida et al., 2000; Ng and Hanrahan, 2005; Lumsdaine and Georgiev, 2009; Venkataraman et al., 2013) have been applied to depth estimation.

Based on depth information acquisition, another basic algorithm is light field refocusing, by which the DOF can be controlled. In this way, the objects at a certain depth can be imaged, while objects at other depths are blurred, which is called synthetic aperture imaging in some research (Vaish et al., 2006; Joshi et al., 2007; Yang T et al., 2014). The minimum DOF is consistent with the depth estimation accuracy.

Taking Fig. 6 as an example, if the refocused depth is $Z$, the sub-image of aperture $n_X$ will be shifted by $n_X\Delta p$ pixels, in which $\Delta p$ is the parallax. Then, adding and averaging these shifted sub-images, the refocused image, whose viewpoint is the same with aperture $n_X = 0$, will be obtained. When $\Delta s = 0.05$ m and $Z = 10$ m, the DOF of the refocused image is 2 m according to Eq. (32).

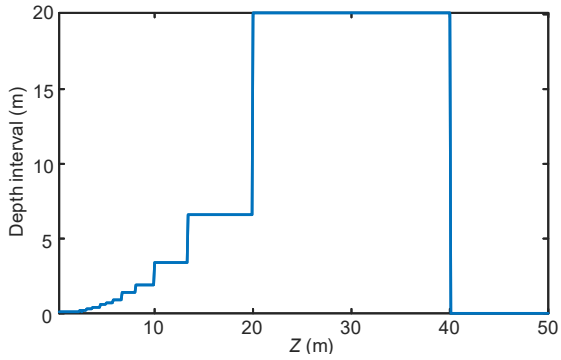


**Fig. 19 The depth interval for Fig. 18b**

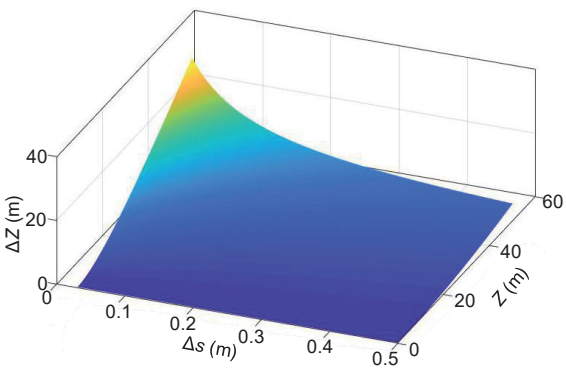


**Fig. 20 The depth estimation accuracy**

In short, as the parallax $\Delta p$ increases, the refocusing effect will be more obvious. Yang JC et al. (2002), Ng and Hanrahan (2005), Wilburn et al. (2005), Levoy et al. (2006), and Lumsdaine and Georgiev (2009) studied light field refocusing based on their multi-aperture optical imaging systems.

In summary, the planar structure is suitable for light field imaging such as superresolution reconstruction, depth estimation, and refocusing. However, application for depth estimation or refocusing is possible only when the object depth is not too large.

### 6.2 Application analysis of the convex structure

It is apparent that convex multi-aperture optical imaging systems have the advantage of a larger whole FOV compared to the planar structure. Most of the convex multi-aperture optical imaging systems focus on wide FOV imaging. The basis of wide FOV imaging is image stitching. In general, the image stitching algorithm is composed of two steps: the first step is image registration, which is the core of image stitching, and the second step is image fusion (Zhang ZZ et al., 2018). However, traditional methods are time-consuming and thus cannot guarantee real-time wide FOV imaging. Luckily, for convex multi-aperture optical imaging systems, the relative position of each sub-image obtained is known, and this offers a solution to the problem. In this subsection, the advantage of convex structure in image stitching is analyzed.

Fig. 13 is taken as an example again. Suppose that five apertures are fixed on a spherical frame whose radius is $R$ and that the latitude interval between adjacent apertures is $\Delta\theta$. When $R = 0.1$ m and $\Delta\theta = 20°$, if the depth of the curved object surface $Z$ increases from 0.3 m to 50 m, the sub-image overlap ratio $\eta_\theta$ will approach $1-\Delta\theta/(2\phi_0) = 0.5413$ according to Eq. (21). To find the critical depth $Z_{\rm opt}$ at which $\eta_\theta$ is approximately constant, take the derivative of Eq. (21) using Eq. (33). The variation trends of $\eta_\theta$ and $\mathrm{d}\eta_\theta/\mathrm{d}Z$ are shown in Fig. 21.

$$\frac{\mathrm{d}\eta_\theta}{\mathrm{d}Z} = \frac{\Delta\theta R\sin\phi_0}{2Z\sqrt{Z^2-(R\sin\phi_0)^2}\left(\phi_0-\arcsin\frac{R\sin\phi_0}{Z}\right)^2}. \tag{33}$$

Judging from Fig. 21b, as $\mathrm{d}\eta_\theta/\mathrm{d}Z$ decreases monotonically with $Z$, when $Z \geq 9$ m ($\mathrm{d}\eta_\theta/\mathrm{d}Z \leq$

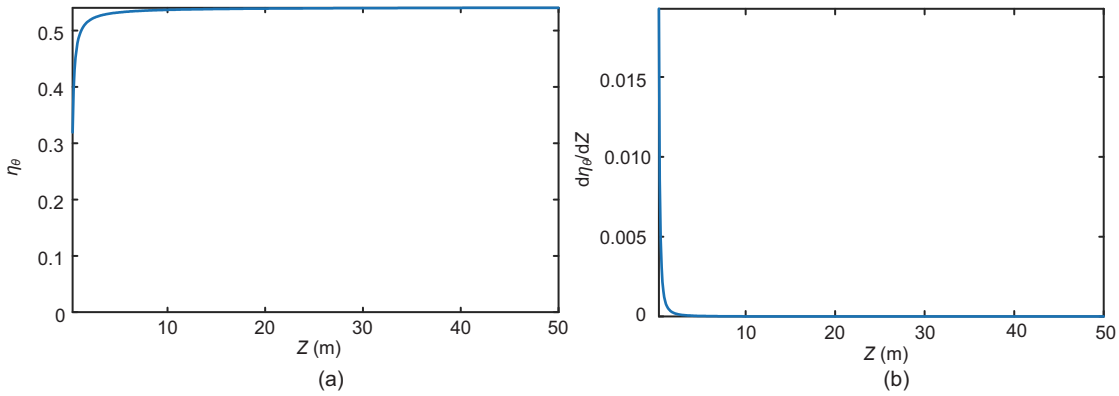


**Fig. 21 The variation trend of $\eta_\theta$ (a) and $d\eta_\theta/dZ$ (b)**

$10^{-5}$), $\eta_\theta$ is approximately constant at $1-\Delta\theta/(2\phi_0)$. The conclusion is easily drawn that to obtain a panoramic image when the depth of the curved object surface $Z$ is large enough, the sub-image overlap ratio can be regarded as $1-\Delta\theta/(2\phi_0)$, and then the non-overlapping parts of adjacent sub-images are spliced.

To find the critical depth $Z_{\rm opt}$ where direct stitching can be carried out when $Z \geq Z_{\rm opt}$, the tolerable error $\Delta e$ is expressed in pixels by Eq. (34). In this way, as the tolerable error is offered, $Z_{\rm opt}$ is computed using Eq. (35). Following the previous analysis, the variation trends of $Z_{\rm opt}$ with $\Delta e$ are shown in Fig. 22.

$$\begin{aligned}\Delta e &= \left(\eta_\theta - \left(1-\frac{\Delta\theta}{2\phi_0}\right)\right)N_u \\ &= \frac{N_u\Delta\theta\arcsin\frac{R\sin\phi_0}{Z}}{2\phi_0\left(\phi_0-\arcsin\frac{R\sin\phi_0}{Z}\right)},\end{aligned} \tag{34}$$

$$Z_{\rm opt} = \frac{R\sin\phi_0}{\sin\frac{2\Delta e\phi_0^2}{N_u\Delta\theta+2\Delta e\phi_0}}. \tag{35}$$

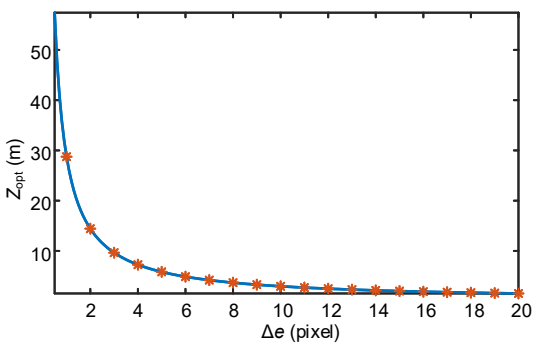


**Fig. 22 The variation trend of $Z_{\rm opt}$ with $\Delta e$**

The deeper the curved object surface is, the more accurate the direct stitching will be. If tolerable error $\Delta e$ is less than 1 pixel, then the critical depth will be computed as $Z_{\rm opt} = 28.7502$ m. Therefore, Eq. (35) gives the design solution for a convex multi-aperture optical imaging system that fits a particular curved object surface depth for direct stitching. Based on the principle of this method, Golish et al. (2012), Cao et al. (2014), and Popovic et al. (2014) performed rapid wide FOV imaging on their proposed ACEs.

Based on the wide FOV, Brady et al. (2012), Afshari et al. (2013), and Cao et al. (2015) achieved panoramic imaging. Leitel et al. (2014), Pang et al. (2017), and Wu SD et al. (2019) used the proposed ACEs to detect motion based on light flow. Some convex multi-aperture optical imaging systems (Guo et al., 2012; Shi et al., 2017; Yu et al., 2019) were applied to fast target location.

In short, the convex structure has the advantage of real-time image stitching, which is suitable for wide FOV imaging. Based on the wide FOV, convex multi-aperture optical imaging systems have great application value in surveillance and reconnaissance, image navigation, multi-target detection, and tracking.

## 7 Conclusions and future work

In this review, some typical multi-aperture optical imaging systems were enumerated and categorized. Then, the light field acquisition models were summarized according to their different structures. Based on mathematical models, the key indexes

of different multi-aperture optical imaging systems (e.g., FOV, sub-image overlap ratio, and common overlapping region) were computed easily.

In the future, multi-aperture optical imaging systems will be changed from rigid to flexible, which means zoomable aperture and a flexible substrate. For microlens arrays, researchers are constantly designing zoomable microlenses and optimizing imaging sensors from flat to curved. For lens module arrays, flexible substrates are worth studying. In addition, considering the one common imaging sensor structure, optical transferring systems need to be upgraded so that the sub-image error can be as small as possible. In short, multi-aperture optical imaging systems will be smaller and more flexible, and have higher resolution.

Concurrently, application research continues to advance. Multi-aperture optical imaging systems will be more widely used in the fields of computational photography, surveillance and reconnaissance, image navigation, 3D reconstruction, and so on.

## Contributors

Qiming QI and Hongqi FAN structured the outline of the paper. Qiming QI drafted the paper. Zhengzheng SHAO and Ping WANG helped check the first two sections. Ruigang FU helped organize the paper. Qiming QI and Hongqi FAN revised and finalized the paper.

## Compliance with ethics guidelines

Qiming QI, Ruigang FU, Zhengzheng SHAO, Ping WANG, and Hongqi FAN 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, Jacques L, Bagnato L, et al., 2013. The PANOPTIC camera: a plenoptic sensor with real-time omnidirectional capability. *J Signal Process Syst*, 70(3):305-328. https://doi.org/10.1007/s11265-012-0668-4

Aurenhammer F, 1991. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput Surv*, 23(3):345-405. https://doi.org/10.1145/116873.116880

Barsky BA, Horn DR, Klein SA, et al., 2003. Camera models and optical systems used in computer graphics: Part I, object-based techniques. Int Conf on Computational Science and Its Applications, p.246-255. https://doi.org/10.1007/3-540-44842-X_26

Bishop TE, Zanetti S, Favaro P, 2009. Light field super-resolution. IEEE Int Conf on Computational Photography, p.1-9. https://doi.org/10.1109/ICCPHOT.2009.5559010

Bolles RC, Baker HH, Marimont DH, 1987. Epipolar-plane image analysis: an approach to determining structure from motion. *Int J Comput Vis*, 1(1):7-55. https://doi.org/10.1007/BF00128525

Born M, Wolf E, Bhatia A, et al., 1999. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9781139644181

Brady DJ, Gehm ME, Stack RA, et al., 2012. Multiscale gigapixel photography. *Nature*, 486(7403):386-389. https://doi.org/10.1038/nature11150

Cao AX, Shi LF, Shi RY, et al., 2014. Image processing algorithm study of large FOV compound eye structure. *Acta Photon Sin*, 43(5):510005 (in Chinese). https://doi.org/10.3788/gzxb20144305.0510005

Cao AX, Shi LF, Deng QL, et al., 2015. Structural design and image processing of a spherical artificial compound eye. *Optik*, 126(21):3099-3103. https://doi.org/10.1016/j.ijleo.2015.07.094

Cao AX, Wang JZ, Pang H, et al., 2018. Design and fabrication of a multifocal bionic compound eye for imaging. *Bioinspir Biomim*, 13(2):026012. https://doi.org/10.1088/1748-3190/aaa901

Carles G, Downing J, Harvey AR, 2014. Super-resolution imaging using a camera array. *Opt Lett*, 39(7):1889-1892. https://doi.org/10.1364/OL.39.001889

Chai JX, Tong X, Chan SC, et al., 2000. Plenoptic sampling. Proc 27th Annual Conf on Computer Graphics and Interactive Techniques, p.307-318. https://doi.org/10.1145/344779.344932

Cheng Y, Cao J, Zhang YK, et al., 2019. Review of state-of-the-art artificial compound eye imaging systems. *Bioinspir Biomim*, 14(3):031002. https://doi.org/10.1088/1748-3190/aaffb5

de Berg M, van Kreveld M, Overmars M, et al., 2001. Computational Geometry: Algorithms and Applications (2nd Ed.). Springer-Verlag.

Deng ZF, Chen F, Yang Q, et al., 2016. Dragonfly-eye-inspired artificial compound eyes with sophisticated imaging. *Adv Funct Mater*, 26(12):1995-2001. https://doi.org/10.1002/adfm.201504941

Duparré J, Dannberg P, Schreiber P, et al., 2005. Thin compound-eye camera. *Appl Opt*, 44(15):2949-2956. https://doi.org/10.1364/Ao.44.002949

Duparré J, Radtke D, Tünnermann A, 2007. Spherical artificial compound eye captures real images. Proc SPIE 6466, MOEMS and Miniaturized Systems VI, Article 64660K. https://doi.org/10.1117/12.696258

Durand F, Holzschuch N, Soler C, et al., 2005. A frequency analysis of light transport. *ACM Trans Graph*, 24(3):1115-1126. https://doi.org/10.1145/1073204.1073320

Georgiev T, Chunev G, Lumsdaine A, 2011. Superresolution with the focused plenoptic camera. Proc SPIE, Computational Imaging IX, Article 7873. https://doi.org/10.1117/12.872666

Geyer C, Daniilidis K, 2000. A unifying theory for central panoramic systems and practical implications. Proc 6th European Conf on Computer Vision, p.445-461. https://doi.org/10.1007/3-540-45053-X_29

Golish DR, Vera EM, Kelly KJ, et al., 2012. Development of a scalable image formation pipeline for multiscale gigapixel photography. *Opt Expr*, 20(20):22048-22062. https://doi.org/10.1364/OE.20.022048

Gong XW, Yu WX, Zhang HX, et al., 2013. Progress in design and fabrication of artificial compound eye optical systems. *Chin J Opt*, 6(1):34-45 (in Chinese). https://doi.org/10.3788/CO.20130601.0034

Gortler SJ, Grzeszczuk R, Szeliski R, et al., 1996. The lumigraph. Proc 23rd Annual Conf on Computer Graphics and Interactive Techniques, p.43-54. https://doi.org/10.1145/237170.237200

Guo F, Zheng YP, Wang KY, 2012. Lenses matching of compound eye for target positioning. Proc SPIE 8420, 6th Int Symp on Advanced Optical Manufacturing and Testing Technologies: Optical System Technologies for Manufacturing and Testing, Article 84200B. https://doi.org/10.1117/12.979925

Hahne C, Aggoun A, Haxha S, et al., 2014a. Baseline of virtual cameras acquired by a standard plenoptic camera setup. 3DTV-Conf: the True Vision-Capture, Transmission and Display of 3D Video, p.1-3. https://doi.org/10.1109/3DTV.2014.6874734

Hahne C, Aggoun A, Haxha S, et al., 2014b. Light field geometry of a standard plenoptic camera. *Opt Expr*, 22(22):26659-26673. https://doi.org/10.1364/OE.22.026659

Hao YP, Li L, 2015. New progress in structure design and imaging systems of artificial compound eye. *Laser Infr*, 45(12):1407-1412 (in Chinese). https://doi.org/10.3969/j.issn.1001-5078.2015.12.001

Hartley R, Zisserman A, 2004. Multiple View Geometry in Computer Vision (2nd Ed.). Cambridge University Press, UK. https://doi.org/10.1017/CBO9780511811685

Heikkila J, Silven O, 1997. A four-step camera calibration procedure with implicit image correction. Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.1106-1112.

Isaksen A, McMillan L, Gortler SJ, 2000. Dynamically reparameterized light fields. Proc 27th Annual Conf on Computer Graphics and Interactive Techniques, p.297-306. https://doi.org/10.1145/344779.344929

Johannsen O, Sulc A, Goldluecke B, 2016. What sparse light field coding reveals about scene structure. IEEE Conf on Computer Vision and Pattern Recognition, p.3262-3270. https://doi.org/10.1109/CVPR.2016.355

Joshi N, Avidan S, Matusik W, et al., 2007. Synthetic aperture tracking: tracking through occlusions. IEEE 11th Int Conf on Computer Vision, p.1-8. https://doi.org/10.1109/ICCV.2007.4409032

Kannala J, Brandt SS, 2006. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Trans Patt Anal Mach Intell*, 28(8):1335-1340. https://doi.org/10.1109/tpami.2006.153

Kim C, Zimmer H, Pritch Y, et al., 2013. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans Graph*, 32(4):1-12. https://doi.org/10.1145/2461912.2461926

Land MF, 1989. Variations in the structure and design of compound eyes. In: Stavenga DG, Hardie RC (Eds.), Facets of Vision. Springer, Berlin, p.90-111. https://doi.org/10.1007/978-3-642-74082-4_5

Leitel R, Brückner A, Buß W, et al., 2014. Curved artificial compound-eyes for autonomous navigation. Proc SPIE 9130, Micro-Optics, Article 91300H. https://doi.org/10.1117/12.2052710

Levoy M, 2006. Light fields and computational imaging. *Computer*, 39(8):46-55. https://doi.org/10.1109/mc.2006.270

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. *ACM Trans Graph*, 25(3):924-934. https://doi.org/10.1145/1141911.1141976

Li L, Yi AY, 2012. Design and fabrication of a freeform microlens array for a compact large-field-of-view compound-eye camera. *Appl Opt*, 51(12):1843-1852. https://doi.org/10.1364/AO.51.001843

Liang CK, Shih YC, Chen HH, 2011. Light field analysis for modeling image formation. *IEEE Trans Image Process*, 20(2):446-460. https://doi.org/10.1109/TIP.2010.2063036

Lim JG, Ok HW, Park BK, et al., 2009. Improving the spatial resolution based on 4D light field data. Proc 16th IEEE Int Conf on Image Processing, p.1169-1172. https://doi.org/10.1109/ICIP.2009.5413719

Lin HT, Chen C, Kang SB, et al., 2015. Depth recovery from light field using focal stack symmetry. IEEE Int Conf on Computer Vision, p.3451-3459. https://doi.org/10.1109/ICCV.2015.394

Lin ZC, Shum HY, 2004. A geometric analysis of light field rendering. *Int J Comput Vis*, 58(2):121-138. https://doi.org/10.1023/B:VISI.0000015916.91741.27

Lindlein N, Leuchs G, 2012. Geometrical optics. In: Träger F (Ed.), Springer Handbook of Lasers and Optics. Springer, Berlin, p.35-87.

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

Luo JS, Guo YC, Wang X, 2015. Development of a multi-focusing artificial compound eye with decreasing focal length. *Acta Photon Sin*, 44(10):1022002 (in Chinese). https://doi.org/10.3788/gzxb20154410.1022002

McMillan L, Bishop G, 1995. Plenoptic modeling: an image-based rendering system. Proc 22nd Annual Conf on Computer Graphics and Interactive Techniques, p.39-46. https://doi.org/10.1145/218380.218398

Moon P, Spencer DE, 1953. Theory of the photic field. *J Franklin Inst*, 255(1):33-50. https://doi.org/10.1016/0016-0032(53)90727-3

Ng R, 2005. Fourier slice photography. *ACM Trans Graph*, 24(3):735-744. https://doi.org/10.1145/1073204.1073256

Ng R, Hanrahan P, 2005. Light field photography with a hand-held plenoptic camera. Proc SPIE 6342, Int Optical Design Conf, Article 63421E. https://doi.org/10.1117/12.692290

O'Shea DC, Zajac A, 1986. Elements of modern optical design. *Phys Today*, 39(5):87-88. https://doi.org/10.1063/1.2815008

Pang K, Fang FZ, Song L, et al., 2017. Bionic compound eye for 3D motion detection using an optical freeform surface. *J Opt Soc Am B*, 34(5):B28-B35. https://doi.org/10.1364/JOSAB.34.000B28

Popovic V, Seyid K, Akin A, et al., 2014. Image blending in a high frame rate FPGA-based multi-camera system. *J Signal Process Syst*, 76(2):169-184. https://doi.org/10.1007/s11265-013-0858-8

Shi CY, Wang YY, Liu CY, et al., 2017. SCECam: a spherical compound eye camera for fast location and recognition of objects at a large field of view. *Opt Expr*, 25(26):32333-32345. https://doi.org/10.1364/OE.25.032333

Soler C, Subr K, Durand F, et al., 2009. Fourier depth of field. *ACM Trans Graph*, 28(2):1-12. https://doi.org/10.1145/1516522.1516529

Song YM, Xie YZ, Malyarchuk V, et al., 2013. Digital cameras with designs inspired by the arthropod eye. *Nature*, 497(7447):95-99. https://doi.org/10.1038/nature12083

Suo JL, Ji XY, Dai QH, 2012. An overview of computational photography. *Sci China Inform Sci*, 55(6):1229-1248. https://doi.org/10.1007/s11432-012-4587-6

Tanida J, Kumagai T, Yamada K, et al., 2000. Thin Observation Module by Bound Optics (TOMBO): an optoelectronic image capturing system. Proc SPIE 4089, Optics in Computing, Article 2000. https://doi.org/10.1117/12.386797

Tardif JP, Sturm P, Roy S, 2006. Self-calibration of a general radially symmetric distortion model. Proc 9th European Conf on Computer Vision, p.186-199. https://doi.org/10.1007/11744085_15

Vaish V, Wilburn B, Joshi N, et al., 2004. Using plane + parallax for calibrating dense camera arrays. Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.2-9. https://doi.org/10.1109/CVPR.2004.257

Vaish V, Levoy M, Szeliski R, et al., 2006. Reconstructing occluded surfaces using synthetic apertures: stereo, focus and robust measures. IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.2331-2338. https://doi.org/10.1109/CVPR.2006.244

Venkataraman K, Lelescu D, Duparré J, et al., 2013. Picam: an ultra-thin high performance monolithic camera array. *ACM Trans Graph*, 32(6):1-13. https://doi.org/10.1145/2508363.2508390

Wang TC, Chandraker M, Efros AA, et al., 2018. SVBRDF-invariant shape and reflectance estimation from a light-field camera. *IEEE Trans Patt Anal Mach Intell*, 40(3):740-754. https://doi.org/10.1109/TPAMI.2017.2680442

Wang YQ, Yang JG, Guo YL, et al., 2019. Selective light field refocusing for camera arrays using bokeh rendering and superresolution. *IEEE Signal Process Lett*, 26(1):204-208. https://doi.org/10.1109/LSP.2018.2885213

Wang YW, Cai BL, Lu Y, et al., 2017. Optical system design of artificial compound eye based on field stitching. *Microw Opt Technol Lett*, 59(6):1277-1279. https://doi.org/10.1002/mop.30525

Wanner S, Goldluecke B, 2012a. Globally consistent depth labeling of 4D light fields. IEEE Conf on Computer Vision and Pattern Recognition, p.41-48. https://doi.org/10.1109/CVPR.2012.6247656

Wanner S, Goldluecke B, 2012b. Spatial and angular variational super-resolution of 4D light fields. Proc 12th European Conf on Computer Vision, p.608-621. https://doi.org/10.1007/978-3-642-33715-4_44

Wanner S, Goldluecke B, 2014. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans Patt Anal Mach Intell*, 36(3):606-619. https://doi.org/10.1109/TPAMI.2013.147

Wen C, Ma T, Wang C, et al., 2019. Progress in research on the compound eye structure and visual navigation of insects. *Chin J Appl Entomol*, 56(1):28-36 (in Chinese). https://doi.org/10.7679/j.issn.2095-1353.2019.004

Weng J, Cohen P, Herniou M, 1992. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans Patt Anal Mach Intell*, 14(10):965-980. https://doi.org/10.1109/34.159901

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

Williem W, Park IK, 2016. Robust light field depth estimation for noisy scene with occlusion. IEEE Conf on Computer Vision and Pattern Recognition, p.4396-4404. https://doi.org/10.1109/CVPR.2016.476

Wu G, 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

Wu SD, Jiang T, Zhang GX, et al., 2017. Artificial compound eye: a survey of the state-of-the-art. *Artif Intell Rev*, 48(4):573-603. https://doi.org/10.1007/s10462-016-9513-7

Wu SD, Zhang GX, Zhu M, et al., 2018a. Geometry based three-dimensional image processing method for electronic cluster eye. *Integr Comput-Aided Eng*, 25(3):213-228. https://doi.org/10.3233/ICA-180564

Wu SD, Zhang GX, Jiang T, et al., 2018b. Multi-aperture stereo reconstruction for artificial compound eye with cross image belief propagation. *Appl Opt*, 57(7):B160-B169. https://doi.org/10.1364/AO.57.00B160

Wu SD, Zhang GX, Neri F, et al., 2019. A multi-aperture optical flow estimation method for an artificial compound eye. *Integr Comput-Aided Eng*, 26(2):139-157. https://doi.org/10.3233/ICA-180593

Yang JC, Everett M, Buehler C, et al., 2002. A real-time distributed light field camera. Proc 13th Eurographics Workshop on Rendering, p.77-86.

Yang T, Zhang YN, Yu JY, et al., 2014. All-in-focus synthetic aperture imaging. Proc 13th European Conf on Computer Vision, p.1-15.

Yu XD, Zhang YJ, Wang YY, et al., 2019. Optical design of a compound eye camera with a large-field of view for unmanned aerial vehicles. *Acta Photon Sin*, 48(7):0722003 (in Chinese). https://doi.org/10.3788/gzxb20194807.0722003

Yuan W, Li LH, Lee WB, et al., 2018. Fabrication of microlens array and its application: a review. *Chin J Mech Eng*, 31(1):16. https://doi.org/10.1186/s10033-018-0204-y

Zhang C, Chen T, 2003. Spectral analysis for sampling image-based rendering data. *IEEE Trans Circ Syst Video Technol*, 13(11):1038-1050. https://doi.org/10.1109/tcsvt.2003.817350

Zhang JM, Chen Y, Tan HQ, et al., 2020. Optical system of bionic compound eye with large field of view. *Opt Prec Eng*, 28(5):1012-1020 (in Chinese). https://doi.org/10.3788/OPE.20202805.1012

Zhang YK, Du JL, Shi LF, et al., 2010. Artificial compound-eye imaging system with a large field of view based on a convex solid substrate. Proc SPIE 7848, Holography, Diffractive Optics, and Applications IV, Article 78480U. https://doi.org/10.1117/12.869482

Zhang ZY, 2000. A flexible new technique for camera calibration. *IEEE Trans Patt Anal Mach Intell*, 22(11):1330-1334. https://doi.org/10.1109/34.888718

Zhang ZZ, Qiu S, Jin WQ, et al., 2018. Image mosaic of bionic compound eye imaging system based on image overlap rate prior. Proc SPIE 10846, Optical Sensing and Imaging Technologies and Applications, Article 108462C. https://doi.org/10.1117/12.2505433

Zhou PL, Yu HB, Zhong Y, et al., 2020. Fabrication of waterproof artificial compound eyes with variable field of view based on the bioinspiration from natural hierarchical micro-nanostructures. *Nano-Micro Lett*, 12(1):166. https://doi.org/10.1007/s40820-020-00499-x

Zhu H, Wang Q, Yu JY, 2017. Light field imaging: models, calibrations, reconstructions, and applications. *Front Inform Technol Electron Eng*, 18(9):1236-1249. https://doi.org/10.1631/FITEE.1601727

Zhu L, Zhang YL, Sun HB, 2019. Miniaturising artificial compound eyes based on advanced micronanofabrication techniques. *Light Adv Manuf*, 2(1):84-100. https://doi.org/10.37188/lam.2021.007

Qiming QI was born in 1994. He received his MS degree in information and communication engineering from National University of Defense Technology, Changsha, China in 2021. His research interests are bionic vision application and optical automatic target recognition.

Hongqi FAN, corresponding author of this invited paper, received his BS degree in mechanical engineering and automation from Tsinghua University, Beijing, China in 2001, and his PhD degree in information and communication engineering from National University of Defense Technology, Changsha, China in 2008. He is currently a professor at National University of Defense Technology. His research interests include information fusion, target tracking, signal processing, and intelligent guidance systems.

## Appendix A: Performance calculation for planar ACE

Suppose that the object plane is $Z$ away from the optical center plane. For the single aperture whose label is $n_X$, every pixel coordinate $u$ corresponds to a point on the object plane. $x$ is used to denote the position of the object point:

$$\begin{aligned} x &= n_X \Delta s + (u - c_u)\delta_u \frac{Z}{f} \\ &= n_X \Delta s + (u - c_u)\frac{Z}{f_u}. \end{aligned} \tag{A1}$$

Therefore, the size of the sub-image object region is $(N_u - 1)Z/f_u$. For two adjacent apertures, if their sub-images overlap, the size of a single aperture's object region is larger than its interval $\Delta s$. In this case, only when object depth $Z$ is larger than the threshold, as the following equation shows, will adjacent sub-images overlap:

$$N_u \delta_u \frac{Z}{f} \geq \Delta s \Rightarrow Z \geq \frac{f_u \Delta s}{N_u}. \tag{A2}$$

Further, the sub-image overlap ratio can be computed as follows:

$$\eta_X = 1 - \frac{\Delta s}{N_u \delta_u \frac{Z}{f}} = 1 - \frac{f_u \Delta s}{N_u Z}. \tag{A3}$$

When $n_X$ and $u$ are taken to be the maximum and minimum, the edge coordinates of the whole object region are

$$\begin{cases} x_{\min} = n_{X(\min)}\Delta s + \dfrac{Z}{f_u}(1 - c_u), \\ x_{\max} = n_{X(\max)}\Delta s + \dfrac{Z}{f_u}(N_u - c_u). \end{cases} \tag{A4}$$

Note that the pixel resolution is $\Delta x = Z/f_u$, while the subpixel shift is not considered. With $\lfloor \cdot \rfloor$ denoting rounding down, the number of pixels representing the whole object region is

$$N_x = \frac{x_{\max} - x_{\min}}{\Delta x} = \left\lfloor (N_X - 1)\Delta s \frac{f_u}{Z} \right\rfloor + N_u - 1. \tag{A5}$$

## Appendix B: Performance calculation for spherical ACE

As shown in Fig. 13, five apertures are arranged on the spherical surface whose radius is $R$, and the angle between adjacent apertures is denoted as $\Delta\theta$.

If the adjacent sub-images overlap, $\Delta\theta < 2\phi_0$ should be satisfied. In this way, according to the sine theorem, the minimum scene depth $Z_{\min}$ is computed as shown below:

$$\frac{\sin(180^\circ-\phi_0)}{Z_{\min}/\cos\frac{\Delta\theta}{2}}=\frac{\sin\left(\phi_0-\frac{\Delta\theta}{2}\right)}{R}$$
$$\Rightarrow Z_{\min}=R\frac{\sin\phi_0\cos\frac{\Delta\theta}{2}}{\sin\left(\phi_0-\frac{\Delta\theta}{2}\right)}. \tag{B1}$$

The angle of the overlapping region corresponding to $O_G$, denoted as $\alpha$, can be obtained:

$$\frac{\sin(180^\circ-\phi_0)}{Z}=\frac{\sin\left(\phi_0-\frac{\Delta\theta+\alpha}{2}\right)}{R}$$
$$\Rightarrow \alpha=2\phi_0-2\arcsin\frac{R\sin\phi_0}{Z}-\Delta\theta. \tag{B2}$$

Thus, the sub-image overlap ratio can be computed as follows:

$$\eta_\theta=\frac{\alpha}{2\frac{\Delta\theta+\alpha}{2}}=1-\frac{\Delta\theta}{2\phi_0-2\arcsin\frac{R\sin\phi_0}{Z}}. \tag{B3}$$

Then, for the loop $0<n_\theta<N_\theta-1$, the projection radius of an object plane in the non-overlapping part is

$$\begin{aligned}R'&=Z\sin\left(\frac{\Delta\theta-\alpha}{2}+\theta_{\mathrm{G}}\right)\\&=Z\sin\left(\Delta\theta-\phi_0+\arcsin\frac{R\sin\phi_0}{Z}+\theta_{\mathrm{G}}\right).\end{aligned} \tag{B4}$$

Therefore, the number of apertures located at loop $n_\theta$ should satisfy

$$\begin{aligned}N_{\phi(n_\theta)}&\geq\left\lceil\frac{360^\circ}{2\arctan\frac{(Z-R)\tan\theta_0}{R'}}\right\rceil\\&=\left\lceil\frac{180^\circ}{\arctan\frac{(Z-R)\tan\theta_0}{Z\sin\left(\theta_{\mathrm{G}}+\Delta\theta-\phi_0+\arcsin\frac{R\sin\phi_0}{Z}\right)}}\right\rceil.\end{aligned} \tag{B5}$$

As for the loop $n_\theta=N_\theta-1$, $N_{\phi(n_\theta)}$ should satisfy

$$N_{\phi(N_\theta-1)}\geq\left\lceil\frac{180^\circ}{\arctan\frac{\tan\theta_0\sin\varphi}{\tan\phi_0\sin(\varphi+\theta_{\mathrm{G}})}}\right\rceil. \tag{B6}$$

Here, $\varphi=\arctan\frac{(Z-R)\tan\phi_0}{Z}$.