IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON COLLABORATION FOR INTERNET OF THINGS

Received August 10, 2018, accepted September 4, 2018, date of publication September 17, 2018, date of current version October 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870394

Research on Depth Estimation Method of Light Field Imaging Based on Big Data in Internet of Things From Camera Array

YUE WU
Changsha High-Tech Industrial Development Zone, Changsha 410205, China
National University of Defense Technology, Changsha 410073, China
e-mail: wuyuework@163.com

ABSTRACT In recent years, optical field imaging technology has received extensive attention in the academic circle for its novel imaging characteristics of shooting first and focusing later, variable depth of field, variable viewpoint, and so on. However, the existing optical field acquisition equipment can only acquire a limited number of discrete angle signals, so image aliasing caused by under sampling of optical field angle signals reduces the quality of optical field images. Based on the camera array system as a platform, this paper studies the optical field imaging and depth estimation method based on the Big Data in Internet of Things obtained from camera array around the angle sampling characteristics of the optical field data set, and has achieved some innovative research results in the following aspects. On the basis of analyzing the characteristics of different depth clues in the optical field data set, a depth estimation method combining parallax method and focusing method is proposed. First, this paper analyzes the disparity clues and focus clues contained in the multi-view data set and the light field refocusing image set of the camera array, respectively, and points out the differences and relationships between the two depth clues extraction methods in the light field sampling frequency domain space, that is, the disparity method focuses on the energy concentration characteristics near the frequency domain spatial angle axis, while the focus method focuses on the high frequency proportion of energy distribution on the angle axis. Then, the weighted linear fusion method based on image gradient is used to fuse the two calculation results, which improves the accuracy and robustness of depth estimation. Finally, the results of depth estimation experiments on different sets of scenes show that compared with the method based on a single depth cue, the method in this paper has higher accuracy in depth calculation in discontinuous areas of scene depth and similar texture areas.

INDEX TERMS Big data in Internet of Things(IoT), camera array, light field imaging, image aliasing, depth estimation.

I. INTRODUCTION

Research shows that more than 80 % of human perception of the real world comes from visual information. Based on the acquisition of visual information, the human brain can quickly establish an understanding of the characteristics of objects in the scene, such as shape, distance, posture, color, etc [1]. This rapid perception ability of three-dimensional scenes based on visual information enables human beings to efficiently observe, recognize and summarize the changes and laws of nature. The acquisition of visual information and the perception of the real three-dimensional world through visual information almost run through all historical stages of human development [2]. Therefore, it is of great significance to study the acquisition of visual information and the three-dimensional perception technology.

Since the French Joseph Nicephore Niepce obtained the world's first photo on photographic materials in 1827, photography has become an important way for human beings to obtain visual information [3]. The word photography comes from Greek and means "drawing by light". Therefore, how to record light information in space is one of the fundamental problems in photography. The traditional photography method uses sensors or films to record the light intensity distribution of an imaging plane in space in order to obtain a visual image, and its essence is a two-dimensional projection process of a target light set in space.

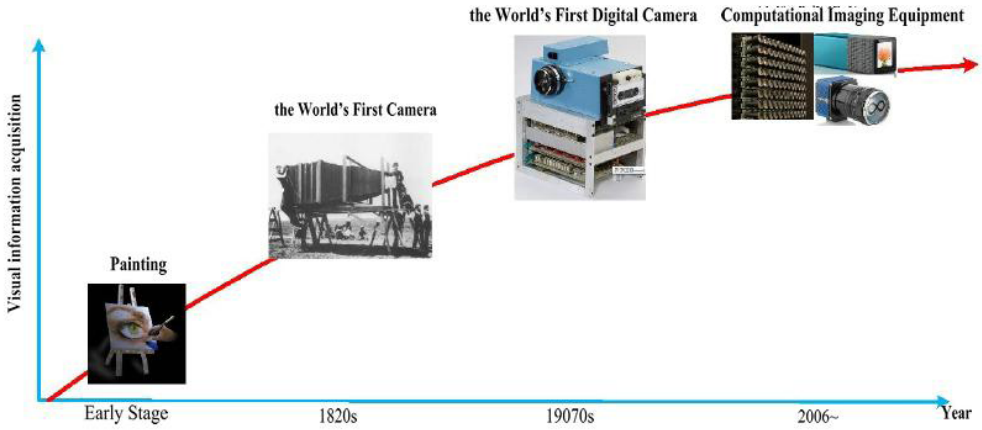


FIGURE 1. Development roadmap of visual information acquisition.

Projecting high-dimensional light signals is a typical dimensionality reduction sampling process. Therefore, there is a serious loss of information in the traditional photography process [4]. In addition, due to the limitation of processing materials and processes of optical imaging devices, the traditional imaging system has certain limitations. On the basis of inheriting the traditional photography theories and methods, the emerging computational photography [5] has further researched and innovated imaging theories, improved imaging equipment, and overcome various limitations of traditional imaging. Computational photography involves computer vision, signal processing, modern optics and many other fields, which will lead to another technological revolution in the field of digital imaging technology as shown in Figure 1.

Light field imaging is one of the most representative theories in computational photography [6]. Different from the traditional imaging device, the optical field imaging device can not only record the light intensity distribution of the imaging plane, but also can record the angle signals of light rays, that is, the optical field imaging device can record the position and angle signals of light rays in space at the same time. Based on the acquisition of optical field data, researchers design corresponding calculation methods to realize optical field imaging according to specific imaging requirements. This calculation imaging process makes the light field imaging effect no longer determined solely by the imaging parameters at the time of shooting, but can be flexibly changed by later calculation. Obviously, the dependence of its imaging process on physical devices can be further reduced, which provides a new idea for designing a new imaging system [7]. Based on the theory of light field imaging, several research institutions at home and abroad have designed different light field imaging systems, such as camera array, microlens array light field camera, coded aperture light field camera, etc.

The recording of light angle signals is the most essential characteristic of optical field devices [8]. The acquisition of ray angle signals is closely related to the calculation of scene depth and improvement of imaging performance. The introduction of the angle signal makes the optical field imaging method realize the effects that are difficult to obtain in traditional imaging, such as digital refocusing view-point switching, extension of depth of field all-in-focus, etc. In addition, due to the high degree of coupling between angle signals and scene depth information, scene depth estimation research based on light field data has also received extensive attention from many scientific research institutions at home and abroad. However, existing optical field devices can only acquire discrete angle signals, and there is a compromise between the resolution of optical field angle and position signal acquisition, which leads to the problem of angle signal undersampling and affects the quality of optical field imaging [8]. In addition, there are occlusion and illumination differences in the natural scene and noise in the signal acquisition process of the sensor, which makes the scene depth estimation based on light field data have some problems such as low reconstruction accuracy and non-robust matching algorithm, and finally restricts the application and development of light field imaging theory.

To sum up, light field imaging theory has become an important theoretical growth point and technical breakthrough point in the field of digital imaging and computer vision, and has expanded the development direction of the next generation imaging technology. The research on scene depth estimation method based on light field angle signals will provide a new solution to the core problem of computer vision, and further promote the integration of computational photography and computer vision [9]. At present, the quality of light field imaging and the accuracy of scene depth calculation in the above research hotspots are not ideal, which restricts the development of existing theories and applications.

For a long time, image-based depth estimation has been one of the focuses in the field of computer vision [10]. Aiming at different depth clues in human visual system (such as parallax, blur, shadow, texture, perspective, etc.), researchers proposed different scene depth estimation methods. However, most depth estimation methods still rely on parallax clues.

On this basis, the region-based and feature-based matching method is even more a classic method in image depth estimation. At present, combining the scene segmentation theory and the global optimization theory, the method based on disparity clues is recognized as superior to the method based on other depth clues, and the accuracy of scene depth estimation results is also higher [11]. However, the existing disparity method still has the problem that the depth estimation process is not robust. One of the reasons is the complexity of the natural scene, such as occlusion, repeated texture, depth discontinuity and so on. Another reason is that the acquisition system has insufficient viewing angle and low signal-to-noise ratio for scene sampling. For example, the traditional binocular or multi-eye system has less viewing angle and image noise and blur are common.

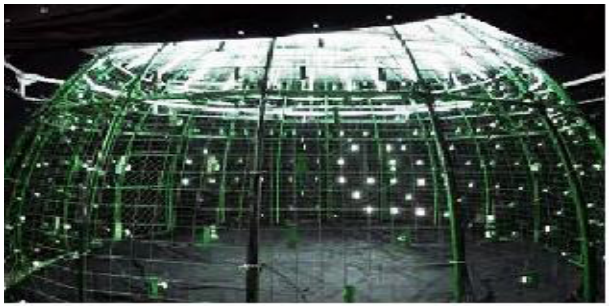
Besides, Big Data in IoT has become one of the hottest topics in recent years. There are different interpretations of ''Big Data in IoT''. The potential of Big Data in IoT is to use the large quantity of data for discovering new knowledge and unexpected relationships. Most of taxis traveling on the urban roads have been equipped with Global Positioning System (GPS) over many years of construction and operation of ITS (Intelligent Transportation Systems). These taxis yield large-scale GPS data while the collection lasts a long period of time (such as one year). It will cover the whole network and contain the traffic information of each road. Thus, it is suitable to use the data to model the relationship among different roads. What is more important is that the GPS data will be collected much easier with the wide use of GPS equipment. The source of the data is confined to taxis, while drivers or passengers with mobile phones can also provide data.

Aiming at the shortcomings of the existing optical field imaging theory and the depth clues contained in the optical field data set, this paper takes the typical wide baseline optical field camera array system based on obtained Big Data in IoT as a platform to carry out research work. The typical light field acquisition system of camera array is firstly analyzed systematically in Section 2, based on which an 8×8 camera array system architecture schematic is presented. Then, the principle of Optical Field Imaging Based on microlens array is studied in Section 3, to demonstrate its advantages compared with traditional imaging system. Finally, a novel depth estimation method combined parallax and focusing method (DEPFM) is proposed in Section 4, of which the accuracy and robustness in terms of depth estimation is demonstrated by a series of case studies in Section 5. Section 6 concludes the paper.

II. CAMERA ARRAY SYSTEM

Camera array is a typical light field acquisition system, and its main advantage lies in the large equivalent aperture of the generated image, which makes the depth of field change characteristics of camera array light field imaging more obvious, i.e. the depth resolution is higher. Compared with portable light field cameras, camera array light field imaging has many advantages such as easy change of aperture arrangement and controllable sampling density [12]. Therefore, the research work in this paper is carried out on camera arrays. Next, the camera array system built to meet the research work of this paper and the corresponding light field imaging theory are mainly introduced.

(a)



(b)

FIGURE 2. (a) Birdcage system. (b). Birdcage system.

The camera array system consists of a plurality of cameras or cameras arranged in a distributed manner, which is used for collecting scene light.

According to the different acquisition signal attributes and the different acquisition target scenes, common camera array distribution methods include spherical, circular, and planar types, as shown in Fig. 2. Figure 2(a) and (b) respectively show the ''birdcage system'' set up by the University of Southern California and Tsinghua University to collect scene illumination changes [13]. Figure 3 shows planar camera array systems built by Stanford University to obtain high-resolution light locations [14].

Since the work in this paper is based on the light field data, an 8×8 camera array system is built in this study, in which 64 IH046C CCD cameras are used, and each camera has effective resolution of 752×576 pixels-equipped with a COMPUTOR M1214 12 mm fixed focus lens. Each camera has a horizontal field of view angle of 37 and a vertical field of view angle of 30. In order to calibrate the camera array accurately, the camera angle of view needs to have a certain overlap area, so the camera array system adopts a relatively tight arrangement mode. Cameras are distributed at equal intervals in a plane manner, and the center of the cameras are horizontally spaced 7 cm apart and vertically spaced 9 cm apart, and the overall size of the array is 70 cm wide by 100 cm high. At the same time, 6 servers for synchronous

FIGURE 3. Stanford university camera array.

data collection are used, of which the architecture is shown in Figure 4.

III. RESEARCH ON OPTICAL FIELD IMAGING BASED ON MICROLENS ARRAY

The conventional camera imaging system collects light through a lens and then photoelectrically converts the light by a sensor to form an image, which is shown in Fig. 5(a). However, the optical field imaging of camera array is quite different from traditional camera imaging, in which light rays at different angles of spatial points are firstly recorded by cameras at different angles, and then the light rays are digitally focused through a light field calculation process to form an image as shown in Fig. 5(b).

In the camera array system, the spatial position (optical center position) and image plane position of each camera can be obtained by precise calibration, and the plane where the optical center of each camera is located and the plane where the sensor is located correspond to the two planes of the two-plane optical field model. As shown in Fig. 6, the UV plane represents the angular plane formed by the camera light center, and the ST plane represents the image plane formed by the sensor.

Each point (u_i, v_j) in the UV plane corresponds to a subset of the ST plane, which is formed by projecting images acquired by cameras in the (u_i, v_j) position onto the ST plane, that is, a certain angle of the target scene sampled by camera array. In theory, any ray passing through plane ST and plane UV can be made up of $r\ (u, v, s, t)$. To image a point p in front of the camera array, all rays emitted by p through plane ST and plane UV must be indexed first, which are accumulated to obtain an imaging result. Since the index and accumulation processes are both calculation processes, the light field focus is different from the refraction and convergence of the traditional lens, which can be adjusted at will after acquiring the light field data (digital refocusing) [15].

Camera array light field imaging can simulate the imaging effect of a traditional camera with a larger aperture, and therefore it is also called Synthetic Aperture Photography (SAP). Similar to the real aperture imaging system, the camera array light field imaging also has a depth of field phenomenon. That is, when the target object is located in the depth of field range of the focus plane, it can be clearly imaged, otherwise the imaging becomes blurred. The reason for the aforementioned phenomenon is that when the sensor plane deviates from the ideal image plane where the lens focuses, a diffusion region in the sensor plane will be formed by image points, and if the region is larger than the size of one sensor unit, imaging interference will occur between different image points to result in blurred images, which is shown in Figure 7. Therefore, the depth of field of imaging is related to the aperture size, image plane distance, sensor position, and sensor unit size. Due to the larger equivalent aperture, the depth of field of the camera array system becomes smaller, thus having obvious focusing and defocusing effects.

In the above-mentioned light field model, it is assumed that both the light center plane and the image plane are ideal cases. However, the light centers of different cameras in the real camera array do not meet the strict plane distribution characteristics. There are also differences in their respective image planes, so the positional relationship between cameras needs to be further considered in actual imaging process.

Before the camera array is used for light field imaging, the camera array system must be calibrated accurately. If the distortion of the lens is not considered, the imaging position of any point in space on the image can be represented by a pinhole model as shown in Fig.8. According to the relevant knowledge of camera model:

$$\begin{aligned}(x, y, 1)^T &= \begin{bmatrix} \frac{f}{f_x} & -\frac{f}{f_x \cdot \tan\theta} & x_0 \\ 0 & \frac{f}{f_y \cdot \sin\theta} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \\ &\quad \cdot (X_w, Y_w, Z_w, 1)^T \\ &= K\,(R, t) \cdot (X, 1)^T = P \cdot (X, 1)^T \end{aligned} \tag{1}$$

Where (x, y) is the coordinate of the image formed by the X point in the image coordinate system, f is the focal length of the camera, f_x and f_y represents the effective focal lengths of the x axis and the y axis respectively. (x_0, y_0) is the image main point position; K is perspective projection matrix, which is only related to the camera's internal structure, i.e. the camera's internal parameters. The position and orientation of the camera in the world coordinate system are described as $[R\ t]$, which are called camera's external parameters.

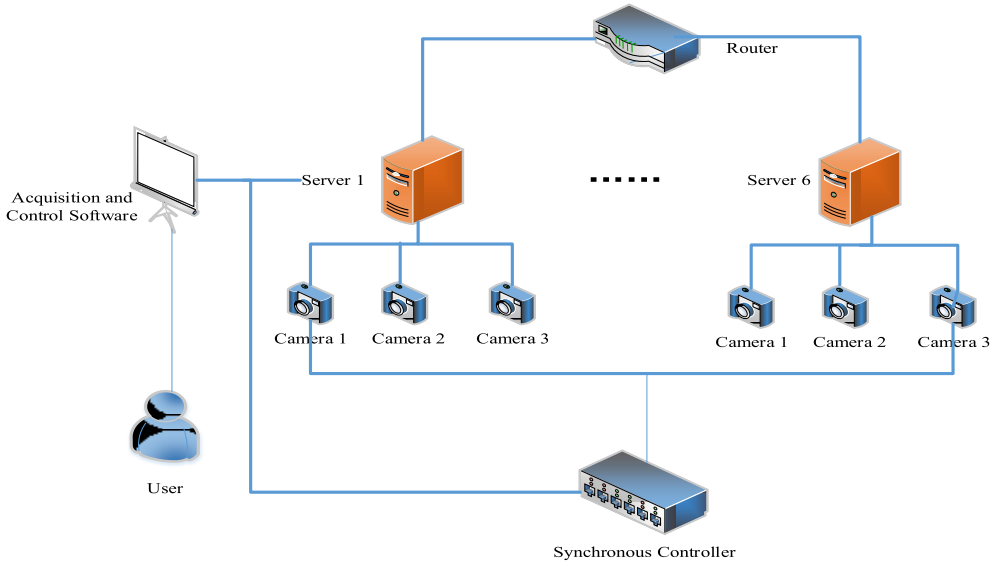


FIGURE 4. Camera array system architecture schematic.

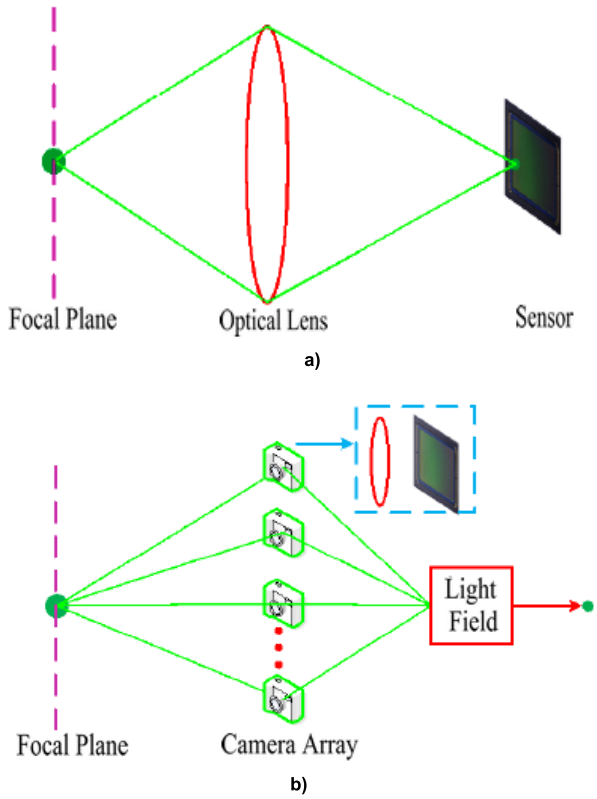


FIGURE 5. Comparison between traditional imaging and optical field imaging. a) Traditional imaging system. b) Camera array imaging system.

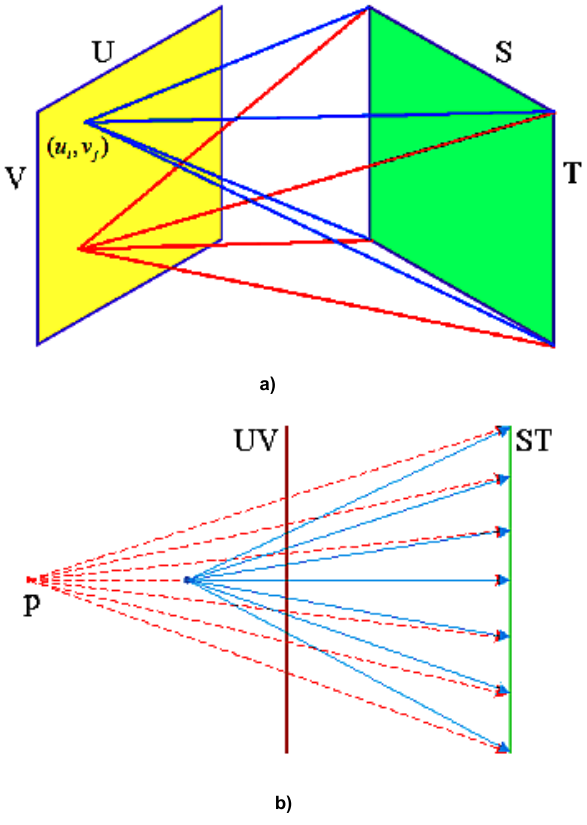


FIGURE 6. Principle of optical field biplane angle sampling and focusing. a) Optical field biplane angle sampling schematic. b) Principle of optical field biplane refocusing.

According to the theory of multi-view geometry, the process of transforming from the camera focus plane to the imaging plane is essentially a perspective transformation [16]. After accurate camera calibration, the perspective transformation corresponding to each camera in the camera array can be determined. The perspective transformation itself is a special projective transformation. Therefore, different camera image planes in the camera array can induce

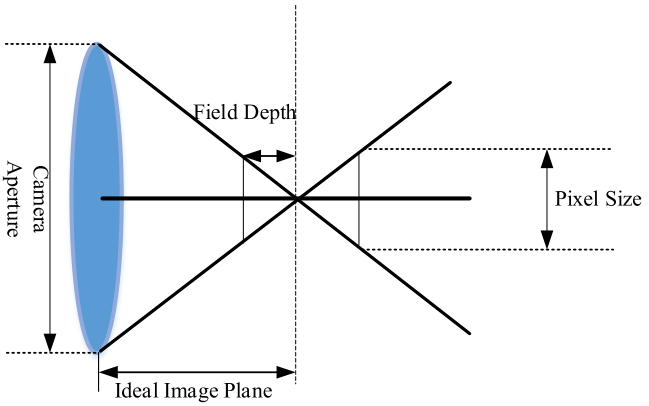


FIGURE 7. Schematic of imaging depth of field.

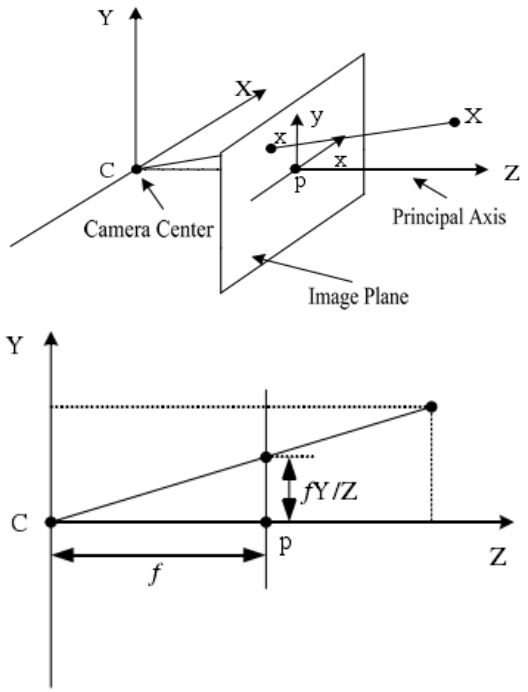


FIGURE 8. Schematic diagram of camera parametric model.

corresponding projective mapping relationship from a virtual focus plane in the scene as shown in Figure 9.

Assuming that the parameter matrix in the left camera is K, and the parameter matrix in the right camera is K'. The position of a point p in the left and right camera coordinate system is $Q = (X, Y, Z)^T$ and $Q' = \left(X', Y', Z'\right)^T$, sand the image point of point p in the left and right camera image coordinate systems is represented by $q = (x, y, 1)^T$ and $q' = \left(x', y', 1\right)^T$ respectively. According to the camera model knowledge:

$$q \cong K \cdot Q \quad q' \cong K' \cdot Q' \tag{2}$$

Assuming that the coordinate of the light center of the left camera in the right camera coordinate system is t, which is the translation vector of the left camera coordinate system relative to the right camera coordinate system. Similarly,

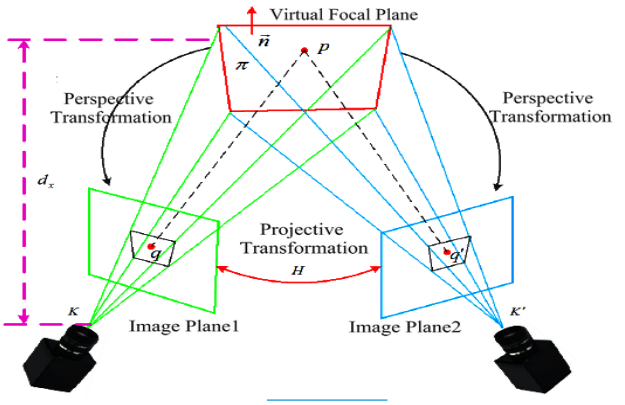


FIGURE 9. Plane transformation schematic.

assuming that the relative rotation matrix between the left camera coordinate system and the right camera coordinate system is R, and then there are:

$$Q' = R \cdot Q + t \tag{3}$$

As shown in Fig.9, the virtual focus plane is π, and $\vec{n}$ is the normal vector of the plane. Assuming that point p is on plane π, d_x is the distance from the optical center of the right camera to the plane π:

$$w = \frac{\vec{n}}{d_x}; \quad w^T \cdot Q = 1 \tag{4}$$

It can be seen from Eq. (3) and (4) that for any point p in plane π, the coordinates of the point p in two different camera coordinate systems satisfy following relationship:

$$Q' = R \cdot Q + t = R \cdot Q + t \cdot \left(w^T \cdot Q\right) = \left(R + t \cdot w^T\right) \cdot Q \tag{5}$$

According to the principle of light field refocusing, rays of different angles emitted from a certain point in space are recorded by different cameras in the camera array respectively, so the number of angle samples in light field imaging of the camera array is equal to the number of cameras. The essence of optical field focusing imaging is: according to the homography induced by different depth planes, the corresponding imaging points of cameras with different angles are digitally accumulated and re-averaged [17]. The imaging characteristic of the camera array system is that the imaging focal length can be changed digitally and flexibly. Since the depth of field range reflects the sensitivity of the imaging system to changes in focus depth, while the depth of field range is linearly related to the aperture size, so the wide baseline camera array system has higher depth resolution. However, compared with the traditional imaging system, the camera array system still has the problem of poor imaging quality. An important reason that affects the imaging quality is that the optical field signals acquired by the camera array system are discrete in the angular domain, and this kind of undersampling problem in the angular domain of the optical field will cause the aliasing phenomenon of light field images.

IV. CAMERA ARRAY MULTI-CLUE SCENE DEPTH ESTIMATION METHOD

Depth estimation is a key link in 3d reconstruction based on images and also an important basis for anti-aliasing rendering of light field images. Since there are abundant depth clues hidden in the imaging results of different focus settings of the camera array, as well as the multi-view and wide baseline characteristics of camera array system, the study of scene depth estimation based on camera array can provide new ideas for traditional image-based 3d reconstruction.

Depth estimation based on images has always been a hot topic in computer vision research. Similar to the perception of depth by the human visual system, image-based depth estimation can utilize a variety of depth clues such as parallax, perspective, blur, occlusion, motion information, etc. Depending on the number of cameras, the existing methods can be divided into three types: monocular method, binocular method and multi-eye method [18]. The classic representative of monocular method is zoom method, where the depth calculation is based on image defocus caused by camera aperture, and the core of the method is the focus (or defocus) measurement problem. The classic representation of binocular and multi-eye methods is parallax method, which is based on the parallax caused by the change of image viewpoint, of which core is the matching problem.

The camera array data set is similar to the traditional multi-view data set, which is a two-dimensional image set of the target scene recorded by multiple cameras at different viewing angles. The camera array is characterized that cameras are generally distributed on the same plane and are arranged densely. The differences of viewing angles between images captured by different cameras are small, and the image scale transformation is not large. Therefore, the target scene information recorded by the multi-view data of the camera array is more complete, and the scene depth clues contained therein are also more abundant. As mentioned in the previous section, the imaging effect of a virtual large aperture camera can be obtained by using the camera array, while the refocusing in light field imaging is a digital calculation process, and the imaging point is clearly coupled with the change of blur and the depth of the target scene. In addition, the disparity depth clues between cameras with different viewing angles in the camera array data set make it possible to estimate the scene depth using the camera array data set without relying on a single depth clue. If the camera array can be made full use of to obtain different depth clues in the information, the accuracy and robustness of scene depth estimation can be improved.

A. PARALLAX DEPTH CUE

There is a visual angle difference between different cameras in the camera array, and the distance between the image point coordinates of the same spatial point and different cameras is parallax. As shown in the Figure.10, the left camera and the right camera are respectively in three-dimensional space (0, 0, 0) and (B, 0, 0), and the coordinates of the same spatial

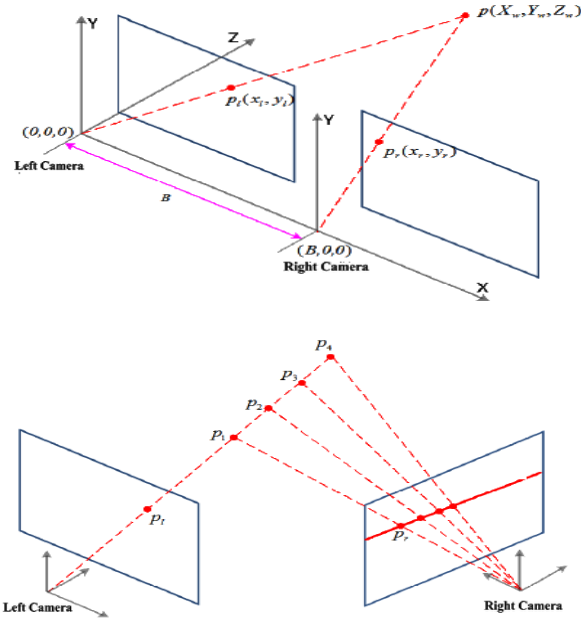


FIGURE 10. Depth calculation principle based on parallax.

point $p\ (X_w, Y_w, Z_w)$ on the images of the left camera and the right camera are $p_l\ (x_l, y_l)$ and $p_r\ (x_r, y_r)$ respectively.

According to the theory of the camera model in the chapter, when the focal length f of the camera and the depth measurement baseline B formed by the left and right cameras are known, it is easy to deduce that the horizontal parallax is:

$$d = x_l - x_r = f\frac{X_w}{Z_w} = f\frac{X_w - B}{Z_w} = f\frac{B}{Z_w} \tag{6}$$

On the other hand, if the horizontal parallax d is known, the depth information of the target point can also be inverted by the Eq.(6):

$$Z_w = f\frac{B}{d} \tag{7}$$

It can be seen that parallax d is inversely proportional to depth Z_w, and parallax d is directly proportional to baseline length B, that is, the smaller the depth, the larger the parallax; the wider the baseline, the larger the parallax, so the wide baseline measurement system usually has higher depth resolution.

B. FOCUS DEPTH CUE

Compared with traditional cameras, the camera array imaging has the characteristics of smaller depth of field and refocusing, so the clarity and fuzziness of its imaging points contain more abundant depth clues, which is called focus depth clues in this study. In the aperture imaging theory, when the object distance a, image distance b and focal length f of a certain spatial point meet the classical imaging formula Eq.(8), a clear image of the target point can be obtained, otherwise defocusing blur will occur [19].

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \tag{8}$$

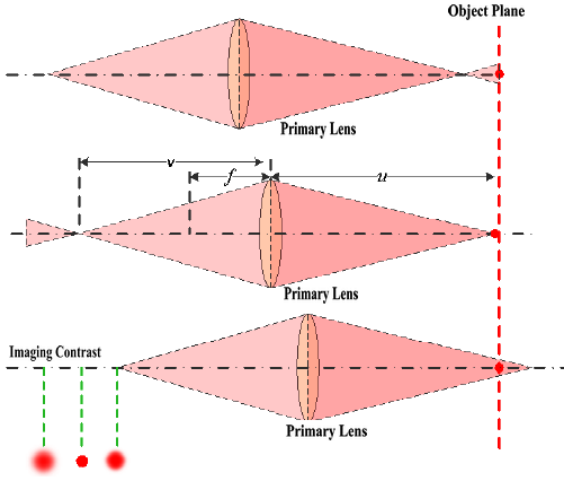


FIGURE 11. Depth map obtained by focusing method.

Zoom method is a classical algorithm based on focus depth clues, which can be classified as Depth From Focus (DFF) and Depth From Defocus (DFD) according to different calculation principles. Since the focusing method is essentially a search process without problem of ambiguity, so the accuracy and reliability of depth calculation are generally better than defocus method. In addition, the focusing method has the characteristics of simple principle and fast calculation, and therefore the focus method is used to extract the focus clues of the image.

In terms of the focusing method, it is necessary to adjust the relative positions of the camera and the target object several times in the depth direction to obtain image sequences of different focusing situations, which is shown in Figure 11. Furthermore, an appropriate focus measure method is used to calculate the clarity of the imaging point, and the clearest imaging position is selected as the depth calculation result. The focus method is essentially a search method, and its core is to design a reliable focus measurement [20].

C. CHARACTERISTIC ANALYSIS OF PARALLAX METHOD AND FOCUSING METHOD

From a surveying point of view, using parallax method and focusing method to estimate scene depth is essentially the same, because the measuring baseline in parallax method and aperture size in focusing method have the same physical essence. However, the solution process and depth calculation results of the two methods are obviously different. Therefore, based on the analysis model of light field sampling space and EPI (Epipolar Plane Image) image in the previous chapter, the reasons for the differences between parallax method and focus method are expected to be further explored.

Fig. 12 shows the optical field sampling diagram of parallax cue and focus cue, where the V axis still represents the light field angle sampling, the T axis represents the light field position sampling, and the angle of the light field sampling distribution at different target points will tilt with the

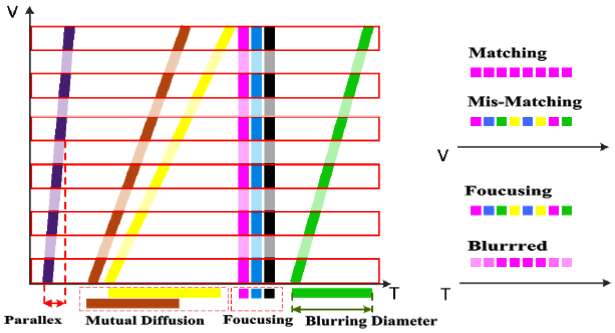


FIGURE 12. Optical field sampling diagram of parallax cue and focus cue.

change of its depth. On the left side of Fig.12, when the angle sampling is inclined, the signal of the target point will affect different positions on the T axis, and the size of its influence area is the diameter of the fuzzy circle formed by point diffusion. When the signal of the same target point is observed at different angles on the V axis, the horizontal displacement of its sampling point on the T axis represents the magnitude of parallax, which is shown in the sampling distribution of the leftmost purple point in Fig.12.

Due to sample tilt, the sampling of the same position and different target points on the T axis will affect each other, which shows the diffusion effect of the light field image in the unfocused region [21]. In parallax method, it is often assumed that the surface of the target object is diffuse reflection surface, and sampling at different angles at the same spatial point has consistency, which is called lambertian hypothesis. If all angle samples corresponding to the same position of the T axis meet the similarity criteria defined by the parallax method, then the angle samples at that position are considered to match; On the other hand, if the consistency of different angle samples is weak, then the angle samples do not match.

Unlike parallax method, more attention will be paid to the size of the area affected by the sampling distribution of the same target in focusing method. If the inclination angle of the sampling distribution at the same point is larger, the area affected by it will increase, and the diffusion effect between different sampling points is obvious [22]. At this time, the imaging at this position of the T axis is relatively blurred. On the other hand, the sampling between different points does not affect each other, and the change of adjacent sampling points is significant, then the region shows focusing characteristics.

From the abovementioned analysis, it can be seen that the sampling characteristics of parallax method and focus method are not consistent. Parallax method focuses on measuring the characteristics of the sampling points at the same position on the T axis changing with the angle, while focusing method mainly measures the intensity of the mutual influence and diffusion effect between the sampling points at adjacent positions on the T axis. Parallax method is based on matching features of points, while focusing method is based on focusing features of regions. The previous analysis qualitatively

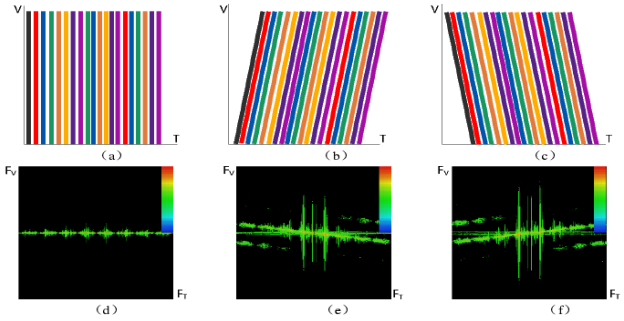


FIGURE 13. EPI frequency domain analysis of parallax sampling characteristics.

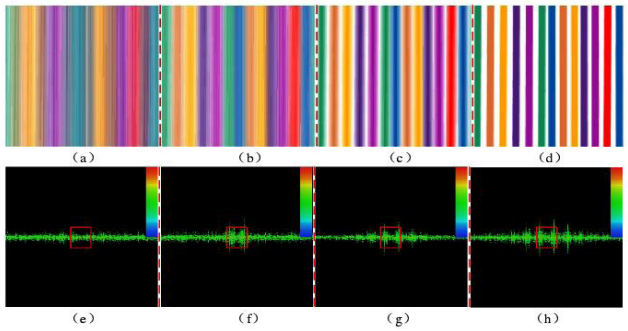


FIGURE 14. EPI frequency domain analysis of focusing sampling characteristics.

describes the difference between parallax method and focusing method in airspace. In order to quantitatively evaluate the relationship between the sampling features concerned by the two methods, Figs. 13 and Fig.14 respectively describe the frequency domain characteristics of the two methods' two-dimensional EPI patterns under different circumstances.

Figs. 13 a)-c) shows the light field sampling distributions when the target points are at different distances from the focal plane, and the characteristics of the spectral response are shown in Fig. d)- f). In the Fig.13, F_T is the frequency domain correspondence of the T axis. When the target object is in the focal plane position, the energy will be concentrated in the narrow band region of $F_T = 0$. The light field sampling at a certain position on the T axis has angle invariance, and the parallax-based method can obtain a strong matching response at this moment. On the contrary, when the object is far away from the focal plane, the light field sampling changes obviously with the angle, so it is difficult to obtain a stable matching result based on parallax [23].

Unlike the matching method, the focusing method uses the gradient of the image area as the basis for judging the degree of focus. When the target object is clearly focused, there is no signal interaction between adjacent pixels of the image. However, when the target object is not clearly focused, there is a diffusion effect of adjacent pixels. In order to analyze this phenomenon, the light field samples in Fig. 13 (a) is diffused to different degrees in the direction of the T axis, which is shown in Fig. 14 (a)∼(d). When there is a complex occlusion relationship between different depth points, the diffusion effect in the direction of T axis is difficult to intuitively express. Therefore, the mutual influence of points at the same depth is only considered in Fig.14, to illustrate the focus method's emphasis on measuring the regional characteristics of diffusion effects.

Besides, as can be seen from the spectra of Fig.14 (e) ∼ (h), the energy in the narrow band region $F_T = 0$ has changed significantly. When the pixels in this region are in a non-diffusion state, the proportion of high frequency components in the frequency band region is significantly increased, and the texture in this image region is clearer at this time. On the other hand, due to the mutual diffusion between pixels in the region, the texture in the image region becomes blurred.

To sum up, parallax method and focus method have the same surveying essence, but there are differences in the sampling characteristics that they focus on when extracting depth clues. Therefore, a depth estimation method that combines parallax method and zoom method is proposed in this study, to improve the accuracy and robustness of depth calculation.

D. DEPTH ESTIMATION METHOD COMBINED PARALLAX AND FOCUSING METHOD (DEPFM)

The flow chart of the method proposed in this section is shown in Fig.15. As shown in Fig.15, the camera array is used as a platform, and the optical field imaging method described in section 2 is used to synthesize a series of optical field images focused on different depths after collecting the data of the target scene, which is also called confocal image sequence. Furthermore, based on the multi-view data set of the camera array and the confocal image sequence data set, the parallax method and the focusing method are combined in this method, and the global optimization method is used to calculate the scene depth finally.

Due to the differences in brightness and color between different camera images, a matching method that is insensitive to the overall changes in brightness when performing matching calculations on camera array data sets is used in this study. Since the Normalized Cross Correlation method (NCC) has the feature of de-averaging, the calculation results are not sensitive to illumination changes or overall changes in the brightness of the image area. Therefore, the NCC method is used to extract the matching metric function in the parallax cue process. The radii of the matching window in the horizontal and vertical directions are denoted as n and m respectively, then the matching metric of a pair of image points $I_a(x, y)$ and $I_b(x, y + d)$ in the cameras a and b is [24]:

$$NCC(x, y, d) = \frac{1}{N-1} \sum_{k=1}^{N} NCC_k(x, y, d) \tag{9}$$

The physical meaning of calculating NCC lies in seeking out the angle between the directions formed by pixels in the different distribution regions, and the larger the value is, the higher the distribution schedule is. Unlike the

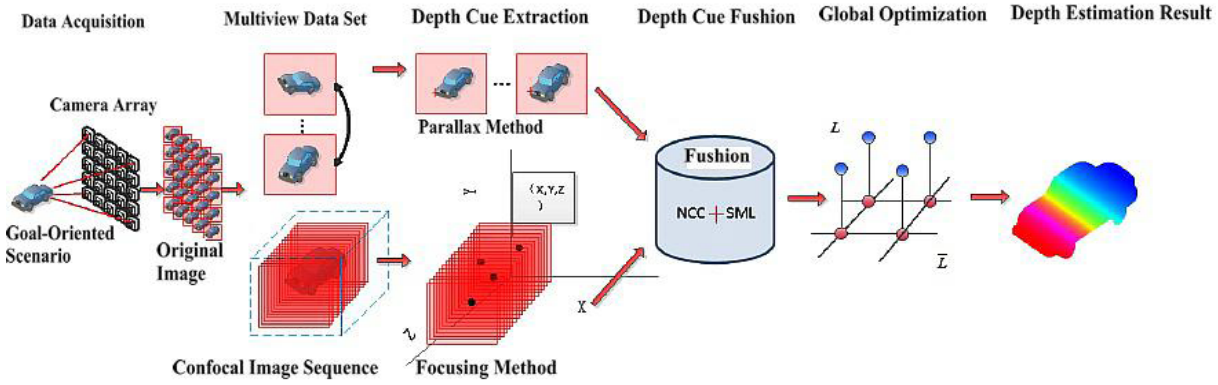


FIGURE 15. Flow chart of depth estimation method combsined parallax and focusing method.

FIGURE 16. 10 sets of light field images of different focal planes.

classical laplacian operator, the improved Laplacian summation operator Sum-Modified-Laplacian (SML) is used to extract focus depth clues in the light field image, based on which the independent calculation of different directions can be conducted, to avoid the problem that the horizontal and vertical calculations may cancel out.

$$\begin{cases} ML\,(x,y,d) = \left\|2I'_d\,(x,y) - I'_d\,(x-1,y) - I'_d\,(x+1,y)\right\| \\ \qquad\qquad + \left\|2I'_d(x,y) - I'_d\,(x,y-1) - I'_d\,(x,y+1)\right\| \\ SML\,(x,y,d) = \sum_{i=-n}^{n}\sum_{j=-m}^{m} ML\,(x+i,y+j,d) \end{cases} \tag{10}$$

Where I'_d represents a candidate confocal image corresponding to parallax d.

According to the characteristics of the two kinds of depth clues, the matching candidate points that obviously do not meet the angle invariance are considered to be eliminated firstly, and then the calculation results of the two kinds of depth clues are comprehensively utilized to calculate the scene depth:

$$F\,(x,y,d) = \alpha \cdot NCC\,(x,y,d) + (1-\alpha) \cdot SML\,(x,y,d) \tag{11}$$

Where α is the normalized coefficient.

V. CASE STUDY

In this section, the proposed depth estimation method combined with disparity and focus fusion is expected to be verified on virtual data and real data respectively.

As shown in Figure 16, a relatively simple virtual scene is described, where the three books in the scene are in different depth planes respectively. In this section, 10 groups of light field images with different focus settings are extracted from the 100 groups of images, and three different imaging positions *a, b, c* are randomly selected for analysis.

Based on three depth estimation method of NCC, SML and the proposed DEPFM, Fig. 17 (a)∼(c) show fusion effect of the calculation results the position point *a, b, c* in the Fig.16 respectively. In this case, the actual focus position of point *a* is located on the 8th group of images. However, the extreme points obtained by SML method and NCC method are different. Therefore, the difference between SML and NCC method in the calculation process of depth clues in different regions is demonstrated.

As shown in Figure 17 (c), SML method is more robust in areas with strong high frequency components, while NCC will show deviation in calculation results in areas with complex texture or repeated texture. Based on Fig.17, the characteristics of parallax method and focus method in the process of obtaining scene depth and the possibility of merging two methods are preliminarily verified.

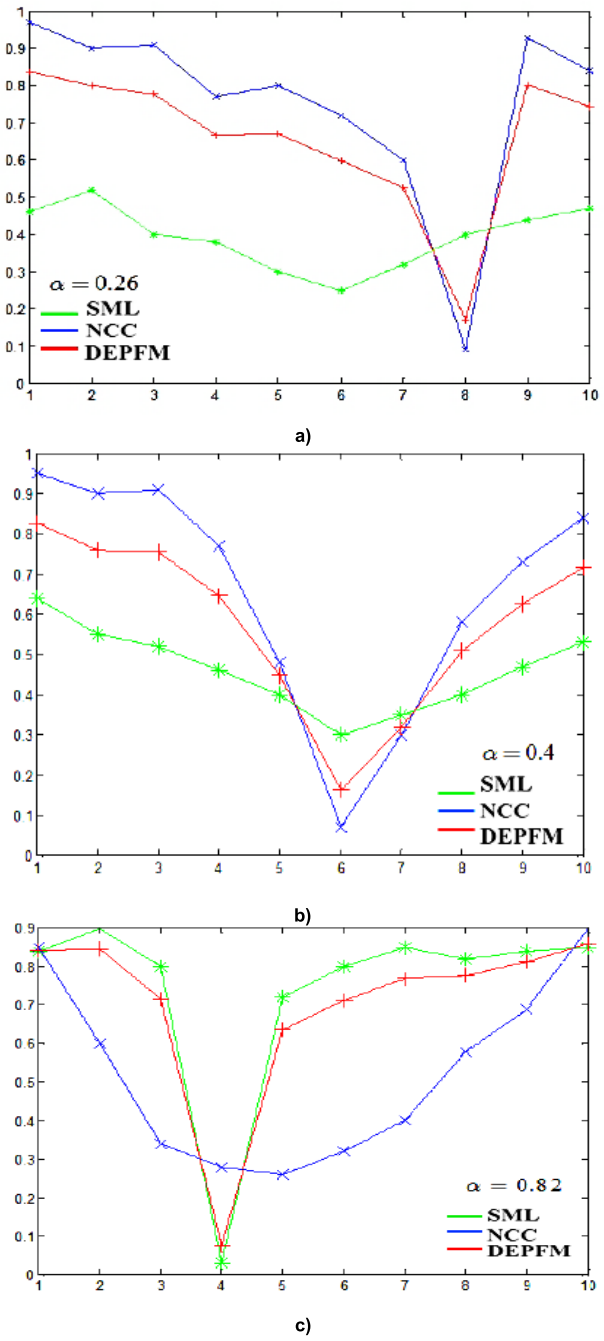


FIGURE 17. Fusion analysis of parallax and focus calculation results. a) Depth calculation result at point *a*. b) Depth calculation result at point *b*. c) Depth calculation result at point c.

Furthermore, the depth estimation experiments under virtual scene conditions are carried out. As shown in Fig.18 (a), in addition to the data contained in Fig.17, a group of virtual scenes with complicated relationships of objects and occlusions are added to the experiment, and accurate knowledge of the depth of the virtual scene can facilitate quantitative analysis of subsequent experimental results. Fig.18 (c) shows the fusion weights of the results calculated by SML and NCC in the two groups of virtual experiments.

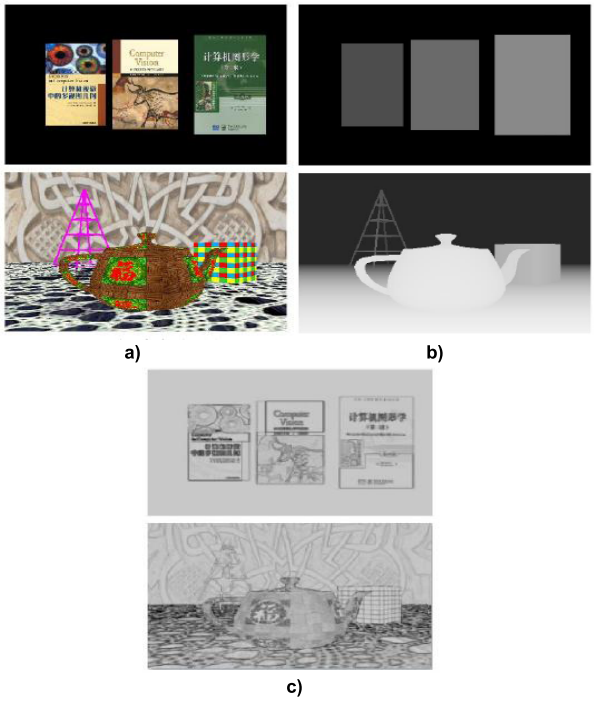


FIGURE 18. Two sets of virtual experiment scenes. a) two sets of virtual scenes. b) depth map of virtual scenes. c) diagram of fusion weight.

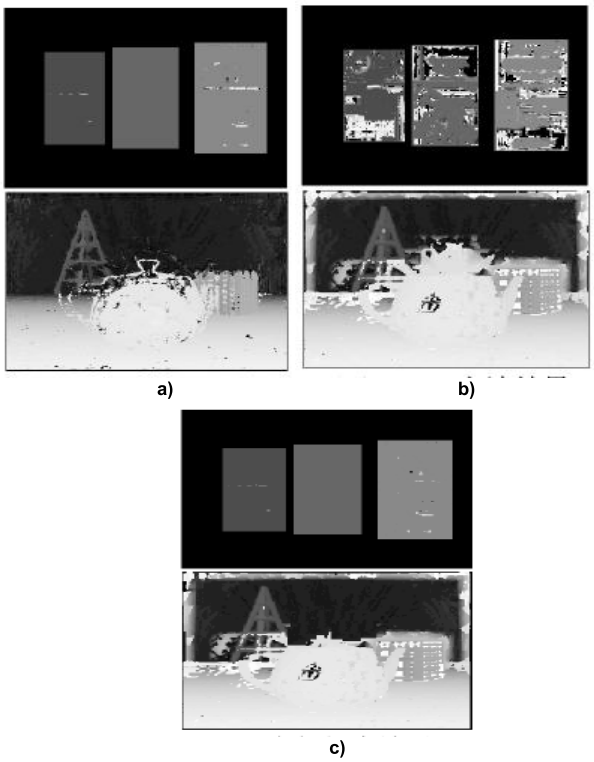


FIGURE 19. Two sets of virtual scene depth calculation results. a) Results obtained on ncc. b) Results obtained on sml. c) Results obtained on depfm.

In the virtual scene, the NCC, SML and DEPFM methods are used to calculate the depth, and the results are analyzed quantitatively as shown in Figure 19 and Figure 20.

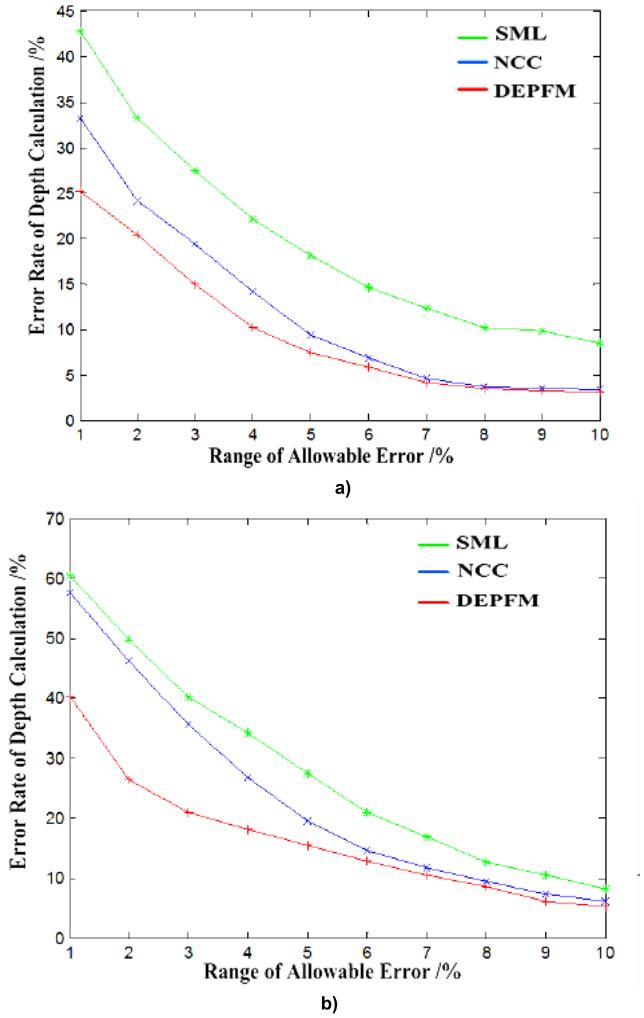


FIGURE 20. Quantitative analysis for two sets of virtual depth calculation results. a) Depth calculation result at point *a*. b) Depth calculation result at point *b*.

From Fig.19, it can be obtained that for more complex scenes, the NCC method is less robust to edges with discontinuous calculation depth, while SML is more robust to edge regions. As shown in Figure 19 (c), more accurate depth calculation results can be obtained on part of the spout of a teapot by the fusion method proposed in this paper.

In order to eliminate the uncertainty caused by the global optimization method for depth cue fusion, Figure.20 quantitatively compares the results of NCC, SML, and DEPFM method when global optimization is not carried out, based on which the DEPFM is demonstrated to be superior to other two methods obviously. As a result, the proposed DEPFM can be used to achieve ideal depth calculation results for both groups of virtual scenes. Under the condition that the allowable error range is less than 10 % of the total depth, the accuracy of the calculation results is above 90 %.

Based on the aforementioned multi-groups of experiments on virtual scene and real scene, the depth estimation method combined with disparity clues and focus clues proposed in this chapter has high calculation accuracy is demonstrated further. Compared with NCC matching method based on parallax and SML method based on focus evaluation, the DEPFM method proposed in this paper has superior robustness to the method based on a single depth cue for the situation of depth discontinuity area, texture similarity in the target scene.

VI. CONCLUSION

In recent years, large-scale and complex traffic data have been collected with wide deployment of ITS. It is usually isolated or has some dispersed links. This raises a new challenge to extract or mine additional information from these large and complex data. In the field of Big Data in IoT, there is no general approach to this problem. However, a powerful idea is to integrate the data to compose a network by the links reflecting their relationships. The information of the data is hidden in the network. Thus, the useful information or solution methods can be obtained based on this network. Starting from the theory of optical field imaging and the depth estimation based on optical field, this paper focuses on the depth cue fusion in the depth estimation process based on optical field Big Data in IoT.

This chapter analyzes the differences between parallax method and focusing method in two-dimensional light field space, and proposes a depth estimation method by combining parallax method and focusing method. This method constructs a unified depth cue extraction model for camera array datasets, and achieves high-precision scene depth estimation under the framework of global optimization. The experimental results show that the proposed method is robust to depth discontinuities and complex texture regions. On the basis of analyzing the characteristics of different depth clues in the optical field data set, a depth estimation method combining parallax method and focusing method (DEPFM) is proposed. Firstly, this paper analyzes the disparity clues and focus clues contained in the multi-view data set and the light field refocusing image set of the camera array respectively, and points out the differences and relationships between the two depth clues extraction methods in the light field sampling frequency domain space, that is, the disparity method focuses on the energy concentration characteristics near the frequency domain spatial angle axis, while the focus method focuses on the high frequency proportion of energy distribution on the angle axis. Then, the weighted linear fusion method based on image gradient is used to fuse the two calculation results, which improves the accuracy and robustness of depth estimation. Finally, the results of depth estimation experiments on different sets of scenes show that compared with the method based on a single depth cue, the method in this paper has higher accuracy in depth calculation in discontinuous areas of scene depth and similar texture areas.

REFERENCES

[1] S. Choi and S.-W. Min, "Depth estimation method using depth-of-field imaging with a retroreflector," *Opt. Express*, vol. 26, no. 5, pp. 5655–5664, 2018.

[2] T. Tao, Q. Chen, S. Feng, Y. Hu, and C. Zuo, "Active depth estimation from defocus using a camera array," *Appl. Opt.*, vol. 57, no. 18, pp. 4960–4967, 2018.
[3] S. Zhang, H. Sheng, D. Yang, J. Zhang, and Z. Xiong, "Micro-lens-based matching for scene recovery in lenslet cameras," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1060–1075, Mar. 2017.
[4] H. Sheng, S. Zhang, X. Cao, Y. Fang, and Z. Xiong, "Geometric occlusion analysis in depth estimation using integral guided filter for light-field image," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5758–5771, Dec. 2017.
[5] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2146–2159, May 2018.
[6] Z. Ma, Z. Cen, and X. Li, "Depth estimation algorithm for light field data by epipolar image analysis and region interpolation," *Appl. Opt.*, vol. 56, no. 23, pp. 6603–6610, 2017.
[7] J. Chen, J. Hou, Y. Ni, L.-P. Chau, "Accurate light field depth estimation with superpixel regularization over partially occluded regions," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4889–4900, Oct. 2018.
[8] M. Feng, Y. Wang, J. Liu, L. Zhang, H. F. M. Zaki, A. Mian, "Benchmark data set and method for depth estimation from light field images," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3586–3598, Jul. 2018.
[9] W.-Y. Lee, C.-Y. Li, and J.-Y. Yen, "Integrating wavelet transformation with Markov random field analysis for the depth estimation of light-field images," *IET Comput. Vis.*, vol. 11, no. 5, pp. 358–367, Aug. 2017.
[10] D. Wan and J. Zhou, "Multiresolution and wide-scope depth estimation using a dual-PTZ-camera system," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 677–682, Mar. 2009.
[11] L. Zhu *et al.*, "Single image depth estimation based on convolutional neural network and sparse connected conditional random field," *Opt. Eng.*, vol. 55, no. 10, p. 103101, 2016.
[12] K. Okazaki, "Experimental study to improve of transillumination image by depth estimation using light source of two wavelengths," *Amer. Mineral.*, vol. 89, nos. 1–2, pp. 654–662, 2016.
[13] W. Chen, M. Zhang, and Z. Xiong, "Nonsingle viewpoint omni-stereo depth estimation via space layer labeling," *Opt. Eng.*, vol. 50, no. 4, p. 047005, 2011.
[14] P. Lang *et al.*, "Bulge growth and quenching since z = 2.5 in candels/3D-HST," *Astrophys. J.*, vol. 788, no. 1, pp. 1858–1871, 2014.
[15] K. Yang, L. Xu, Q. Yang, and R. Duan, "Striation-based source depth estimation with a vertical line array in the deep ocean," *J. Acoust. Soc. Amer.*, vol. 143, no. 1, pp. EL8–EL12, 2018.
[16] A. Abashian *et al.*, "Backward production in at 8 GeV/," *Astron. Astrophys.*, vol. 573, no. 1, pp. 163–176, 1976.
[17] H.-Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C.-K. Tang, "Pop-up light field: An interactive image-based modeling and rendering system," *ACM Trans. Graph.*, vol. 23, no. 2, pp. 143–162, 2004.
[18] E. C. Shang, "Source depth estimation in waveguides," *J. Acoust. Soc. Amer.*, vol. 77, no. 4, pp. 1413–1418, 1998.
[19] Q. Wu, G. Xu, M. Li, L. Chen, X. Zhang, and J. Xie, "Human pose estimation method based on single depth image," *IET Comput. Vis.*, vol. 12, no. 6, pp. 919–924, 2018.
[20] F. Acero *et al.*, "Detection of gamma rays from a starburst galaxy," *Science*, vol. 326, no. 5956, pp. 1080–1082, 2009.
[21] S. Ahlen *et al.*, "The case for a directional dark matter detector and the status of current experimental efforts," *Int. J. Mod. Phys. A*, vol. 25, no. 1, pp. 1–51, 2010.
[22] P. G. Pérez-González *et al.*, "Ultraviolet through far-infrared spatially resolved analysis of the recent star formation in M81 (NGC 3031)," *Astrophys. J.*, vol. 648, no. 2, p. 987, 2006.
[23] S. C. Chapman *et al.*, "The properties of microjansky radio sources in the hubble deep field-north, SSA 13, and SSA 22 fields," *Astrophys. J.*, vol. 585, no. 1, pp. 57–66, 2003.
[24] R. Houborg, M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik, "Drought indicators based on model-assimilated gravity recovery and climate experiment (GRACE) terrestrial water storage observations," *Water Resour. Res.*, vol. 48, no. 7, pp. 2515–2521, 2012.

YUE WU received the B.S. degree in computer science and the M.S. degree in management science and engineering from the National University of Defense Technology, Changsha, Hunan, China, in 1997 and 2001, respectively, where he is currently pursuing the Ph.D. degree with the College of Information System and Management. His current research interests include image super-resolution and image processing.