# Light field imaging based on a parallel SVM method for recognizing 2D fake pedestrians[*]

**JIA Chen**[1], **ZHANG Yao**[1]**, **SHI Fan**[1,2]**, **and ZHAO Meng**[1]

*1. Key Laboratory of Computer Vision and System of Ministry of Education, Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China*

*2. The MOE Key Laboratory of Weak-Light Nonlinear Photonics, Nankai University, Tianjin 300457, China*

(Received 7 April 2021; Revised 25 May 2021)

It is novel to apply three-dimensional (3D) light field imaging technology to recognize two-dimensional (2D) fake pedestrians. In this paper, we propose a parallel support vector machine (SVM) method based on 3D light field imaging (light field camera) and machine learning techniques. A light field (LF) camera with robust sensors, which is able to record rich 3D information, is used as hardware equipment. Histogram of oriented gradient (HOG) feature extraction algorithm and SVM classification method are used to recognize the real and 2D fake pedestrians efficiently. Besides, we carry out an experiment on our improved LF pedestrian dataset. The experimental results of parameter optimization study show that in the case of 400 training samples (200 positive samples and 200 negative samples), 120 to 420 testing samples, and an HOG cellsize as 8×8, the best recognition accuracy with polynomial kernel function is improved by more than 2% compared with the previous method. The best accuracy is 99.17%. Otherwise, the recognition accuracy of more than 98.00% will be obtained even under other experimental conditions.

**Document code:** A **Article ID:** 1673-1905(2022)01-0048-6

**DOI** https://doi.org/10.1007/s11801-022-1047-4

The rapid development of light field (LF) imaging technology provides theoretical support and technical guidance for the research of pedestrian recognition[1,2]. Nowadays, pedestrian recognition technology is widely used in the fields of intelligent robots, video surveillance, and automotive safety. It has been developed from the original experimental research to the large-scale application[3,4]. However, the wrong recognition results occur frequently as well, as shown in Fig.1, the pedestrian traffic signs are recognized as real pedestrians. The primary reason is that in the process of videos or image acquisition, due to the limitations of two-dimensional (2D) imaging equipment, the depth information of pedestrians is missing, making it impossible to express the object model from all angles.

The difference between 2D and three-dimensional (3D) imaging devices is that the detector of traditional 2D imaging devices can only respond to the intensity of light resulted in losing the direction information. In contrast, the 3D imaging device can record the light intensity as well as the detection. Meanwhile, it can generate the depth maps of the scene[5]. As a new type of 3D imaging device, the LF camera has been widely used in scientific research[6,7]. By combining lens, microlens array, and photosensor, an LF sensor is constructed, which can obtain the LF information of the scene. As shown in Fig.2, the microlens array is an integer array and contains multiple units, and a set of 2D arrays can represent it. At the same time, there is a conjugation relationship between the ultraviolet (UV) surface of the main lens and photosensor. Especially, when each beam of light in the lens passes through the microlens array, sub-aperture images of microlens are left on the photosensor. When the complete light path distribution in the camera is obtained, the light will be re-projected to a new plane, and this process is equivalent to the manual focusing of 2D cameras. In essence, it can be realized by mathematical calculation, so it also called digital focusing[8]. Through this process, the position images of different image planes can be acquired with only one exposure under the support of a large-aperture by LF camera. The 3D information of the scene can be easily obtained. Nowadays, compared with other RGB-D cameras, such as time of flight (TOF), binocular vision, etc[9,10], it has many advantages, such as a single sensor, high resolution, high-cost performance, low environmental restriction, low noise for imaging and sensing, comfortable operating in the later stage, etc. Therefore, this paper uses Lytro-Illum, the second generation of consumer-grade LF cameras produced by Lytro Company, to carry out the experimental research on 2D fake pedestrians. It should be noted that the format of the imported data generated by the LF camera is. LF

[*] This work has been supported by the National Natural Science Foundation of China (Nos.61906133, 62020106004, 92048301 and 61703304).

[**] E-mails: zytju221@tju.edu.cn; shifan@email.tjut.edu.cn

picture (.LFP) is a unique output format of the Lytro-Illum camera. At the same time, there are two methods to resolve the format. The first is the official Lytro software produced by Lytro company, one type of desktop personal computer software. The second is LFP reader program based on MATLAB toolbox, an open-source/method that can be used by anyone. In our experiment, we choose the former method, because Lytro software has the advantage of simple operation and fast processing speed[11]. The requirement of processing speed is essential to our research. Meanwhile, the second method needs to analyze the obtained data (.RAW/.JOSN) in MATLAB, it takes a long time to process only one LF image, which is not in line with the original intention of our banquet design.

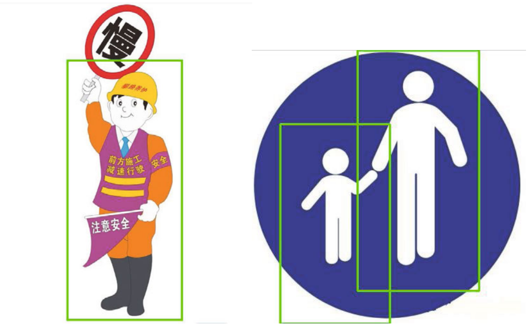


**Fig.1 The wrong recognition results of 2D fake pedestrians (traffic signs)**

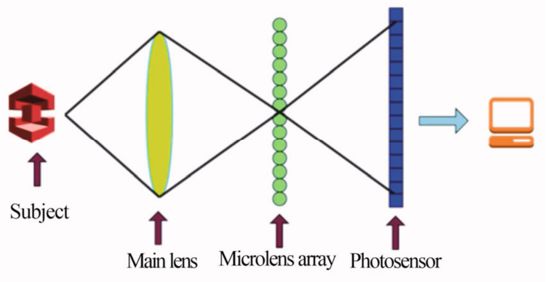


**Fig.2 Schematic diagram of LF camera imaging**

In previous work, our research group first verified the feasibility of the LF camera in the study of two-dimensional fake pedestrian recognition, but this work is not perfect. The proposed method is only a serial-based classic support vector machine (SVM). There is still room for improvement. So, in this paper, we propose a new parallel strategy based on LF camera and SVM. At the same time, to make our method more universal, we improved the data set and considered more realistic scenarios. Further, based on the optimization of machine learning parameters, we carried out sufficient experimental verification. Therefore, compared with the previous feasibility study, this paper has different strategies for machine learning and the data set used in the experiment, and the final experimental results have also greatly improved.

Compared with the first pedestrian LF dataset constructed by the method of Ref.[12], the positive and negative samples in the dataset were further improved by taking into account the situations that often occur in the real scene. For positive samples, more 2D fake pedestrian images in mutual occlusion are added, and the single pedestrian images in the simple scene are reduced. For negative samples, more traffic sign images and signboard images are added, and the liquid crystal display (LCD) images are reduced.

Based on light field imaging and machine learning techniques, a new parallel 2D fake pedestrian recognition framework was proposed. This is different from all previous researches in the field of LF pattern recognition. Meanwhile, compared with the previous method, the results in this paper are much better than those before.

Based on our recognition framework and the improved dataset, the effects of histogram of oriented gradient (HOG) feature parameter cellsize and different kernel functions were studied on experimental results, the best effect was achieved when set the cellsize as 8×8, and kernel function is polynomial.

In our process of extracting HOG features, the input images come from the improved pedestrian LF dataset we constructed, the size of each image is 64×128. Firstly, the input images are normalized to reduce the impact of illumination, although we have already considered the factor in the acquisition of our dataset. Secondly, the images are divided into small spatial regions (cellsize), and then we use the one-dimensional horizontal gradient operator [−1, 0, 1] and the one-dimensional vertical gradient operator $[-1, 0, 1]^T$ to calculate the horizontal and vertical gradients at each point $(x, y)$ of the cell. Finally, the gradient of one pixel in the corresponding image is shown as

$$G_x(x, y)=H(x+1, y)-H(x-1, y), \tag{1}$$

$$G_y(x, y)=H(x, y+1)-H(x, y-1), \tag{2}$$

where $G_x(x, y)$ is the horizontal gradient of one pixel, $G_y(x, y)$ is the vertical gradient of one pixel, and $H(x, y)$ is the pixel value of one pixel.

Therefore, the gradient value of pixel points can be expressed as Eq.(3) and the gradient direction can be expressed as Eq.(4).

$$G(x,y)=\sqrt{G_x^{\,2}(x,y)+G_y^{\,2}(x,y)}, \tag{3}$$

$$\theta(x,y)=\tan^{-1}\left[\frac{G_x(x,y)}{G_y(x,y)}\right]. \tag{4}$$

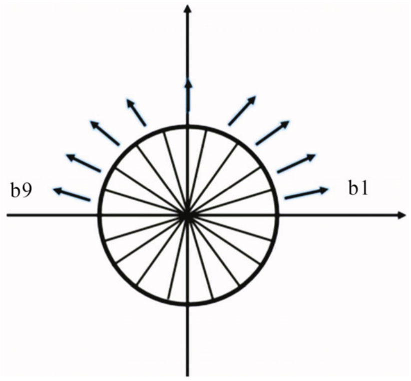


**Fig.3 Schematic diagram of gradient direction projection**

SVM is an efficient machine learning algorithm, which is widely used in the research of classification problems[13]. In our experiment, for the study of 2D fake pedestrian recognition, it essentially belongs to a nonlinear problem. Therefore, for the samples in our dataset ($x_1$, $y_1$), ($x_2$, $y_2$) …($x_n$, $y_n$), $x_i \in R^n$, $y_i \in [-1, +1]$ indicates the class to which $x_i$ belongs to, we use kernel function for nonlinear transformation operation and map samples to high-dimensional feature space $H$. According to Mercer[14], the optimal decision function can be expressed as Eq.(5). Through this transformation, we have found the optimal classification plane in $H$, as shown in Fig.4.

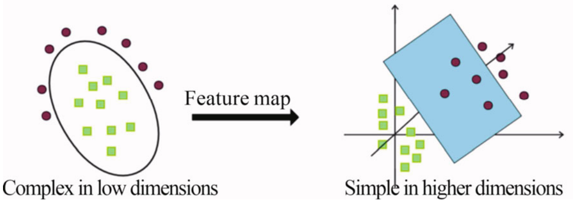


**Fig.4 The principle of SVM**

$$f(x) = \operatorname{sgn}\left\{\sum_{i=1}^{l} \alpha_i y_i K(x \cdot x_i) + b\right\}. \qquad (5)$$

Fig.5 is the recognition framework. The hardware device used in the experiment is Lytro-Illum, and the software is Lytro desktop software[15].

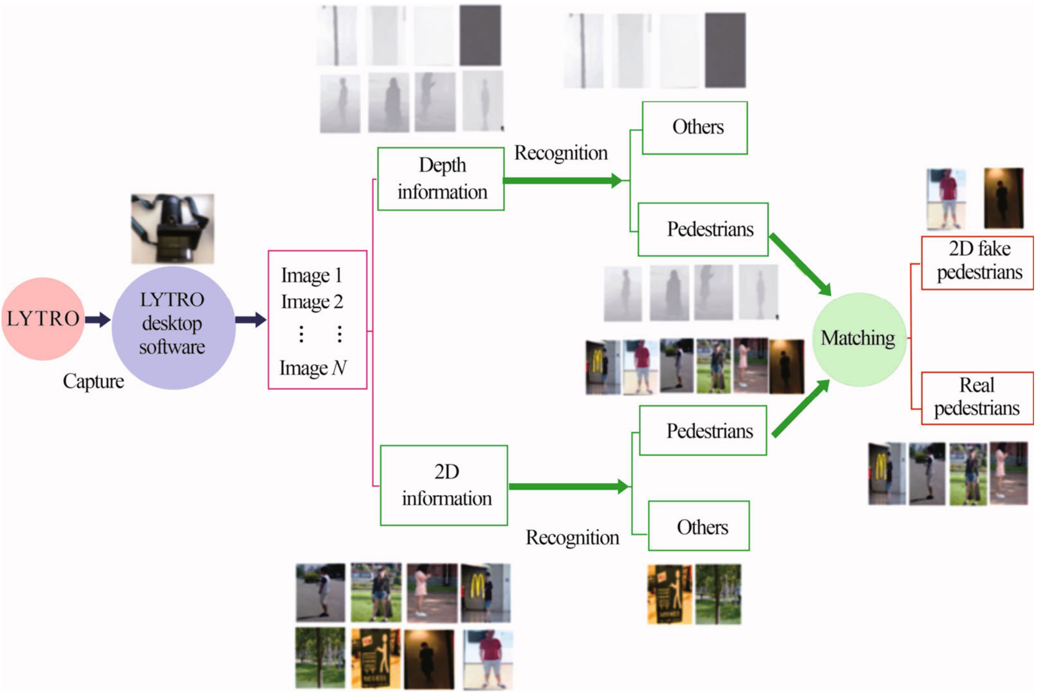


**Fig.5 The proposed recognition framework**

At the beginning of the experiment, we obtain a large number of images in the real scene through Lytro-Illum. The images include real pedestrian images, 2D fake pedestrian images, and other images.

After that, the acquired images are processed by Lytro software. And the images containing 2D information and depth information are obtained, respectively.

Then, the parallel method of HOG&SVM was applied to recognize 2D images and the depth images separately. In this process, the SVM first algorithm maps the extracted two-dimensional features to the three-dimensional space, then finds the optimization model with the maximum interval by calculating the geometric interval between the hyperplane and the support vector. Through this process, the pedestrian images (including the real pedestrians and 2D fake pedestrians) are obtained, and the interference of other life scenes is eliminated. At the same time, the training samples do not need to be retained in the training process, and the final model is only related to the support vector. The method of HOG&SVM is shown in Fig.6.

Finally, we matched the pedestrian images of 2D information and depth information. If the images contain the same pedestrians, we will believe that the pedestrians in the images are the real ones.

The increasing interest in pedestrian recognition leads to research on the diversity of pedestrian datasets. Since the properties of the benchmark dataset are still defective, Ref.[12] first used the Lytro LF camera to build a pedestrian LF dataset, which contained 1091 samples. This dataset played a vital role in the application of LF imaging technology to the study of the 2D fake pedestrian problem. Moreover, it proved the rationality and feasibility of our proposed recognition framework. However, through the research, we found that the dataset also has some limitations; the main limitations are as follows.

1. The dataset contains large positive samples of a single pedestrian, and the dataset got fewer pedestrian scenes in mutual occlusion.

2. For negative samples in the dataset, we collected large LCD images and high-definition (HD) images. While taking into account the reality, the samples of traffic sign images and signboard images (2D images) appear more frequently than LCD and HD images in the real scene.

With this in mind, we improved the previous dataset. For the positive samples, we increased the number of pedestrian images under mutual occlusion and reduced one single pedestrian images in real life, as shown in Fig.7. For the negative samples, we reduced the LCD images and HD images and added many traffic signs images and signboard images (2D images). Some added images are shown in Fig.8, the improvement of the dataset is still based on the Lytro LF camera and Lytro software. In the end, the dataset contains 500 positive samples and 500 negative samples. Meanwhile, considering the requirement of reducing hardware consumption and reaching real-time recognition, both 2D images and depth images were normalized to 128×64 in the experiments.

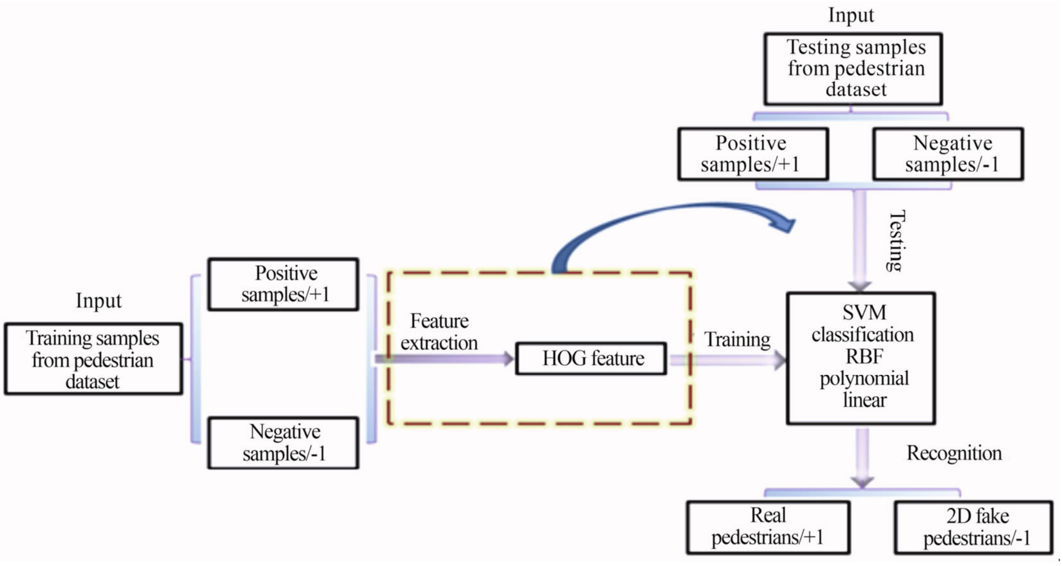
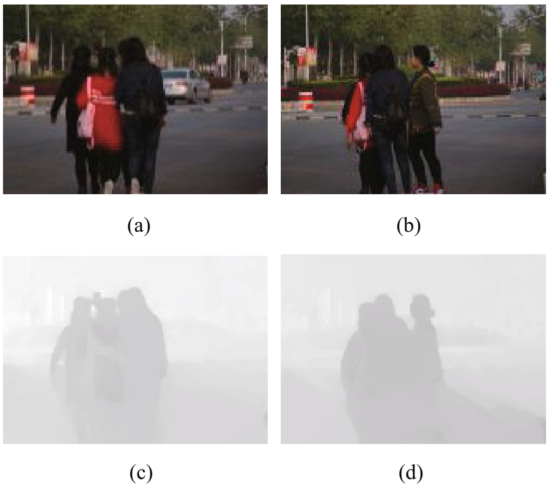


**Fig.6 Our model based on HOG&SVM**


(a) (b)

(c) (d)

**Fig.7 Some increased samples in our dataset: (a) (b) Pedestrians in mutual occlusion (positive samples); (c) (d) Depth images to (a) (b)**

(a) (b)

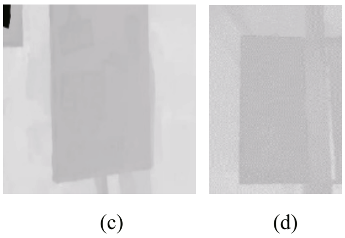

(c) (d)

**Fig.8 Some increased samples in our dataset: (a) (b) Traffic signs (negative samples); (c) (d) Depth images to (a) (b)**

In the process of extracting HOG features, we chose the cellsize as 4×4, (Ref.[12] set the cellsize as 8×8), the number of testing samples we set were 120, 180, 240, 360 and 420, the training samples were 400, and the experimental results are shown in Tab.1.

In Ref.[12], the best recognition accuracy under different kernel functions we can get is less than 97.00%. In this paper, when we select 400 training samples and 120 testing samples, and the HOG parameter cellsize set as 4×4, the best recognition accuracy we can get is 98.33% (polynomial kernel). At the same time, when other testing samples are selected (such as 300 or 360 testing samples with the linear kernel), the recognition accuracy is greater than 97.50%. Even the worst recognition accuracy is greater than 95.50% (360 testing samples with radical base function (RBF) kernel). To analyze the basic reason, if the correct ROI area is not correctly extracted in Ref.[12], the next step of classification will be affected. But for the method proposed in this paper, we adopt a parallel method to recognize 2D and depth pedestrian information at the same time, which will avoid the wrong recognition of real pedestrians.

**Tab.1 The experimental results (cellsize 4×4)**

| Testing<br>Accuracy | 120 | 180 | 240 | 300 | 360 | 420 |
|---|---|---|---|---|---|---|
| RBF | 96.67 | 96.11 | 96.25 | 96.67 | 95.56 | 95.24 |
| Polynomial | 98.33 | 97.22 | 96.67 | 97.33 | 96.39 | 97.14 |
| Linear | 97.50 | 96.67 | 97.50 | 98.00 | 97.50 | 97.38 |

For our experimental research, we found that the selection of HOG parameter cellsize has a great influence on the establishment of the training model. Therefore, in order to further test the performance of the proposed method and make the experiment more convincing, we selected and tested the influence of different cellsizes on experimental results. The two parameters we choose are 8×8 and 16×16, and the corresponding experimental results are shown in Tab.2 and Tab.3, respectively. When cellsize as 8×8 is selected, the recognition accuracy has been further improved, reaching a maximum of 99.17% (polynomial kernel, 120 testing samples). Moreover, when 420 testing samples are selected, the highest recognition accuracy can reach 98.57% (polynomial kernel). When cellsize as 16×16 is selected, we can get the highest recognition accuracy of 98.33% (120 testing samples, RBF kernel). When other testing samples are selected (such as 180 or 240 testing samples with polynomial kernel and RBF kernel), the recognition accuracy is also greater than 97.00%. In addition, through the detailed analysis of the experiment, we also found that when cellsize was selected as 4×4 and 8×8, the recognition results of polynomial kernel function are better than RBF kernel function and linear kernel function on average, when cellsize was selected as 16×16, the recognition results of RBF kernel function are better than polynomial kernel function and linear kernel function on average. The main reason is that with the increase of cellsize, the dimensions of the extracted features gradually decrease, and the mathematical form of the feature is relatively simple, the RBF kernel function is more expressive when dealing with simple features.

**Tab.2 The experimental results (cellsize 8×8)**

| Testing<br>Accuracy | 120 | 180 | 240 | 300 | 360 | 420 |
|---|---|---|---|---|---|---|
| RBF | 97.50 | 96.67 | 96.67 | 97.00 | 95.83 | 96.90 |
| Polynomial | 99.17 | 98.33 | 98.33 | 98.00 | 97.78 | 98.57 |
| Linear | 98.33 | 97.78 | 97.92 | 99.00 | 96.39 | 95.95 |

**Tab.3 The experimental results (cellsize 16×16)**

| Testing<br>Accuracy | 120 | 180 | 240 | 300 | 360 | 420 |
|---|---|---|---|---|---|---|
| RBF | 98.33 | 97.78 | 97.08 | 97.33 | 96.39 | 96.90 |
| Polynomial | 97.50 | 97.78 | 97.08 | 97.00 | 96.11 | 97.14 |
| Linear | 97.50 | 97.78 | 96.25 | 96.33 | 95.28 | 93.57 |

In summary, we have first proposed a new parallel SVM method to solve the problem of recognizing 2D fake pedestrians. Then, for the research of 2D pedestrian recognition, we have improved the previously established pedestrian LF dataset. Finally, in our improved dataset and under three different kernel functions, we studied the influence of cellsize on the experimental results during the HOG feature extraction process. The final experimental results show that in the case of 120 testing samples, 400 training samples, and an HOG cellsize as 8×8, the best recognition accuracy of the method can reach 99.17% (polynomial kernel). Meanwhile, when the cellsize are 4×4 and 16×16, the highest recognition accuracy will also be exceeded by 98.00%, reaching 98.33%. This further work demonstrates that the prospect of using light field imaging and SVM to solve the problem of recognizing 2D fake pedestrians.

## Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

## References

[1] BILAL M. Algorithmic optimisation of histogram intersection kernel support vector machine-based pedestrian detection using low complexity features[J]. IET computer vision, 2017, 11(5)：350-357.

[2] LI F, ZHANG R, YOU F. Fast pedestrian detection and dynamic tracking for intelligent vehicles within V2V cooperative environment[J]. IET image processing, 2017, 11(10)：833-840.

[3] BRAUN M, KREBS S, FLOHR F, et al. Eurocity persons：a novel benchmark for person detection in traffic scenes[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(8)：1844-1861.

[4] BILAL M, KHAN A, KHAN M U K, et al. A low-complexity pedestrian detection framework for smart video surveillance systems[J]. IEEE transactions on circuits and systems for video technology, 2016, 27(10)：2260-2273.

[5] LIU S, LI Y F. Precision 3-D motion tracking for binocular microscopic vision system[J]. IEEE transactions on industrial electronics, 2019, 66(12)：9339-9349.

[6] SEPAS-MOGHADDAM A, PEREIRA F, CORREIA P L. Ear recognition in a light field imaging framework：a new perspective[J]. IET biometrics, 2018, 7(3)：224-231.

[7] JEON H G, PARK J, CHOE G, et al. Depth from a light field image with learning-based matching costs[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(2)：297-310.

[8] MIGNARD-DEBISE L, IHRKE I. A vignetting model for light field cameras with an application to light field microscopy[J]. IEEE transactions on computational imaging, 2019, 5(4)：585-595.

[9] ANAND C, JAINWAL K, SARKAR M. A three-phase, one-tap high background light subtraction time-of-flight camera[J]. IEEE transactions on circuits and systems I：regular papers, 2019, 66(6)：2219-2229.

[10] DING Y, ZHAO Y, CHEN X, et al. Stereoscopic image quality assessment by analysing visual hierarchical structures and binocular effects[J]. IET image processing, 2019, 13(10)：1608-1615.

[11] LFP (light field photography) file reader[EB/OL]. (2014)[2021-05-20]. http：//code.behnam.es/python-lfp-reader.

[12] JIA C, SHI F, ZHAO Y, et al. Identification of pedestrians from confused planar objects using light field imaging[J]. IEEE access, 2018, 6：39375-39384.

[13] DENG F, GUO S, ZHOU R, et al. Sensor multifault diagnosis with improved support vector machines[J]. IEEE transactions on automation science and engineering, 2015, 14(2)：1053-1063.

[14] RAGHAVENDRA R, RAJA K B, BUSCH C. Presentation attack detection for face recognition using light field camera[J]. IEEE transactions on image processing, 2015, 24(3)：1060-1075.

[15] Lytro Inc[EB/OL]. (2014-06-02)[2021-05-20]. http：// www.lytro.com/.