



Multi-viewpoint Rendering Optimization of Indoor Scene Based on Binocular Vision

He Jing (✉)

School of Road Bridge and Architecture, Chongqing Vocational College of Transportation,
Chongqing 402247, China

Abstract. The traditional multi-viewpoint rendering method for indoor scenes has poor lighting and shadow effects, resulting in too dark or bright indoor scene multi-viewpoint rendering. Therefore, an optimization method for indoor scene multi-viewpoint rendering based on binocular vision is proposed. This research plans light and shadow effects based on the visual relationship between point light sources and indoor scenes; sets rendering points based on binocular vision; and renders indoor scenes with multiple viewpoint angles. The simulation experiment results show that compared with the traditional rendering method, the light and shadow effect of the studied method is excellent, and the rendered indoor scene is suitable for light and dark.

Keywords: Binocular vision · Indoor scene · Multi-viewpoint rendering · Optimization method

1 Introduction

As an important factor affecting the brightness of indoor scenes, light plays a decisive role in the optimization of rendering of indoor scenes. The applicable light environment of indoor scenes, along with the development and upgrading of construction sites, has also formed its own unique system. However, with the continuous improvement of urban living standards, the use efficiency of point light source lamps is getting higher and higher, and the lighting intensity and type are different. Literature [1] studied the optimization of multi-viewpoint rendering of indoor scenes based on perception. For the input 3D indoor scene model, a novel viewpoint perception measurement function based on geometric information, structural information and aesthetic information was first defined; and then a The multi-view overall optimization model; then the simulated annealing algorithm is used to optimize the user-specified number of multi-view sets, and finally the image is rendered. The experimental results show that for the problem of multi-view rendering optimization of indoor scenes, the proposed method can be adaptively obtained A series of comprehensive and beautiful rendering pictures. Literature [2] studies the optimization of semantic segmentation network for indoor scenes based on rgb-d images. The network processes the original data by training separately and gradually fusing, and adds a strengthening supervision module in the decoding stage to

effectively improve the accuracy of semantic segmentation. At the same time, the anti residual decoding method and jump structure are introduced to reduce the information loss. Reference [3] studies the method of building indoor and outdoor scenes based on BIM and GIS. Taking IFC and citygml as the research objects, IFC Standard building scene organization model is transformed into citygml standard through geometric reconstruction and multi-level LOD semantic mapping, and the visual rendering performance is optimized by model simplification and case processing. Finally, through the transformation of the terrain, the visual expression of the building indoor environment in the three-dimensional geographical environment is realized.

But the above three methods of rendering optimization, for the preset effect of point light disturbance is not ideal, so according to this problem in conventional methods, this paper proposes a multi view rendering optimization of indoor scene based on binocular vision. Binocular vision simulates the principle of human vision, which is a technical means of using computer to perceive distance passively. Observe an object from two or more points, obtain images under different vision, and calculate the offset distance between pixels according to the matching relationship of pixels between images, so as to obtain the three-dimensional information of indoor scene. Taking binocular vision as the innovation point, the original multi view rendering method of indoor scene is optimized.

2 Multi-Viewpoint Rendering Optimization Method of Indoor Scene Based on Binocular Vision

2.1 Get the Visual Relationship Between the Point Light Source and the Indoor Scene, and Plan the Lighting Effect

As a kind of light, point light source has two kinds of lighting characteristics: strong light and weak light. Under the strong light point light source, the indoor environment will be bright and dark, but under the weak light, the effect is not obvious. Under the action of strong light of point light source, every side of indoor objects is affected, giving people a strong visual experience. Therefore, it is necessary to obtain the visual relationship between the point light source and the indoor scene. The visual relationship between light source and space is very close. Only by controlling the effect of light and shadow can the residents feel the color difference of the building space. Therefore, according to the relevant data such as the intensity of the point light source, the disturbance area range of the point light source is determined

$$c_i = \sin(\alpha + \beta) \sum_{i=1, j=1}^n \pi r_j d_i \quad (1)$$

In the formula: α represents the illumination angle of the point light source; β represents the light scattering angle; i represents the light source position; j represents the illumination position; r_j represents the area radius; d_i represents the flash distance. According to the disturbance area of the above-mentioned point light source, the disturbance position of the indoor scene is determined, and the visual relationship between the two can be expressed by the visual mapping function:

$$\Delta_T = \frac{\gamma a \ln(s - s')}{qc_i m_j k} \tag{2}$$

In the above, γ is the optical path of point light source; a is the optical path path; s is the positive relationship; s' is the relationship error; q is the light intensity; m_j is the spatial range of the indoor scene affected; k is the indoor opening degree. According to the above formula, the visual relationship between the point light source and the indoor scene is obtained, and the multi-viewpoint light and shadow rendering effect of the indoor scene is guaranteed. The visual relationship between weak light environment and indoor scene is established, and the light and shadow effect of indoor scene is planned according to the new visual relationship. Figure 1 below is the visual relationship curve.

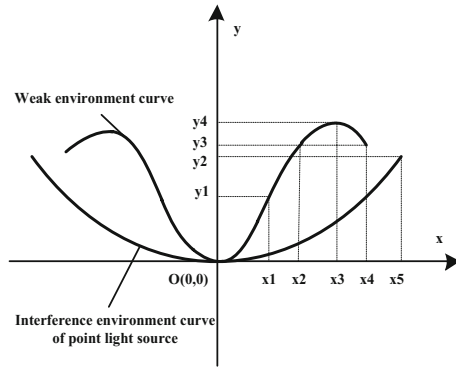


Fig. 1. Diagram of visual relation curve

According to the above figure, there is a corresponding relationship between the interference of the light source and the spatial structure, so the low-light environment established corresponds to the indoor scene. The final result of the relationship is:

$$\Delta_R = \sqrt{\frac{hx_a - 1}{\Delta_T \oplus g(\varepsilon + 1)}}^{-e} \tag{3}$$

In the formula: h is the weak light environment parameter; x_a is the curve node position; \oplus is the mapping symbol; g is the corresponding relationship; ε is the visual relationship parameter, and its value range is $[0, 1]$. According to the calculation results of the above formula, plan the light and shadow effects of indoor scenes under low light conditions. Since the low-light environment is quiet and soft, it is necessary to consider

the light scattering ability of indoor materials and household items under low-light conditions. Therefore, according to the basic characteristics of these coefficients, the light and shadow effect parameters in the indoor low-light environment are calculated. The calculation formula of the parameters is:

$$p = \arctan \frac{2\Delta_R(C - b)}{\omega D} \tag{4}$$

In the formula: C represents the light projection area; b represents the projection edge range; D represents the building space coverage area; ω represents the reflection coefficient [4]. According to the above calculation process, the lighting effect of indoor scene is planned.

2.2 Set Rendering Points Based on Binocular Vision

According to the known camera parameters, such as the overall dimension, corner type, corner coordinates under the coordinate system and the coordinates of these corners, corresponding to the coordinates in the actual environment, the internal and external parameters of the camera are calculated by using the least square method. The data parameters to be calibrated are shown in Table 1 below.

Table 1. Camera data parameter list.

Camera parameters	Expression
Internal parameters	$\begin{bmatrix} u & \gamma & m \\ 0 & v & n \\ 0 & 0 & 1 \end{bmatrix}$
Radial distortion, tangential distortion	s1, s2; p1, p2
External parameters	$W = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix} O = \begin{bmatrix} o_1 \\ o_2 \\ o_3 \end{bmatrix}$

In the above table, u , v , and γ are the scale factors of the camera on the x axis, y axis, and z axis. It should be noted that γ is the non-vertical factor; m and n are the optical centers of the two cameras. They are The internal parameters of the camera. W represents the rotation matrix, and O represents the translation vector [5]. According to the parameters in the above table, set the coordinate of the spatial point in the world coordinate system as $\delta = (x, y, z)^O$, and the corresponding projection coordinate as $\zeta = (m, n)^O$. $\delta' = (x, y, z)^O$ and $\zeta' = (m, n)^O$ are obtained after sorting, so there is a matrix formula:

$$\lambda \delta' = B[W, o]\zeta' = \begin{bmatrix} u & \gamma & m \\ 0 & v & n \\ 0 & 0 & 1 \end{bmatrix} [W, o]\zeta' \tag{5}$$

In the formula: B is the internal parameter matrix of the camera. In order to reduce the calculation error, let the value of z in the calibration angle $\delta = (x, y, z)^O$ be 0:

$$\lambda \begin{bmatrix} m \\ n \\ 1 \end{bmatrix} = B[w_1, w_2, w_3, o] \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix} = B[w_1, w_2, o] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

In the formula: w_1, w_2 and w_3 are 3×1 column vectors; $[k_1, k_2, k_3]$ represents the camera rotation matrix. Suppose there is $\delta = [x, y]^O$, then there is $\delta' = [x, y, 1]^O$. At this time, the above formula can be simplified:

$$\lambda \zeta' = \eta \delta' \quad (7)$$

In the formula: $\lambda = B[w_1, w_2, o]$ is the homography matrix, which represents the corner points extracted from the indoor scene and the mapping of its corresponding pixel points in the coordinate system [6]. Let the data in the matrix be η_1, η_2, η_3 , because w_1 and w_2 are orthogonal:

$$\begin{cases} \eta_1^O B^{-O} B^{-E} \eta_2 = 0 \\ \eta_1^O B^{-O} B^{-E} \eta_1 = \eta_2^O B^{-O} B^{-E} \eta_2 \end{cases} \quad (8)$$

B^{-O} and B^{-E} in the above formula are the projection matrices of the quadratic curve in the image coordinate system. According to the properties of the quadratic curve, let $D = B^{-O} B^{-E}$, the symmetric matrix of the camera can be obtained as:

$$\begin{aligned} D &= \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{u^2} & -\frac{\gamma}{u^2 v} & \frac{m\gamma - nv}{u^2 v} \\ -\frac{\gamma}{u^2 v} & -\frac{1}{v^2} + \frac{\gamma^2}{u^2 v} & -\frac{1}{v^2} - \frac{\gamma[m\gamma - nv]}{u^2 v^2} \\ \frac{m\gamma - nv}{u^2 v} & -\frac{m}{v^2} - \frac{\gamma[m\gamma - nv]}{u^2 v^2} & \frac{[m\gamma - nv]^2}{u^2 v^2} + \frac{m^2}{v^2} + 1 \end{pmatrix} \end{aligned} \quad (9)$$

Take the j column vector of matrix η , then there exists $\eta_j = [\eta_{j1}, \eta_{j2}, \eta_{j3}]^O$, and assume a 6-dimensional vector $\omega = [D_{11}, D_{12}, D_{22}, D_{13}, D_{23}, D_{33}]^{-O}$, then $\eta_j^O = D\eta_s = m_{js}^O \omega$ can be obtained. According to the matrix constraint conditions obtained by the above formula, we can know that:

$$\begin{bmatrix} m_{12}^O \\ (m_{11} - m_{12})^O \end{bmatrix} \omega = 0, \begin{bmatrix} m_{12}^O \\ (m_{11} - m_{12})^D \end{bmatrix} = p \quad (10)$$

Then $p\omega = 0$. When the number of collected images meets the requirements, p is a 6×6 matrix; if p is a full-rank matrix, ω can get a unique solution. At this time, the internal and external parameters of the camera can be calculated:

$$\begin{cases} \eta = D_{33} - \left[D_{13}^2 + m(D_{12}D_{13} - D_{11}D_{23}) \right] / D_{11} \\ u = \sqrt{\eta / D_{11}} \\ v = \sqrt{\eta D_{11} / (D_{11}D_{22} - D_{12}^2)} \\ \gamma = -D_{12}u^2v / \eta \\ m = (D_{12}D_{13} - D_{11}D_{23}) / (D_{11}D_{22} - D_{12}^2) \\ n = \gamma m / u - D_{13}u^2 / \eta \end{cases} \quad (11)$$

Through the above calculation process, the parameters are obtained to realize the setting of rendering point based on binocular vision [7].

2.3 Multi-viewpoint Rendering of Indoor Scenes

Considering the projection relationship between the object's point position and the panoramic image point position, the set rendering viewpoint is corrected. Due to the multi angle binocular vision shooting, the photo center will not be in the same position, resulting in the different perspective deformation state of the same scene in the overlapping area of the two groups of images, resulting in the compression of panoramic image information and distortion. Therefore, under the premise of keeping the parameters unchanged, the image viewpoint is corrected by using the translation concept, as shown in Fig. 2 [8].

In this image, c represents the origin of the viewpoint; c_1 and c_2 represent the optical center of the image sequence; L represents the projection distance; and A_1 and A_2 represent a certain point of the two groups of processed images. Make the optical center of image sequence coincide with the established coordinate system, and meet the following requirements:

$$\sigma_i = x_i y_i \quad (12)$$

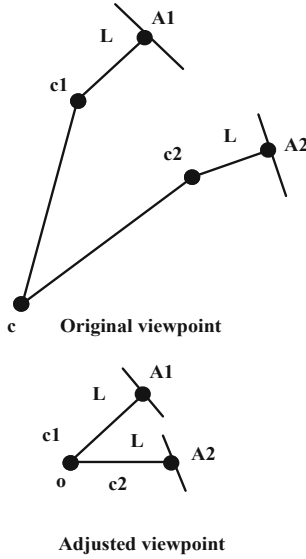


Fig. 2. Schematic diagram of image viewpoint correction process

In the formula: σ_i represents any point on the i th image and the three-dimensional coordinates in the camera coordinate system; y_i represents any point on the i image and the coordinates in the established three-dimensional coordinate system. i represents the camera coordinate system of the i image, compared to the rotation matrix of the three-dimensional coordinates. After the viewpoint is corrected, the mapping of the image plane to the spherical surface is calculated [9]. Assuming that the coordinate of any point i on the image is (x_0, y_0, z_0) and the coordinate of the mapping point c' is (x'_0, y'_0, z'_0) , according to the spherical three-dimensional European coordinate conversion, the stitching function of the panorama is obtained:

$$\begin{cases} x'_0 = \theta \sin \kappa \cdot \frac{y}{\theta} \cos \kappa \cdot \frac{x}{\theta} \\ y'_0 = -\theta \sin \kappa \cdot \frac{y}{\theta} \sin \kappa \cdot \frac{x}{\theta} \\ z'_0 = \theta \cos \kappa \cdot \frac{x}{\theta} \end{cases} \quad (13)$$

In the formula: κ is the positive angle between the plane projection vector and the horizontal axis in the 3D coordinate system; θ is the reverse angle. According to the above formula, the point cloud data of the image is obtained. Then the virtual 3D panoramic environment frame model is established, and the frame components are filled by using image simulation and rendering technology. However, the rendered 3D model does not have the characteristic points of the virtual object, so the surface of the designed model is required to be covered with texture. The process is divided into four stages. Firstly, the virtual item frame built and rendered is filled and rendered, and then the algorithm is used to draw the covering pattern on the surface of the object, and then other components with a single structure are filled, and the overall virtual item is plumped according to the frame structure. By analogy, according to the research content of the first two sections, a complete indoor scene panoramic image is rendered from a multi-viewpoint angle [10]. The key steps in the rendering algorithm are as follows:

```

{
    Private enum
    WavePosition/(Left = 1, Right = 2, Neutral = 3)/
    Private enum
    WaveGestureState/(Left = 1, Right = 2, Neutral = 3)
    Private struct WaveGestureTracker/
    Public int IterationCount;
    Public WaveGestureState State/
    Public Long Timestamp;
    Public WavePosition StartPosition;
    Public WavePosition CurrentPosition;
}

```

The above steps record the recognition status of multi view rendering, so as to realize the multi view rendering optimization of indoor scene based on binocular vision.

3 Experimental Study

In order to verify the effectiveness of this method, a simulation experiment is carried out. The proposed method is taken as group A, the method of literature [1] is taken as group B, the method of literature [2] is taken as group C, and the method of literature [3] is taken as Group D. under the same indoor scene test conditions, the differences between different methods are detected, and the test results of day and night are consistent, which will not affect the experimental results.

3.1 Experiment Preparation

The test platform is built. The computer model is rxt2060. The running memory of the computer is 8 GB, the hard disk capacity is 1t, and the video memory capacity is

6 GB, which meets the requirements of this experiment. Insert 3DMAX software and CAD software to facilitate rendering work. Randomly select an indoor building under construction as the experimental object, use landscape scanner to scan the indoor scene of the building in an all-round way, and establish the 3D indoor model of the building according to the scanning results. Table 2 below shows the basic parameters of the experimental scene.

Table 2. Basic parameters of the experimental scene.

Serial number	Information	Parameter
1	House type	Three rooms and two halls
2	South and north windows	2 group/2 group
3	East and West windows	2 group/2 group
4	Gross floor area	141.63 m ³
5	Horizontal height	3.28 m
6	Number of floors	22 layer
7	Interior decoration	Light color, dark color

Different methods are used to render indoor scenes from different viewpoints, and the lighting effects and rendering shading differences between different test groups are compared.

3.2 Light and Shadow Effect Test

Taking a clear day and night as the test background conditions, through different light environments, compare the light and shadow effects obtained by the four test groups before rendering, as shown in Fig. 3 below.

According to the test results in the figure, in the 24-h light shadow relationship acquisition, only the light shadow coefficient of experimental group A has a strong change, and the light shadow coefficient can reach more than 0.9, which shows that the proposed method captures a stronger light shadow relationship. Although the light and shadow coefficients of control group B, group C and group D were all within 0.8, they were not as good as experimental group A in general. It can be seen that the light and shadow relationship of the three test groups was not complete, and they focused more on the rainy weather with insufficient sunshine.

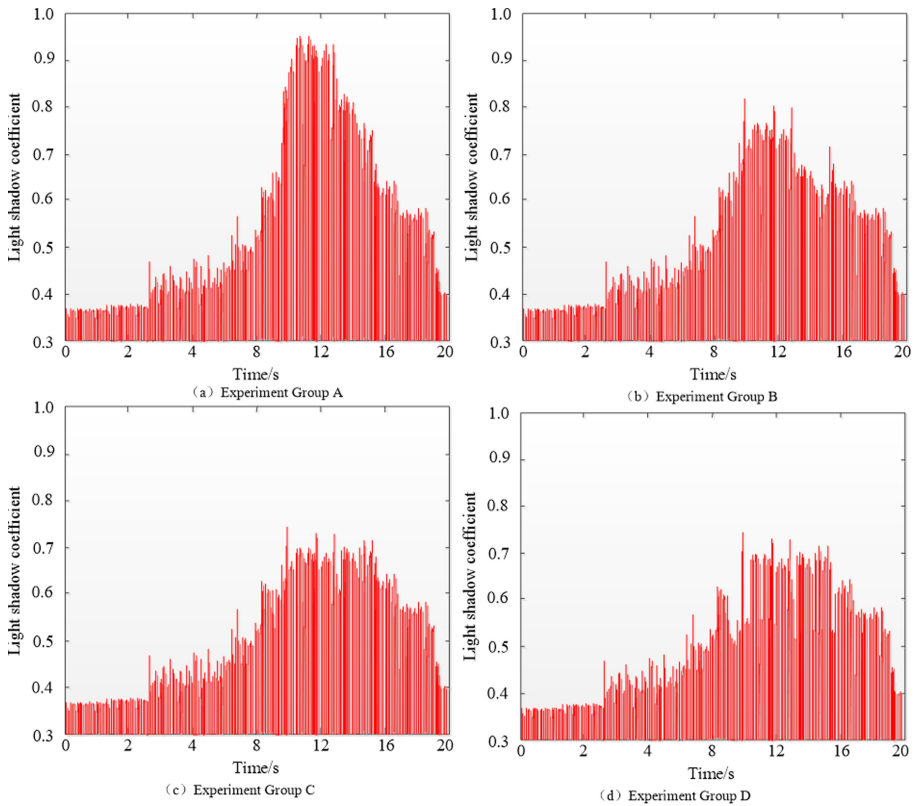


Fig. 3. Comparison of light and shadow effects test

3.3 Rendering Effect Test

The second stage is the rendering effect test. Due to the large differences in the test results of the first stage, it can be seen that the rendering capabilities of the four methods are different. The following Fig. 4 shows the rendering effect test results.

According to the comparison results of the rendering effect test in Fig. 4, it can be seen that the experimental group A has a good rendering effect of the indoor scene with a good light and shadow effect. However, the rendering of indoor scenes is too dim due to poor lighting and shadow effects in experimental group B and experimental group C. At the same time, the light and shadow effect of experimental group D is the worst, so the rendering effect is severely differentiated.

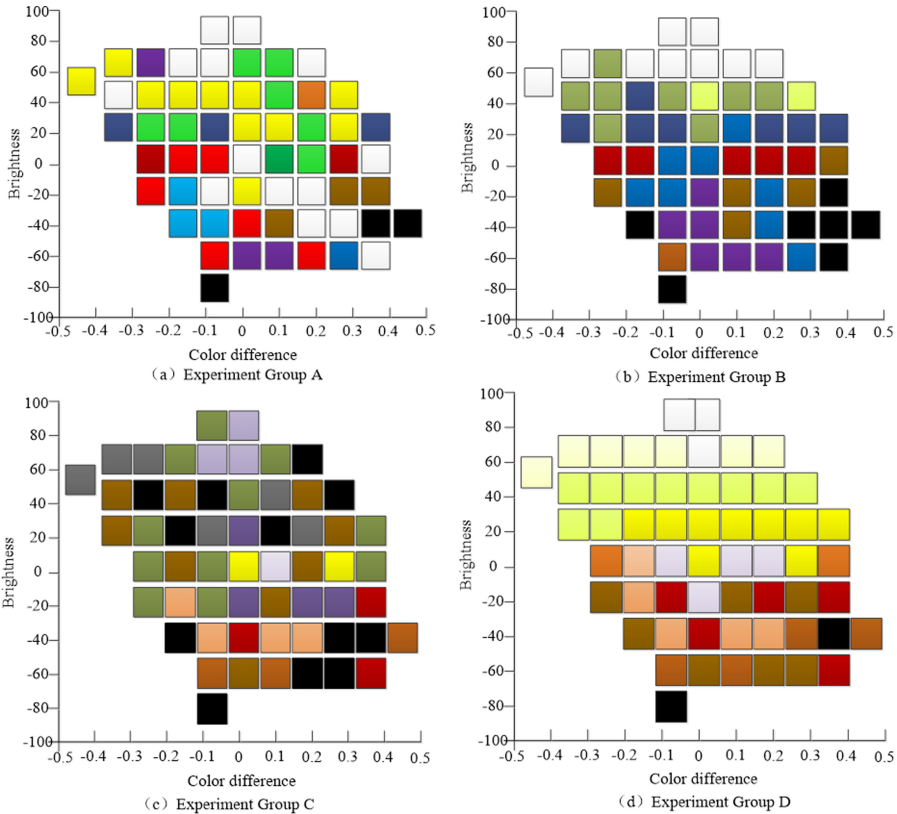


Fig. 4. Comparison of rendering effect test

4 Concluding Remarks

With the help of binocular vision, the optimization design has achieved good results. However, there are still two deficiencies in this study: one is that the calculation is too complex; the other is that the scale of the experiment is not expanded. Therefore, in the future research, we can design a simple algorithm to improve the efficiency of the rendering method, and expand the experiment to add persuasion to the test results.

References

1. Mengyu, J., Ligang, L.: Multi view rendering optimization of indoor scene based on perception. *J Univ. Sci. Technol. China* **48**(2), 140–147 (2018)
2. Ziyu, W., Yingmin, Z., Yongbin, C., et al.: Optimization of indoor scene semantic segmentation network based on rgb-d image. *Automation Information Eng* **2**, 27–32 (2020)
3. An, Z., Li, F., Wan, G., et al.: Construction method of indoor and outdoor scene based on BIM and GIS. *Digital design 1*, 2020, 009 (001):151–153.
4. Chen, X., Li, P.: Analysis of the light and shadow in the modern architecture's expression. *China Illuminating Eng. J* **29**(01), 82–86 (2018)

5. Fu, P., Chen, X., Wu, L.: Research on absolute positioning of binocular vision based on corner. *J. Electr. Measur. Instrument* **32**(03), 1–8 (2018)
6. Liu, Y., Wang, Z., Liu, Y., et al.: Reversing obstacle detection based on binocular vision image. *J. Chongqing Jiaotong Univ.* **37**(3), 92–98 (2018)
7. Sun, S., Sun, S.: Precision matching simulation of binocular stereo vision target. *Comput. Simul.* **35**(11), 413–416 (2018)
8. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intelligent Syst* (2020). <https://doi.org/10.1007/s40747-020-00161-4>
9. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks[J]. *Entropy* **21**(9), 902 (2019)
10. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images[J]. *IEEE Access* **7**, 62412–62420 (2019)