Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 12 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i12.49069

PAPER

Artificial Intelligence Technology for Interactive Mobile Devices and Its Application in 3D Visual Design

Chun Zhang(✉)

Hubei Institute of Fine Arts, Wuhan, China

zhangchun6321@163.com

ABSTRACT

The use of visual rationality in planning and designing indoor 3D spaces can effectively enhance the quality of environmental art design. For the rational planning and design of 3D spaces, it is essential to assign attribute values to the rational evaluation index of 3D space planning, determine the weight vector of rational evaluation factors for indoor 3D space planning, and accomplish the visual rationality planning and design of indoor 3D spaces. 3D reconstruction is one of the primary objectives and focal points of machine vision research. The principle is to extract 3D information from the 2D image captured by the machine's sensors in order to reconstruct a 3D scene model. The traditional method constructs the evaluation index system for indoor 3D space planning to determine the weight of the evaluation index for alternative points. However, it overlooks the weight vector of the evaluation factors, leading to low accuracy in rational planning. This paper optimizes the effect of structured light reconstruction using photometric stereo vision. Firstly, the text delves into the impact of light sources on the 3D reconstruction effect. Subsequently, a fusion algorithm combining 3D and structured light reconstruction methods is introduced, leveraging artificial intelligence (AI) technology. Experimental results show that this algorithm has a superior 3D reconstruction effect.

KEYWORDS

artificial intelligence, 3D visual design, 3D reconstruction

1 INTRODUCTION

With the development of the times, the improvement of industrial levels, and the introduction of new processes and designs, the complexity of products also increases in response to changing market demands. 3D reconstruction engineering is cutting-edge technology in the machine vision and computer vision industries. With the advancement of 2D technology, the structural complexity of many products is increasing daily. There are increasingly more irregular surface products on the market, and there is a need to achieve rapid detection [1]. Three-dimensional reconstruction is a method that allows individuals to comprehend the transition from the

Zhang, C. (2024). Artificial Intelligence Technology for Interactive Mobile Devices and Its Application in 3D Visual Design. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(12), pp. 30–41. https://doi.org/10.3991/ijim.v18i12.49069

Article submitted 2024-02-13. Revision uploaded 2024-03-21. Final acceptance 2024-03-21.

digital world's 2D plane to 3D space. In current detection equipment, the majority of devices utilize a 2D detection method when inspecting curved products. Ordinary light sources are commonly used to illuminate products, but they are only effective for objects with large surface textures. They cannot meet the requirements of high-precision detection, significantly impacting detection efficiency and leading to substantial errors [2]. In order to achieve accurate detection of products with irregular surfaces, 3D reconstruction engineering plays a crucial role within the current scientific and technological framework. As an important branch of 3D technology, accurately and quickly detecting various complex types of surface products is an urgent problem that needs to be addressed. Nowadays, a series of technological revolutions triggered by artificial intelligence are rapidly changing all aspects of life, introducing a plethora of scientific and technological advancements. In various fields, 3D scene reconstruction is crucial. In the process of reconstruction, 3D sensors such as lidar are usually used. However, using a binocular camera is a better option. Binocular cameras are typically a more cost-effective solution.

After decades of research by many scholars, the related problems of 3D reconstruction of objects are still very challenging, and the estimation of the depth of stereo image pairs remains an outstanding issue. Binocular stereo vision technology restores the 3D information of the scene, which involves the process of the binocular camera combined with the corresponding algorithm to obtain depth information [3]. Therefore, obtaining depth information efficiently and accurately to reconstruct the 3D structure of an object is of great significance. The performance of traditional stereo vision algorithms relies on the selection of the cost function, which defines the similarity between two pixels in a stereo image pair. The convolutional neural network learns the similarity of stereo images directly from the data. Although many scholars have made progress in this field, accurately finding stereo correspondence in stereo image pairs remains a significant challenge due to poor lighting conditions and the presence of occluded areas. These problems typically result in the inaccurate calculation of parallax at the object's edge, directly impacting the reconstruction results [4]. Therefore, it is crucial to calculate the depth of information as accurately as possible. In real life, applications in many fields rely on accurately finding stereographic correspondence, making it a subject of great research significance. On the basis of accurate depth information, 3D reconstruction of the image can achieve better results. There are still some unresolved problems and difficulties in this subject that require further efforts in future research. It is essential to conduct more in-depth exploration to achieve a more comprehensive reconstruction effect.

Three-dimensional digital content is the fundamental component of virtual simulation and hybrid reality. The core of creating 3D content is 3D geometric modeling, which is an important fundamental issue in computer graphics. Manual 3D modeling is challenging and laborious, relying heavily on the skills and experience of professionals. Untrained, ordinary users are often unable to perform it effectively [5]. How to enable the general public to conveniently and quickly create and edit 3D content, achieve popular modeling, overcome the bottleneck of 3D content generation, and drive the widespread growth of 3D data has always been a central objective and significant challenge in the field of graphics. Based on the existing theoretical knowledge of three-dimensional reconstruction of binocular vision, this paper focuses on developing a phase unwrapping algorithm based on multi-frequency heterodyne and improving the algorithm to adapt to more complex and dynamic scenes. Based on the purpose and application of the algorithm, a binocular structured light 3D reconstruction system is developed [6]. The modularization of the system enables rapid 3D inspection of curved products and swift commercial visual inspection. This paper optimizes the effect of structured light reconstruction using photometric stereo vision. Firstly, the text explores the influence of light sources on the effectiveness of 3D reconstruction. Subsequently, it proposes a fusion algorithm that combines 3D and structured light reconstruction methods, leveraging artificial intelligence (AI) technology. Experimental results show that this algorithm has a superior 3D reconstruction effect.

2 RELATED WORK

2.1 Advances in 3D vision

Binocular stereovision technology simulates the human visual system to perceive the real world in three dimensions. It achieves parallax by matching two images taken from different perspectives. Compared with active sensing technology, binocular stereo vision technology offers the advantages of simple equipment, low cost, and high efficiency [7]. Therefore, binocular stereo matching technology has been a prominent issue in the field of computer vision for decades and has made significant progress. Camera positioning is a fundamental issue in 3D computer vision. The task is to estimate the position and orientation of the camera in a coordinate system, specifically the camera's attitude, based on the image captured by the camera. The research framework for 3D vision is illustrated in Figure 1.

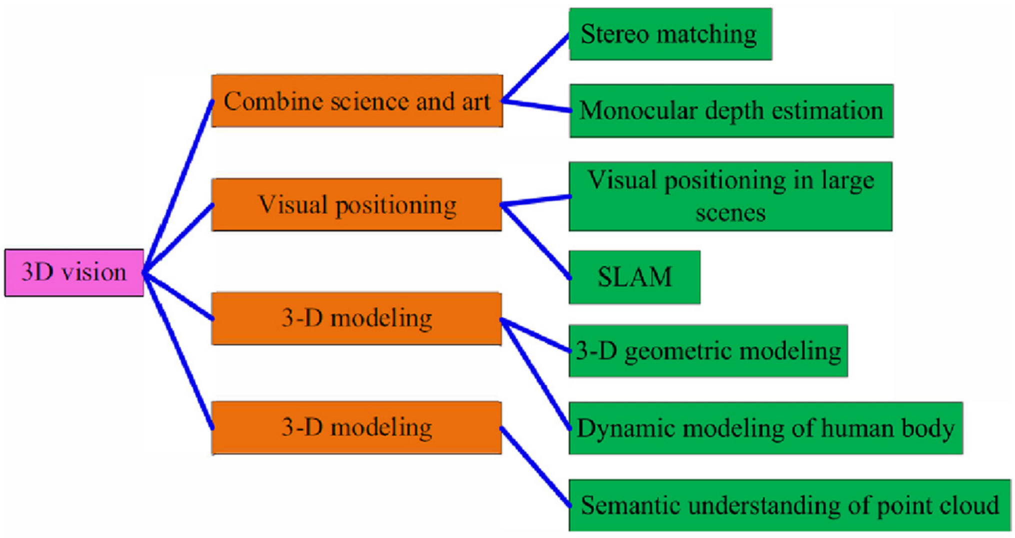

Fig. 1. The research framework for 3D vision

For non-end-to-end stereo matching algorithms, convolutional neural networks are usually employed to replace one or more components of the traditional matching algorithm. Researchers first use a convolutional neural network to calculate the matching cost. This deep twin network is composed of several convolutional layers and fully connected layers, which are used to calculate the similarity between two image blocks. This method achieves the best effect on the binocular data set of that year and proves that the image features extracted by the convolutional neural network are more accurate than the manually designed feature operators [8]. Inspired by this, a lot of stereo vision work uses convolutional neural networks to calculate the matching cost. It explores and proposes a large number of convolutional neural networks with different structures to represent the similarity measurement function between two image blocks, ultimately improving the final outcome. Although these methods have made significant progress in analyzing some binocular data sets, they often demand substantial computing resources and are time-consuming. Researchers approach the stereo matching problem as a multi-category classification issue. This allows the model to implicitly understand the distinctions between image blocks under various parallaxes by learning the probability distribution across all potential parallax values [9]. In these methods, some post-processing techniques are usually employed to further optimize the initial matching cost obtained by the neural networks. These techniques include cross-feature aggregation, semi-global matching, and left-right consistency detection.

In the traditional stereovision process, besides generating matching costs, a neural network can also replace the part. Based on the assumption of locally smoothing the disparity map, some methods incorporate smoothing constraints directly into the network learning process. Researchers propose a framework to predict the penalty cost in traditional algorithms. It proposes a new cost function that includes a path penalty term and an adjacent penalty term. This new cost function enables the neural network to fully utilize the sparse disparity map labels collected in practice, such as the sparse disparity map collected by lidar [10]. However, obtaining the tag value of the penalty cost requires complex processing, making the training process very intricate and time-consuming. Then the researchers proposed a hybrid model that combines a neural network and a conditional random field, and they added a smoothing penalty term. The model utilizes a three-stage network to replace the manually designed parallax optimization function. This network can detect incorrect parallax estimates, use new estimates to replace the incorrect ones, and then optimize the new parallax values again [11]. However, this step of detection, substitution, and optimization requires additional computing resources. Most stereo-matching techniques struggle to accurately capture visual disparities on reflective surfaces using only local constraints. Researchers use prior information about objects to address the uncertainty in estimating disparities on reflective surfaces. This method introduces a 3D model of the car as prior information. However, the introduction of 3D models significantly increases the computational load of the model.

2.2 Overview of 3D reconstruction technology

Three-dimensional reconstruction technology aims to represent the surface information of an object or scene using a 3D point cloud or 3D surface. Three-dimensional reconstruction is a crucial technology used to create visual models of three-dimensional objects on a computer. It involves calculating, manipulating, and analyzing three-dimensional objects within a computer system to represent the real world [12]. Three-dimensional reconstruction technology can effectively simulate the three-dimensional structure and appearance information of the target or scene in various fields such as 3D printing, medical diagnosis, surgery, large-scale scene reconstruction, remote sensing, global information systems, and remote conferences. It plays a prominent role in commercial applications such as 3D television, 3D games, 3D film creation, and industrial automated testing. There are two ways to achieve 3D reconstruction of objects: active and passive. The active modes include mechanical contact mode and radiation non-contact mode. The contact method can obtain the desired data directly and quickly, primarily relying on specific instruments and equipment [13]. The 3D data collected by the equipment through contact with the object has very high accuracy in restoring the truth. However, it is influenced by factors such as system time consumption, space constraints, human operation, and other variables. The application of contact 3D reconstruction is limited in scope. Non-contact methods involve detecting objects without physical contact and obtaining the desired data using magnetic fields, light, sound waves, and other media. This method has been widely used in various fields, and binocular vision technology enables non-contact detection.

The passive method mainly utilizes the camera and lens to capture images by reflecting the natural environment of the measured object, such as sunlight or industrial light sources. Subsequently, it calculates the 3D spatial information of the object using the necessary algorithm [14]. The main methods include the texture recovery shape method, the shadow recovery shape method, and the stereovision method. This paper presents a method for achieving 3D reconstruction based on stereovision. According to the simulation of the human visual system, based on the principle of binocular parallax imaging, the position difference of the corresponding points of the image is obtained, and the 3D information is restored. At present, many researchers have started utilizing coded-structured light methods and color structured light methods to conduct and achieve 3D surface measurements. Researchers have enhanced the existing practical coding methods, which can be categorized into temporal coding, spatial coding, and direct coding. In time coding, there is only a one-bit difference between adjacent code values of gray code [15]. Therefore, the one-bit error will occur at most, making it more stable than binary code, easy to encode, and widely used. However, it also has some shortcomings. For instance, the minimum number of fine stripes should be greater than eight pixels; otherwise, it will lead to gray code decoding issues, resulting in low resolution of the gray code detection method. The phase-shifting method is applied to small and continuous objects with high resolution. However, its disadvantage is that it involves a complex decoding process and is prone to errors. Foreign researchers have made significant advancements in key technologies for 3D surface detection. These include selecting appropriate coding methods, calibrating cameras and projectors, matching feature points, and stitching 3D point clouds. Many of these methods have been applied and validated in practical settings.

2.3 Research progress of 3D image reconstruction methods

Based on different measurement signal sources, non-contact measurement is divided into active measurement and passive measurement. Active measurement methods include structured light, laser ranging, signal interference, etc. Passive measurement methods include single-angle and multi-angle measurements [16]. Laser ranging utilizes the laser as the active light source, working in conjunction with the servo system to capture the movement displacement and measure the distance from the light source to the object surface. This process enables the 3D reconstruction of the surface. The laser focusing method utilizes the laser to consistently focus on a point and employs the principle of equal focal length to determine the relative depth of the object's surface by adjusting the servo system's position. Restricted by the servo system, this method is not suitable for large-scale scene reconstruction but is appropriate for small-scale precision measurement [17]. The principle of the time-of-flight method is simple: it measures half the distance that the laser travels to reach the object's surface and reflect back to the receiver, taking into account the constant speed of light. If the exact time of laser emission and reception is known, the duration of laser movement can be calculated. This duration can then be multiplied by the speed of light to determine the distance from the object to the laser source. This process helps in obtaining the distance from the emission source to the object surface and the depth of the object surface.

Before using photometric stereo, the light source calibration should be carried out first. The purpose of calibrating a light source is to determine its direction or intensity. Generally, one of the three types of information—light and shade information, shadow information, or reflection characteristics of the object surface—can be used, or the three kinds of information can be combined [18]. The near-field light source needs to correct the light intensity of the light source according to the parallel light model, and the light-emitting model of the actual light source needs to be considered, which increases the calibration complexity. Different calibrators can be combined and optimized by utilizing complementary information. There is a certain intersection between the problem of the near-field light source and the field of uncalibrated photometric stereo. Since the near-field light source is challenging to calibrate, uncalibrated photometric stereo can be utilized directly for reconstruction [19]. At the same time, the ideal photometric stereo method utilizes a point light source. However, the effectiveness of the point light source is constrained; if it is too far, the brightness will be too dark, and if it is too close, it will disrupt the distant light source model. Using a near-light source in the photometric stereo algorithm significantly raises the computational complexity. The most direct method of photometric stereo reconstruction is direct integration [20]. This method requires that the curl of the gradient field data be zero or that the gradient is integrable. However, due to the noise factor, the actual surface is not integrable, so restrictions need to be added to make the gradient integrable by necessity.

3 DESIGN OF APPLICATION MODEL

3.1 Visual rationality planning of three-dimensional space

In the process of planning and designing the visual rationality of indoor three-dimensional space, the geometric relationship between human visual space and virtual 3D space is established. The virtual 3D point element is mapped to the image point of the same name on the screen, and the vertical parallax under the standard stereoscopic mode is obtained. A set of design values that affect the rationality of 3D space planning are identified to form the rationality sequence of indoor 3D space planning, completing the planning and design of indoor 3D space visual rationality. Establishing the geometric relationship between human visual space and virtual 3D space, the mathematical expression is as follows:

$$W'_{qwe} = \frac{c'_{\mathrm{wer}} \mp p'_{c}}{(x'_{ko}, x'_{lo})} \oplus x'_{sf} \tag{1}$$

Then, utilize equation (2) to determine the indoor virtual 3D point that corresponds to the image point with the same name on the screen. The mathematical expression is as follows:

$$E'_{\mathrm{RTU}} = \frac{\{z'_{wep}, z'_{\mathrm{uep}}\}}{[k'_{\mathrm{que}}, k'_{\mathrm{eve}}]} k'_{dgh} \times \{\sigma'_{\mathrm{wer}} \oplus v'_{xvj}\} \tag{2}$$

Represent the horizontal projection and vertical projection of the image point on the imaging plane of the left and right eyes, respectively. Then, use equation (3) to calculate the vertical parallax under the standard stereoscopic mode condition. The mathematical expression is as follows:

$$A'_{\mathrm{we}} = \frac{\{l_a, l_g, l_h, l_s\}}{j'_{\mathrm{weer}} \mp x'_{\mathrm{uio}}} * c'_{\mathrm{cop}} \times g''_{\mathrm{erp}} \tag{3}$$

Then, a set of design values that influence the rationality of 3D space planning are determined, and the calculation expression is as follows:

$$e'_{af} = \frac{z'_{hjk} \times m''_{erp}}{\{v'_{sgp}\} \mp m'_{wer}} \oplus A'_{we} \tag{4}$$

The posterior probability distribution is used to establish the rationality sequence for indoor 3D space planning, and its mathematical expression is as follows:

$$E'_{zxk} = \frac{b''_{dhk} \mp I'_{spl}}{x'_{shjj}} \mp f'_{sgkk} \tag{5}$$

Based on the computer results, the planning and design of indoor 3D space can be achieved through visual rationality. To enhance the planning and design of indoor 3D spaces in visual rationality, it is essential to create an image sequence matching list, calculate the transformation parameters between the platform and the camera coordinate systems, minimize the geometric distance from points to the polar line in all perspectives within the indoor 3D space, extract vertical lines in the 3D space, and identify the 3D points within the indoor space. First, collect the indoor 3D space planning image, and its mathematical expression is as follows:

$$A'_{zbb} = \frac{E'_{qwe}}{\theta'_{dg} \times d'_{sui}} \pm \frac{b'_{hup} \mp \{c'_{uip}\} \pm g'_{fh}}{f'_{fj}} \tag{6}$$

Establish the corresponding list between image sequences, and the mathematical expression is as follows:

$$A'_{\mathrm{wep}} = \frac{\Gamma'_{\mathrm{wer}} \times m'_{ry}}{x'_{yup}} \mp d'_{rp} \tag{7}$$

Then, the transformation parameters between the platform and the camera coordinate systems are estimated, and the mathematical expression is as follows:

$$B'_{\mathrm{erp}} = \frac{m'_{opl} \otimes R}{e'_{tyu}} \oplus \frac{b'_{yui} \mp p'_{kp}}{v'_{iop}} \tag{8}$$

Minimize the geometric distance from the point to the line in the indoor three-dimensional space. The mathematical expression is as follows:

$$e''_{opu} = \frac{\{x''_j \times x''_i\} \mp \omega'_{ry}}{B'_{erp}} \mp v'_{opi} \tag{9}$$

Assuming that the common matching point group under the adjacent perspective represents the transformation matrix between the camera and the platform, the vertical line in 3D space is extracted, and the mathematical expression is as follows:

$$e'_{erp} = \frac{l'_{we} \times \{p'_{jjr} * p'_{jor}\}}{A'_{zbb} \cdot B'_{erp}} \times \{A'_{wep}\} \pm e'_{rtu} \tag{10}$$

Establish a matching list between image sequences, estimate the transformation parameters between the platform and the camera coordinate system, minimize the geometric distance from the point to the polar line under all viewing angles in the indoor 3D space, extract the vertical line of the 3D space, and provide the 3D points of the indoor space. This lays the foundation for the realization of the visual rationality planning and the design of the indoor 3D space.

3.2 Three-dimensional reconstruction

With the rapid development of deep learning technology, some scholars have attempted to apply deep learning methods to the field of photometric stereo, resulting in the development of a network known as a deep photometric stereo network. The deep learning method is characterized by autonomous learning and can eliminate the stringent assumptions of photometric stereo regarding the light source model and reflection model. The depth photometric stereo network is shown in Figure 2.

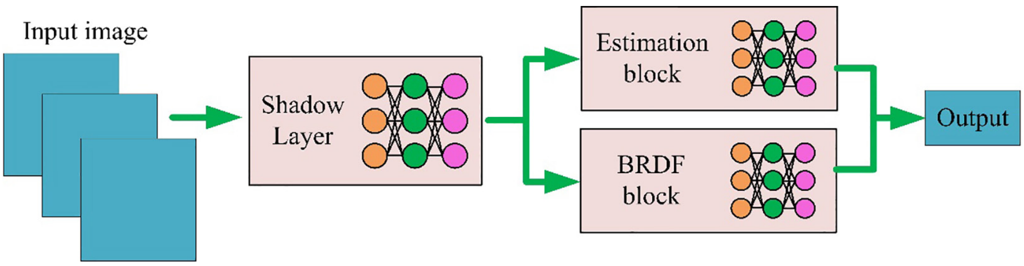

Fig. 2. The depth photometric stereo network

Theoretically, deep learning can be applied to every step of photometric stereo. It involves learning the original image to estimate the surface normal direction and light source information, as well as understanding the relationship between the normal vector and the reconstructed height. Currently, this is the most recent research focus. The optimal network structure for achieving the best results in photometric stereo is a challenging aspect to explore. Especially in industrial applications, darkrooms cannot be used due to the interference of natural light, and the brightness of point light sources is even lower. Therefore, the issue of application value lies in determining which light source can be used to adapt to factory applications without compromising the reconstruction quality. The imaging of a scene depends on many factors, among which the brightness of the scene is related to radiance and the brightness of the image is related to irradiance. Radiance is used to measure the energy emitted by a light source. After the transmission process is lost, the energy will decrease by the time it reaches the surface of the object. Irradiance is a measure of the amount of energy an object receives. The brightness of an image corresponds to the gray value of the image. Irradiance is also known as light intensity in the field of optics. The process flow for 3D reconstruction is illustrated in Figure 3.

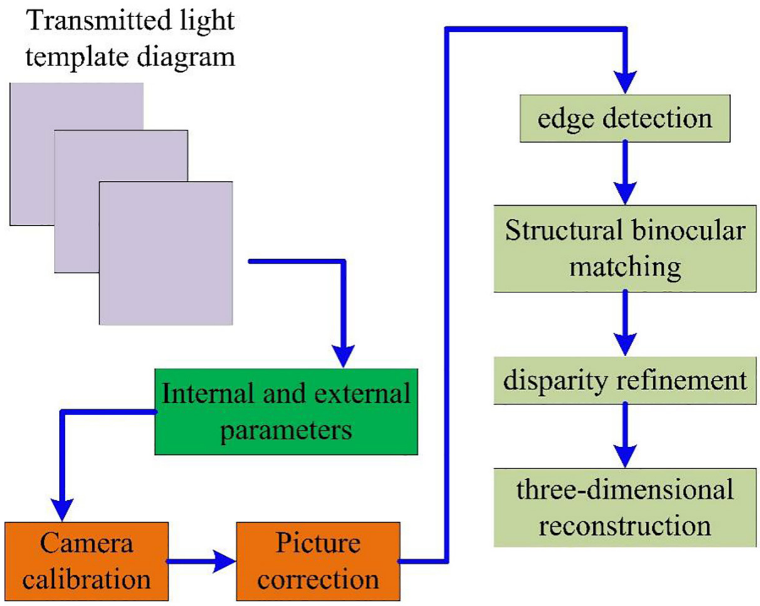

Fig. 3. The process flow of 3D reconstruction

At present, the combination of photometric stereo and other methods has become a trend. There are two main trends: the combination of active measurement and deep learning. Among the active measurement methods, structured light accounts for a relatively large proportion. There are various types of structured light, with the laser speckle method being less researched. Multispectral technology has the capability to offer multi-dimensional information simultaneously, thus holding research potential for enhancing measurement speed. The combination of multispectral technology and structured light technology is worth studying. The application of point light sources and strip light sources. In industrial applications, dark rooms cannot be used due to the interference of natural light, and the brightness of point light sources is insufficient. Therefore, under the premise of not affecting the reconstruction quality, determining the suitable light source for factory applications poses a significant challenge with high practical value.

4 EXPERIMENTS AND RESULTS

Photometric stereo obtains 3D information from the photometric data in the image. Therefore, theoretically, the photometric stereo experiment needs to be conducted in a dark environment. In engineering applications, constructing a dark room will raise the production costs of enterprises. Therefore, this section explores the impact of natural light on the photometric stereo reconstruction effect. In this experiment, we compare the direct reconstruction effect under natural light, the reconstruction effect after subtracting natural light from the original image, and the reconstruction effect without natural light in the darkroom. To prevent overexposure, it is advisable to avoid excessively strong natural light. The experimental data were obtained from the internal, non-public data of a 3D model design and application research institute in China. The experiments were conducted using the PyTorch open-source framework on the Windows 10 operating system and the NVIDIA 3080 GPU to train the network. The statistical results of the error analysis are shown in Figure 4.

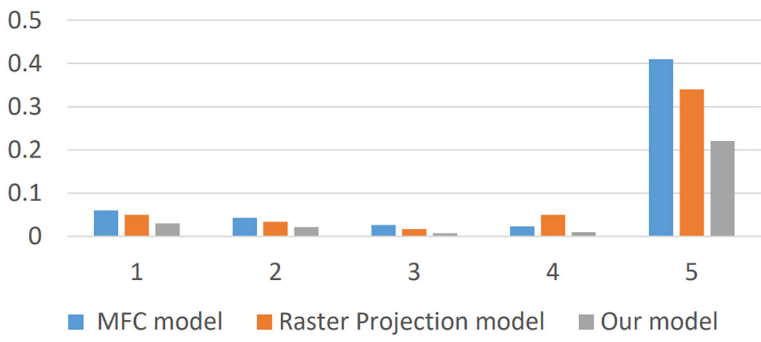

Fig. 4. The statistical results of error analysis

Compare the reflection map and normal vector map with and without natural light. When not heavily influenced by natural light, the reflectance map and normal vector map, obtained by eliminating natural light, exhibit clearer details and greater color contrast. This suggests that the gradient information has been enhanced and refined. Compared with the reconstruction rendering effect, it was observed that the reconstruction height of the experimental group with natural light interference is lower and smoother. In contrast, the reconstruction height of the experimental group with natural light interference removed is higher, and the surface fluctuation of the object is larger. This observation confirms the previous conclusion that the gradient information has been enhanced after eliminating natural light. System time statistics are shown in Figure 5.

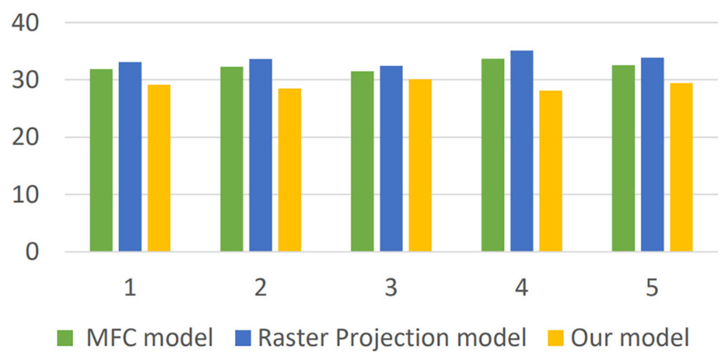

Fig. 5. System time statistics

Multispectral photometric stereo is a convenient method for capturing images by illuminating the object simultaneously with multiple spectral light sources. This approach allows for capturing the object's image in a single instance without the need for sequential lighting. The combination principle of multispectral photometric stereo and structured light is similar to that of white photometric stereo and structured light. The difference lies in the fact that the image sources used to solve the depth map in photometric stereo are distinct. White photometric stereo is acquired by directly illuminating the light source, while multi-spectral photometric stereo is obtained through channel separation technology.

5 CONCLUSION

Three-dimensional reconstruction technology involves using hardware and software architecture to reconstruct 3D spatial coordinates and achieve the visualization of multiple 2D images. Binocular vision is crucial for the development of 3D reconstruction. It involves converting target object information into digital signals using various hardware, processing these digital signals on a computer, and utilizing 3D image processing to reconstruct 3D images. This paper optimizes the structured light reconstruction effect based on photometric stereo vision. Firstly, the text explores the influence of light sources on the effectiveness of three-dimensional reconstruction. Subsequently, a fusion algorithm of 3D and structured light reconstruction methods is proposed, leveraging AI technology. Experimental results show that this algorithm yields superior 3D reconstruction outcomes.

This paper has yielded some research results, but there are still some shortcomings. The system can be expanded by using mark-point splicing or multi-angle splicing of the turntable. Calculate the rotation center of the turntable based on the points on the calibration plate obtained from various angles. Subsequently, apply rigid body transformation to each scanning result using the rotation angle and center to accurately register the point cloud. System running time and matching accuracy are still the key challenges in stereo matching. Finding a fast and efficient scheme is key to the stereo matching algorithm. In the future, we plan to utilize big data technology for implementing 3D visual design applications.

5.1 Funding

Hubei Institute of Fine Arts 2020 Major Research and Cultivation funding project, "Research on Pottery Teaching and Creation Methodology in Chinese Universities," project No. 20ZD09.

6 REFERENCES

[1] B. Tamadazte *et al.*, "CAD model-based tracking and 3D visual-based control for MEMS microassembly," *The International Journal of Robotics Research*, vol. 29, no. 11, pp. 1416–1434, 2010. https://doi.org/10.1177/0278364910376033

[2] B. Zou, Y. Liu, M. Guo, and Y. Wang, "EEG-based assessment of stereoscopic 3D visual fatigue caused by vergence-accommodation conflict," *Journal of Display Technology*, vol. 11, no. 12, pp. 1076–1083, 2015. https://doi.org/10.1109/JDT.2015.2451087

[3] Y. Yang, B. Li, P. Li, and Q. Liu, "A two-stage clustering-based 3D visual saliency model for dynamic scenarios," *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 809–820, 2018. https://doi.org/10.1109/TMM.2018.2867742

[4] M. Bachiller, J. A. Cerrada, and C. Cerrada, "A modular scheme for controller design and performance evaluation in 3D visual servoing," *Journal of Intelligent and Robotic Systems*, vol. 36, pp. 235–264, 2003. https://doi.org/10.1023/A:1023096511738

[5] E. Baek *et al.*, "An exploratory study on visual merchandising of an apparel store utilizing 3D technology," *Journal of Global Fashion Marketing*, vol. 6, no. 1, pp. 33–46, 2015. https://doi.org/10.1080/20932685.2014.971502

[6] S. Johnson *et al.*, "Artifact-based rendering: Harnessing natural and traditional visual media for more expressive and engaging 3D visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 492–502, 2019. https://doi.org/10.1109/TVCG.2019.2934260

[7] R. Richa, A. P. L. Bó, and P. Poignet, "Towards robust 3D visual tracking for motion compensation in beating heart surgery," *Medical Image Analysis*, vol. 15, no. 3, pp. 302–315, 2011. https://doi.org/10.1016/j.media.2010.12.002

[8] W. Song, Y. Fujia, and M. Minami, "3D visual servoing by feedforward evolutionary recognition," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 4, no. 4, pp. 739–755, 2010. https://doi.org/10.1299/jamdsm.4.739
[9] K. Lee, A. K. Moorthy, S. Lee, and A. C. Bovik, "3D visual activity assessment based on natural scene statistics," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 450–465, 2013. https://doi.org/10.1109/TIP.2013.2290592
[10] J. Wang, M. P. Da Silva, P. Le Callet, and V. Ricordel, "Computational model of stereoscopic 3D visual saliency," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2151–2165, 2013. https://doi.org/10.1109/TIP.2013.2246176
[11] Y. Han, "2D-to-3D visual human motion converting system for home optical motion capture tool and 3-D smart TV," *IEEE Systems Journal*, vol. 9, no. 1, pp. 131–140, 2014. https://doi.org/10.1109/JSYST.2014.2322253
[12] R. Opitz, "An experiment in using visual attention metrics to think about experience and design choices in past places," *Journal of Archaeological Method and Theory*, vol. 24, no. 4, pp. 1203–1226, 2017. https://doi.org/10.1007/s10816-016-9310-2
[13] F. Qi *et al*., "3D visual saliency detection model with generated disparity map," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 3087–3103, 2017. https://doi.org/10.1007/s11042-015-3229-6
[14] M. Zhao *et al*., "Orbital collaborative learning in 6G space-air-ground integrated networks," *Neurocomputing*, vol. 497, pp. 94–109, 2022. https://doi.org/10.1016/j.neucom.2022.04.098
[15] C. Chen *et al*., "An intelligent caching strategy considering time-space characteristics in vehicular named data networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19655–19667, 2021. https://doi.org/10.1109/TITS.2021.3128012
[16] C. Chen *et al*., "An edge intelligence empowered flooding process prediction using Internet of things in smart city," *Journal of Parallel and Distributed Computing*, vol. 165, pp. 66–78, 2022. https://doi.org/10.1016/j.jpdc.2022.03.010
[17] N. Yu, Y. Zhai, Y. Yuan, and Z. Wang, "A bionic robot navigation algorithm based on cognitive mechanism of hippocampus," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1640–1652, 2019. https://doi.org/10.1109/TASE.2019.2909638
[18] Q. Jiang *et al*., "Leveraging visual attention and neural activity for stereoscopic 3D visual comfort assessment," *Multimedia Tools and Applications*, vol. 76, pp. 9405–9425, 2017. https://doi.org/10.1007/s11042-016-3548-2
[19] M. A. Garratt, A. J. Lambert, and H. Teimoori, "Design of a 3D snapshot based visual flight control system using a single camera in hover," *Autonomous Robots*, vol. 34, pp. 19–34, 2013. https://doi.org/10.1007/s10514-012-9310-3
[20] A. Dan and M. Reiner, "Reduced mental load in learning a motor visual task with virtual 3D method," *Journal of Computer Assisted Learning*, vol. 34, no. 1, pp. 84–93, 2018. https://doi.org/10.1111/jcal.12216

7 AUTHOR

Chun Zhang, Hubei Institute of Fine Arts, Wuhan 430200, China (E-mail: zhangchun6321@163.com).