Gesture recognition based on binocular vision

Du Jiang [1][*], **Zujia Zheng** [1], **Gongfa Li** [1,3], **Ying Sun** [2], **Jianyi Kong** [1,2], **Guozhang Jiang** [1][2], **Hegen Xiong** [2], **Bo Tao** [2], **Shuang Xu** [1], **Hui Yu** [4], **Honghai Liu** [4] **and Zhaojie Ju** [4].

[1] Key Laboratory of Metallurgical Equipment and Control Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081 ,China

[2] Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

[3] Research Center of Biologic Manipulator and Intelligent Measurement and Control, Wuhan University of Science and Technology, Wuhan 430081, China

[4] School of Computing, University of Portsmouth PO1 3HE, UK

Abstract: A convenient and effective binocular vision system is set up. Gesture information can be accurately extract from the complex environment with the system. The template calibration method is used to calibrate the binocular camera and the parameters of the camera are accurately obtained. In the phase of stereo matching, the BM algorithm is used to quickly and accurately match the images of the left and right cameras to get the parallax of the measured gesture. Combined with triangulation principle, resulting in a more dense depth map. Finally, the depth information is remapped to the original color image to realize three-dimensional reconstruction and three-dimensional cloud image generation. According to the cloud image information, it can be judged that the binocular vision system can effectively segment the gesture from the complex background.

Keywords: Binocular vision; Gesture recognition; Gesture segmentation; Template calibration method

1. Introduction

With the change of lifestyles and the increasing popularization of intelligent devices, the demand for enhancing the existing human-computer interaction experience is particularly urgent. Among them, gesture information can provide users with more realistic and natural interactive experiences. Currently used in gesture recognition devices are data gloves, somatosensory motion sensor, EMG signal acquisition device, monocular camera, binocular camera.

The study of gestures or gesture recognition requires system construction [1]. Due to the immaturity of early hardware and algorithms, the recognition of gestures began with data gloves. In 1989, Surman identified gestures and used them to manipulate virtual objects with data gloves [2]. In 2001. In.cheol and Sung.il enabled glove-based gesture recognition with an accuracy rate of 96.88% [3]. Noor Tubaiz et al used data gloves to achieve a continuous recognition of Arabic sign language, with an accuracy rate of 98.9% [4]. The 16-channel sEMG acquisition system developed by Yinfeng Fang et al can accurately recognize the gesture and the system can be well integrated with the prosthetic device [5]. In practical applications, although the wearable device can accurately recognize gestures, there is a great limitation. Data gloves and EMG collectors are composed of a large number of sensors which are too expensive to be accepted by the average user. At the same time, because the data gloves are in direct contact with the hand and personal differences, each experiment needs to be calibrated so that experiment process is complicated. Last but not least, Data-based gesture recognition leads to a poor human-computer interaction experience, which are factors that limit its use [6]. Because of the development of computer performance and the improvement of the hardware level, vision-based gesture recognition has drawn the attention of researchers. Relative to the data glove, vision-based gesture recognition device is simple, easy maintenance. Hand gesture recognition based on monocular vision has become a common gesture

recognition method. In 2000, Tare et al proposed a random organization mapping algorithm to track and identify the gesture, which effectively reduced the computation time and improved the recognition accuracy [7]. By recognizing the texture features of the key parts of the human hand, Bhuyan et al can recognize the bending motion of the finger by using a monocular camera [8-12]. In the process of using the monocular camera for gesture recognition, the method relies more on image processing algorithms and priori Knowledge. The accuracy of gesture recognition and recognition speed was further improved with ameliorating the sparse algorithm [13-14]. Relative to the data glove, monocular vision-based gesture recognition hardware device is cheap, easy to use and more user-accepted. However, monocular vision is not stable due to hardware limitations and is not suitable for use in complex environments.

The new stereo vision algorithms continue to emerge. Based on binocular vision, the image information obtained by the binocular vision is finer and the viewing angle is wider. Compared with the monocular vision application scene, the images are more flexible and the gestures recognized are more diverse. To realize the vision-based hand recognition naturally, segmentation of gestures in complex environment is a big challenge. Dishi Chen et al proposed an interactive gesture segmentation method that can gesture segmentation in a more complex environment [15]. At the same time, the combination of specific sensors and cameras can further improve the effect of gesture segmentation [16-18]. The use of binocular vision not only capture the image information of the gesture, but also obtain the depth of the target object to help identify information, thus separating the gesture from the complex background. The construction of binocular vision platform can also help effectively plan gesture in virtual space [19]. The wide application of support vector machines and neural networks has also made a great improvement in the accuracy of gesture recognition based on binocular vision [20]. Using depth information, the dynamic transformation process of human hands in three-dimensional space can be accurately located, which greatly enriches the content of gesture operation [21].

At this stage, the application of stereoscopic vision is also multifaceted. The technology is used to measure and locate the three-dimensional dimensions of tiny objects [22-23], face recognition [24-26], object recognition [27-29] and so on. In specific applications, some sensors are combined to improve the accuracy of recognition [30-31]. The use of structured light can make the details of the identified object clearer and reduce the interference of the outside world [32-33].

At present, the gesture recognition based on binocular vision is often realized with professional industrial cameras, sophisticated depth cameras and infrared cameras, which further improves the application cost of the method. Therefore, the use of ordinary cameras to build a binocular vision system to achieve gesture recognition in complex environments is particularly important.

2. Binocular vision system

Binocular vision system compared to monocular or multi-purpose vision system in principle, in line with the principle of bionics, more scientific, flexible and easy to implement. At the same time, passive compared to the initiative does not require additional auxiliary equipment, wider range of applications, lower equipment costs. In this paper, passive binocular vision ranging principle to build a binocular vision system.
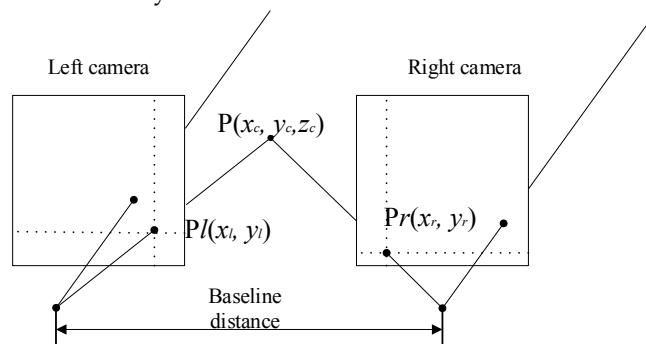
Figure 1 is an imaging principle of two cameras placed in parallel. The distance between the optical axes of two cameras is also an important parameter in binocular vision, the baseline distance. The two cameras simultaneously observe the spatial coordinates of the same feature points P at different positions respectively to obtain the image coordinates of the points P, that is, P_l and P_r. The image coordinates of the two images are $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$. Assuming that the images of the two cameras are on the same plane, so $y_l = y_r = y_u = y$. The expression can be obtained from the geometric relationship.

$$\left\{ \begin{array}{l} x_l = f \dfrac{x_c}{z_c} \\[2mm] x_r = f \dfrac{(x_c - B)}{z_c} \\[2mm] y = f \dfrac{y_c}{z_c} \end{array} \right. \tag{1}$$

In that way, $Disparity = x_l - x_r$. The coordinates of the feature point P under the camera coordinates:

$$\left\{ \begin{array}{l} x_c = \dfrac{B \bullet x_l}{Disparity} \\[2mm] y_c = \dfrac{B \bullet y}{Disparity} \\[2mm] z_c = \dfrac{B \bullet f}{Disparity} \end{array} \right. \tag{2}$$

Therefore, it is possible to precisely match the coordinates of two points in the image to determine the 3D coordinates of the points in which z_c is the camera from the measured object distance. It can be seen from the principle diagram that it is necessary to know the focal length f of the camera and the baseline distance B for obtaining the three-dimensional information. In the process of actual image acquisition, since the imaging principle of the camera is not a real pinhole imaging model, the captured images are not suitable for subsequent stereo matching. Therefore, an external influence factor is needed to adjust the image to meet the requirements [34-36]. This influence factor is also called the distortion coefficient.
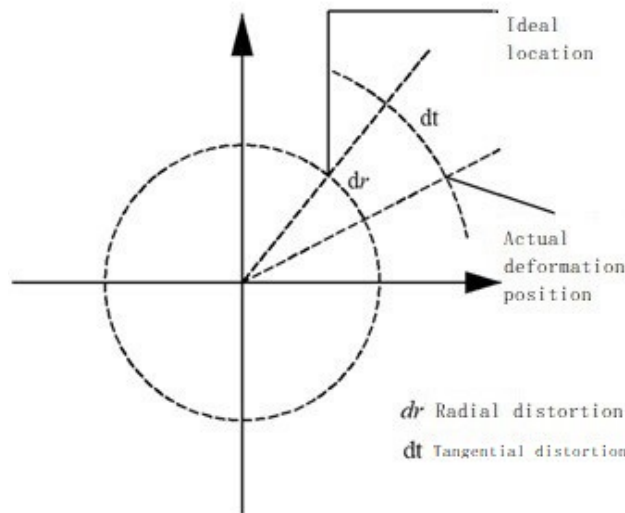

Figure 2. Radial distortion and tangential distortion

Fig.2 reflects the center of the lens distortion is almost zero, the more obvious radial distortion of the edge, the mathematical expression is:

$$\begin{cases} x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \tag{3}$$

(x, y) is the initial position before correction, $(x_{corrected}, y_{corrected})$ is the corrected point. Tangential distortion:

$$\begin{cases} x_{corrected} = x + [2p_1 + p_2(r^2 + 2x^2)] \\ y_{corrected} = y[p_1(r^2 + 2y^2) + 2p_2 x] \end{cases} \tag{4}$$

From (3) and (4), we can get the main distortion parameters of the camera as k_1, k_2, p_1, p_2, k_a. Tab.1 shows the relevant parameters in the camera model.

Table.1 Camera model parameters.

Parameters	Expression	Degree of freedom
Perspective transformation	$A = \begin{bmatrix} a_x & \gamma & u_0 \\ 0 & a_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$	5
Main distortion parameters	k_1, k_2, p_1, p_2, k_a	4
External parameters	$A = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_{91} \end{bmatrix} \quad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$	6

3. Camera calibration

Fig.3 shows a binocular vision system built with two normal web cameras. According to Table 1, to use the binocular vision system for accurate extraction of three-dimensional information of the measuring hand, the camera needs to be calibrated. The model of these cameras are Logitech c270.
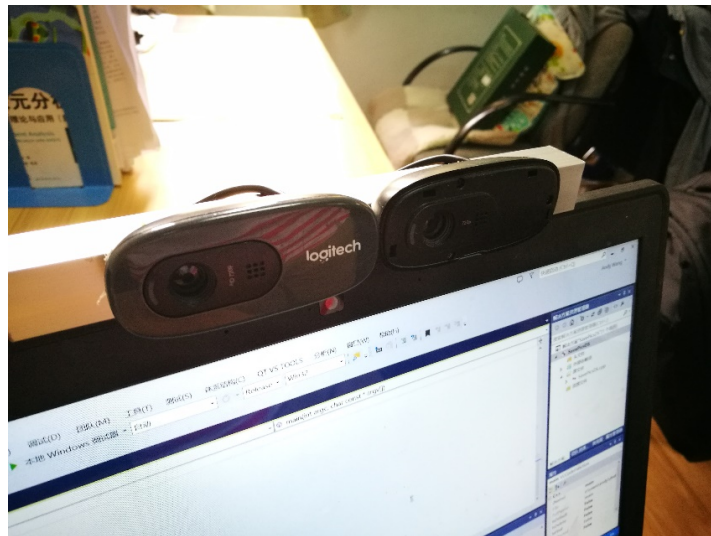

Figure 3. Binocular camera

During the camera calibration process, the camera's internal parameters and external parameters are acquired. The camera parameters are important links in establishing the relationship between the image coordinate system and the world coordinate. Currently used calibration methods are both traditional and self-calibration. In the traditional method, the direct linear transformation (DLT) method is the first time that Abdel-Aziz and Karara directly use the camera model to list a series of linear equations [37-38]. Then get the camera parameters by solving the equation. The method assumes that the camera is in an ideal condition and needs some nonlinear optimization algorithms to make the result more accurate. The perspective transformation matrix method means that the parameters of the camera can be represented by a perspective transformation matrix without considering the influence of distortion parameters. Then, Directly solve the matrix to get the desired result; Tsai's proposed calibration-Tsai two-step method based on Radial Alignment Constraint (RAC) [39], the core of which is to solve the linear Equation to solve the camera's external parameters, and then get the camera's internal parameters, the calibration accuracy of the method is more suitable for accurate measurement, but the experimental conditions are not suitable for a simple visual calibration.

Due to its inherent limitations of traditional calibration methods, Faugeras wanted to calibrate the camera through the surrounding environment combined with prior knowledge in the 1990s [40], which is also called self-calibration. At present, the commonly used camera self-calibration method has basic matrix and intrinsic matrix, based on Kruppa equation, gradual and step-by-step calibration method, and camera calibration method based on active vision. The basic matrix method is to calculate the basic matrix according to the one-to-one correspondence between the object points in the left and right camera images and then to solve the camera's related parameters. Based on the Kruppa equation method [41], the core is that every two images have two Constraints on nonlinear Kruppa equations, Kruppa equations on the polar transformation and image link to get all the images on the image to the corresponding pole distance and LM algorithm for the minimum, you can calculate the camera-related parameters. Compared with the gradual classification, it has more advantages, but its stability is not good. Template- calibration method based checkerboard is simple and high precision, it is often used to calibrate the stereo camera [42-44]. In the template calibration method, some points that can be evenly distributed across the whole image are generally selected. In general, the calibration accuracy of the camera will be higher if the points is more [45-47]. However, as the number of points increases, the amount of operation will increase or even lead to the reduction of calibration accuracy. In the process of camera calibration, we select the feature points of 8*6, which can ensure the accuracy. In the template calibration method, the template has great influence on the calibration accuracy of the camera. Many researchers will choose the industrial high-precision calibration template or dot template, but the price of these templates is very expensive. What we use is a common checkerboard template that uses a square cell of 25*25mm.

Template calibration method is between the two categories which is easier and meets requirement. In this paper, the cameras' internal parameters and external parameters were accurately captured with the template calibration method.
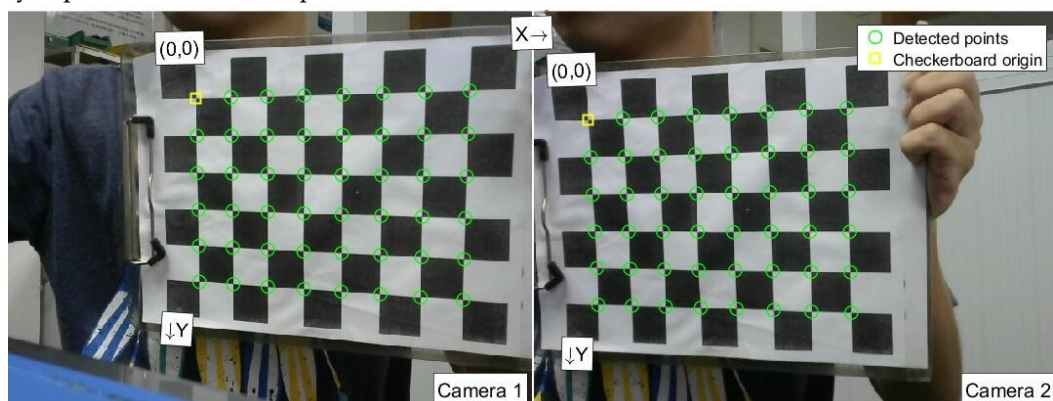

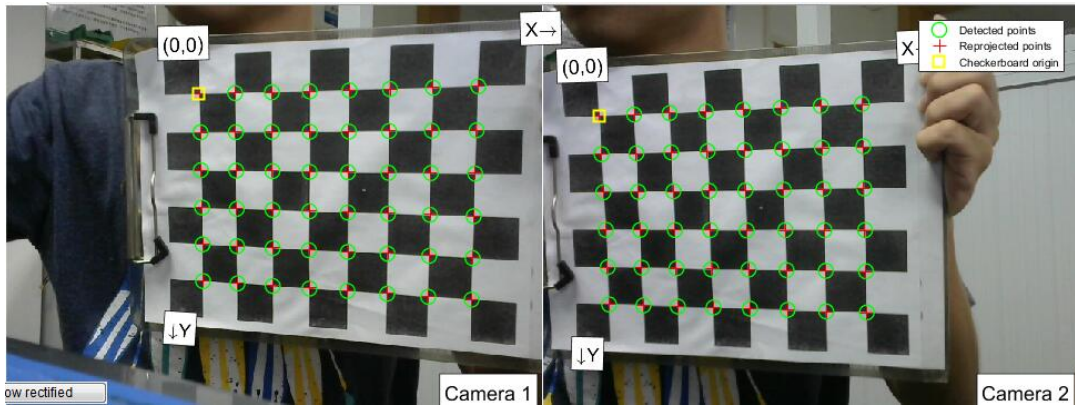**Figure 4.** Initially determine the corner extraction area

Figure 5. Accurately extract corner points

Camera calibration results:

1) Main calibration result of left camera:

focal length: fc = [817.85837   818.74862] ± [3.75478   3.71953]

The main point: cc = [320.26368   246.68164] ± [3.38336   3.26967]

Distortion parameters: kc = [0.03470   0.28927   0.00188   0.00548   0.00000]

               ± [0.01806   0.09767   0.00170   0.00200   0.00000]

Pixel error: err = [0.6652   0.6511]

2) The main camera calibration results:

focal length: fc = [816.58746   815.40532] ± [3.42024   3.44325]

The main point: cc = [313.65229   235.41916] ± [3.10666   3.41942]

Distortion parameters: kc = [0.03265   0.31320   0.00330   0.00259   0.00000]

              ±[0.01774   0.16803   0.00171   0.00159   0.00000]

Pixel error: err = [0.6816   0.6817]

3) Camera external reference calibration result

Rotation vector: om = [-0.07673   -0.04819   1.47491]

Translation vector: T = [20.31733   -53.32342   31.54584]

In the calibration result of the camera, the focal distance is **fc** and the main point coordinates is **cc**. The array **kc** is the internal parameter of the camera, and the 5 parameters in the array correspond to the internal parameters listed in Table 1. The vector **om** and **T** are the external parameters of the camera, which are the relative rotation and translation of the camera. Relevant parameters of the camera are obtained through calibration more accurately, and the error is within an acceptable range.
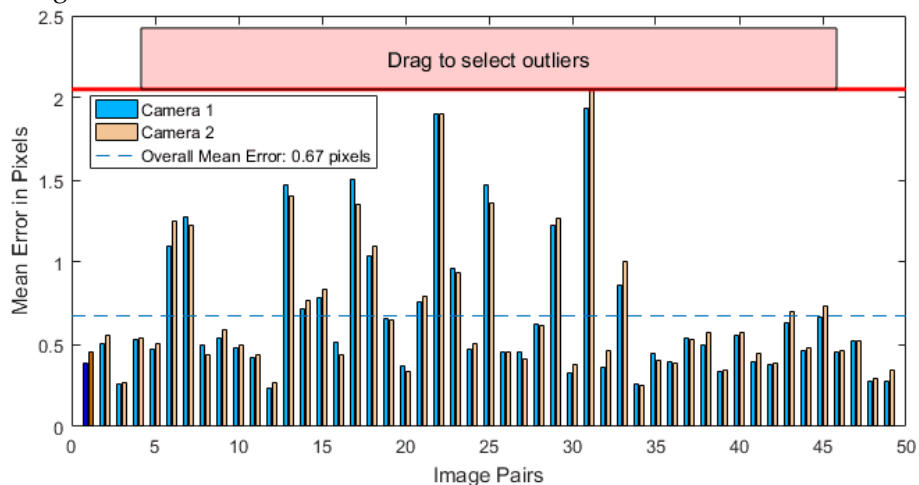

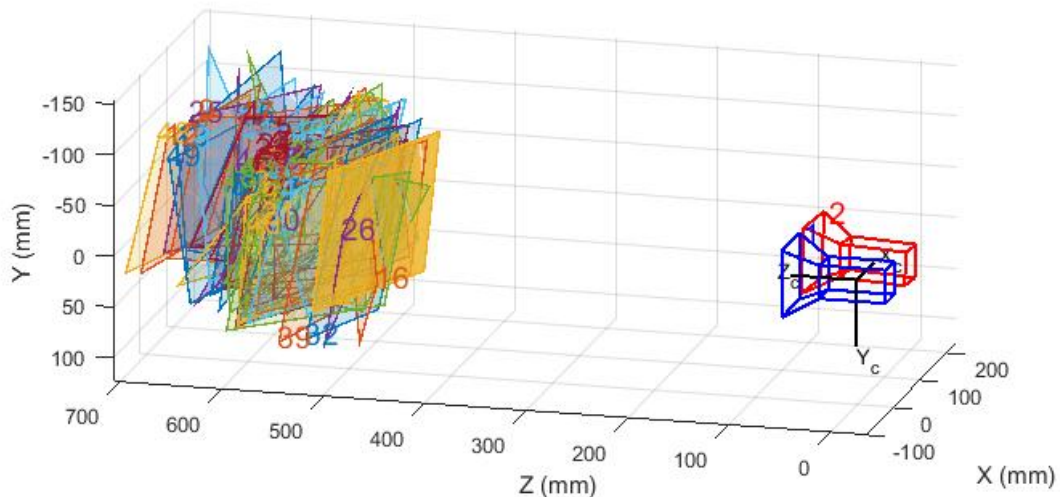**Figure 6.** Camera error analysis histogram

Figure 7. The relationship between the calibration plate and the camera

As can be seen from the histogram of Fig.6, the overall average error is 0.67 pixels and most of the sample errors are below the overall average error line. Figure 7 can be determined by the calibration plate calculated position and the actual position, the calculation result is more accurate.

4. Gesture three-dimensional information extraction

The key technology to realize the gesture three-dimensional information extraction is stereo matching technology, which is also the most challenging part [48]. The stereo matching technology divides the left and right images into small units with characteristic attributes, and then compares the key information of the two images one by one through the characteristic attributes of the small units. Selecting the appropriate matching features and matching criteria is an important guarantee for the accuracy of measurement results. Common matching features include point features, line features, and area features [49-50].

There are many matching algorithms, but none of them can be adapted to all environments. The current common matching algorithms include dynamic programming, image segmentation, regional matching and so on. The matching algorithm is divided into two main categories based on local constraint algorithm and global constraint algorithm. The accuracy of the global constraint algorithm is higher, but the complexity of the algorithm is high and the computational complexity is large, and it does not meet the needs of building this double vision platform. The local constraint algorithm is also divided into the region matching algorithm, the feature matching algorithm and the phase matching algorithm. The feature matching algorithm is preprocessed by the image, and the feature classification is realized, and then the left and right images are matched. Although the matching precision is increased, it needs more complex preprocessing, and the real time is poor. The precision of the phase matching algorithm is the highest, but the problem of phase singularity may occur. The method to solve the phase singularity is more complicated and difficult to realize. Therefore, in order to make the set of binocular vision platform easier to implement and have high real time, we choose the region matching algorithm. The area matching algorithm (BM) is used for stereo matching which has the advantages of high precision, fast speed and easy implementation [51-52]. Using the constructed binocular vision platform, three-dimensional information is extracted from the gesture shown in Fig.8. In view of the building of the binocular vision platform, the main use scene is indoor, so the illumination conditions of the experiment used ordinary fluorescent lamp. And the distance between the hand and the platform is 1m-2m. The scene of gesture collection is a general indoor scene. It doesn't need special treatment to be a solid background. So indoor use of the platform can be more free and flexible without background interference.
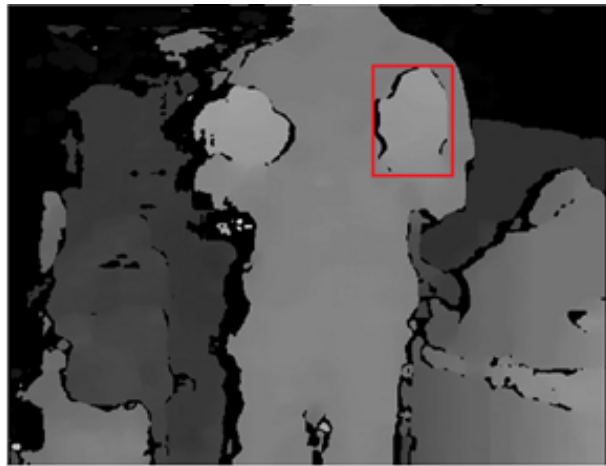
Figure 8. Hand three-dimensional

Figure 9. Gesture depth map information extraction

Figure 10. Gesture 3D cloud illustration

Figure 11. Gestures three-dimensional cloud map part

In extracting the three-dimensional information of the gesture, it can be seen from Fig.9 that the binocular vision system can generate a more dense depth map, but it is difficult to clearly distinguish the depth map from the depth map due to the closer manpower to the body. With 3D reconstruction, the depth map can be transformed into a three-dimensional cloud (Fig.10) where it is easier to see that the gesture information is well extracted (Fig.11). In order to further validate that the proposed binocular vision system can effectively extract the gesture from a complex background, replaced with a pinch gesture (Fig.12) to re-experiment, from the generated three-dimensional cloud map (Fig.13) can verify The built-in binocular vision system makes good use of depth information to distinguish the gesture from the background, and can accurately obtain the hand information (Fig.14).
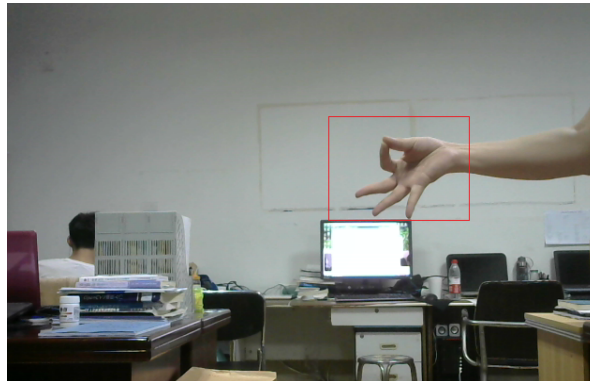

Figure 12. Two finger pinch gesture

Figure 13. Two finger pinch gesture

Figure 14. Pinch gestures three-dimensional cloud part

5. Conclusion

Accurately extracting the human hand's depth information is the premise of realizing the visual gesture recognition. In this paper, we use the binocular vision test technology to extract the image depth information and verify the accuracy of the extracted depth information by reconstructing the gesture image. In this paper, two ordinary network cameras to build a binocular camera, and camera calibration. Then, a fast and effective BM stereo matching algorithm is used to accurately extract the three-dimensional hand information and effectively separate the target object from the complex background. We implemented the binocular vision platform with simple and effective way, and applied it to gesture recognition, which further improved the efficiency of gesture recognition in complex environment. However, in order to further improve the accuracy and adaptability of 3D information, it is necessary to further optimize the related matching algorithm.

References

1. Al-Helali Baligh M. and Sabri A. Mahmoud. Arabic Online Handwriting Recognition (AOHR): A Survey. ACM Computing Surveys (CSUR), 2017, 50(3): 33

2. Sturman David J., David Zeltzer, and Steve Pieper. Hands-on interaction with virtual environments. Proceedings of the 2nd annual ACM SIGGRAPH symposium on User interface software and technology. ACM, 1989: 19-24

3. Kim In-Cheol and Sung-Il Chien. Analysis of 3d hand trajectory gestures using stroke-based composite hidden markov models. Applied Intelligence 2001, 15(2): 131-143

4. Tubaiz Noor, Tamer Shanableh, and Khaled Assaleh. Glove-based continuous Arabic sign language recognition in user-dependent mode. IEEE Transactions on Human-Machine Systems, 2015, 45(4): 526-533

5. Yinfeng Fang, Honghai Liu, Gongfa Li, Xiangyang Zhu. A multichannel surface emg system for hand motion recognition. International Journal of Humanoid Robotics, 2015, 12(02). doi:10.1142/S0219843615500115

6. Kerber Frederic, Michael Puhl, and Antonio Krüger. User-independent real-time hand gesture recognition based on surface electromyography. Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, 2017: 36

7. Tarek El. Tobely, Yuichiro Yoshiki, Ryuichi Tsuda, Naoyuki Tsuruta and Makoto Amamiy Dynamic hand gesture recognition based on randomized self-organizing map algorithm. International Conference on Algorithmic Learning Theory. Springer, Berlin, Heidelberg, 2000: 252-263

8. M.K. Bhuyan, Karl F. MacDorman, Mithun Kumar Kar, Debanga Raj Neog, Brian C. Lovell and Prathik Gadde. Hand pose recognition from monocular images by geometrical and texture analysis. Journal of Visual Languages &

Computing, 2015, 28: 39-55.

9. Qian Yin, Gongfa Li, Jianguo Zhu. Research on the method of step feature extraction for EOD robot based on 2d laser radar. Discrete and continuous dynamical systems-series s, 2015, 8(6): 1415-1421

10. Zhe Li, Gongfa Li, Guozhang Jiang, Yinfeng Fang, Zhaojie Ju, Honghai Liu. Intelligent Computation of grasping and manipulation for multi-fingered robotic hands. Journal of Computational & Theoretical Nanoscience, 2015, 12(12): 6192-6197

11. Zhe Li, Gongfa Li, Ying Sun, Guozhang Jiang, Jianyi Kong, Honghai Liu. Development of articulated robot trajectory planning. International Journal of Computing Science and Mathematics, 2017, 8(1):52-60

12. Weiliang Ding, Gongfa Li, Ying Sun , Guozhang Jiang, Jianyi Kong, Honghai Liu. D-S evidential theory on semg signal recognition. International Journal of Computing Science and Mathematics, 2017, 8(2): 138-145

13. Yang He, Gongfa Li, Yajie Liao, Ying Sun, Jianyi Kong, Guozhang Jiang, Du Jiang, Bo Tao, Shuang Xu and Honghai Liu. Gesture recognition based on an improved local sparse representation classification algorithm. Cluster Computing, 2017: doi: 10.1007/s10586-017-1237-1

14. Bei Li, Ying Sun, Gongfa Li, Jianyi Kong, Guozhang Jiang, Du Jiang, Bo Tao, Shuang Xu and Honghai Liu. Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm. Cluster Computing, 2017 : doi:10.1007/s10586-017-1231-7

15. Disi Chen, Gongfa Li, Ying Sun, Jianyi Kong, Guozhang Jiang, Heng Tang, Zhaojie Ju, Hui Yu and Honghai Liu. An interactive image segmentation method in hand gesture recognition. Sensors, 2017, 17(2): 253

16. Disi Chen, Gongfa Li, Ying Sun, Jianyi Kong, Guozhang Jiang, Jiahan Li and Honghai Liu: Fusion hand gesture segmentation and extraction based on CMOS sensor and 3D sensor. International Journal of Wireless and Mobile Computing, 2017, 12(3): 305-312

17. Gongfa Li, Wei Miao, Guozhang Jiang, Yinfeng Fang, Zhaojie Ju, Honghai Liu. Intelligent control model and its simulation of flue temperature in coke oven. Discrete and continuous dynamical systems - series s, 2015, 8(6): 1223-1237

18. Weiliang Ding, Gongfa Li, Guozhang Jiang, Yinfeng Fang, Zhaojie Ju, Honghai Liu. Intelligent computation in grasping control of dexterous robot hand. Journal of Computational & Theoretical Nanoscience, 2015, 12(12): 6096-6099

19. Miao Wei, Gongfa Li, Guozhang Jiang, Yinfeng Fang, Zhaojie Ju and Honghai Liu. Optimal Grasp Planning of Multi-Fingered Robotic Hands: A Review. Applied & computational mathematics, 2015, 14(3): 238-247

20. Liqian Feng, Sheng Bi, Min Dong, Yunda Liu. A Gesture Recognition Method Based on Binocular Vision System. International Conference on Computer Vision Systems. Springer, Cham, 2017: 257-267

21. Saba Jadooki, Dzulkifli Mohamad, Tanzila Saba, Abdulaziz S. Almazyad and Amjad Rehman. Fused features mining for depth-based hand gesture recognition to classify blind human communication. Neural Computing and Applications, 2017, 28(11): 3285-3294

22. Stroppa L and Cristalli C. Stereo Vision System for Accurate 3D Measurements of Connector Pins' Positions in Production Lines. Experimental Techniques, 2017, 41(1): 69-78

23. Wei Miao, Gongfa Li, Ying Sun, Guozhang Jiang, Jianyi Kong, Honghai Liu. Gesture recognition based on sparse representation. International Journal of Wireless and Mobile Computing, 2016, 11(4):348-356

24. Emambakhsh Mehryar and Adrian Evans. Nasal Patches and Curves for Expression-Robust 3D Face Recognition. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(5): 995-1007

25. Zhaojie Ju, Xiaofei Ji, Jing Li, and Honghai Liu. An integrative framework of human hand gesture segmentation for human-robot interaction. IEEE Systems Journal, 2015, 99:1-11

26. Gongfa Li, Jianyi Kong, Guozhang Jiang , Liangxi Xie, Zhigang Jiang and Gang Zhao. Air-fuel ratio intelligent control in coke oven combustion process. Information-An International Interdisciplinary Journal, 2012, 15(11): 4487-4494

27. Danut Ovidiu Pop, Alexandrina Rogozan, Fawzi Nashashibi and Abdelaziz Bensrhair. Fusion of Stereo Vision for Pedestrian Recognition using Convolutional Neural Networks. ESANN 2017-25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2017

28. Disi Chen, Gongfa Li, Guozhang Jiang, Yinfeng Fang, Zhaojie Ju, Honghai Liu. Intelligent computational control of multi-fingered dexterous robotic hand. Journal of Computational & Theoretical Nanoscience, 2015, 12(12): 6126-6132

29. Gongfa Li, Jia Liu, Guozhang Jiang and Honghai Liu. Numerical simulation of temperature field and thermal stress field in the new type of ladle with the nanometer adiabatic material. Advances in Mechanical Engineering, 2015, 7(4):1-13

30. Joseph W. Starr and B. Y. Lattimer. Evidential Sensor Fusion of Long-Wavelength Infrared Stereo Vision and 3D-LIDAR for Rangefinding in Fire Environments. Fire Technol, 2017, 53, 1961-1983

31. Gongfa Li, Yuesheng Gu, Jianyi Kong, Guozhang Jiang, Liangxi Xie, Zehao Wu,Zhen Li,Yuan He,Po Gao . Intelligent control of air compressor production process. Applied Mathematics & Information Sciences, 2013, 7(3): 1051-1058

32. D. Scharstein, R. Szeliski. High-accuracy stereo depth maps using structured light Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2003, 1: I-I

33. Qing Mei, Jian Gao, Hui Lin, Yun Chen, He Yunbo, Wei Wang, Guanjin Zhang, Xin Chen. Structure light telecentric stereoscopic vision 3D measurement system based on Scheimpflug condition. Optics and Lasers in Engineering, 2016, 86: 83–91

34. J.J. Aguilar, F. Torres and M.A. Lope. Stereo vision for 3D measurement: Accuracy analysis, calibration and industrial applications. Measurement, 1996, 18(4): 193-200

35. Gongfa Li, Heng Tang, Ying Sun, Jianyi Kong, Guozhang Jiang, Du Jiang, Bo Tao, Shuang Xu, Honghai Liu. Hand gesture recognition based on convolution neural network. Cluster Computing, 2017, DOI:10.1007/s10586-017-1435-x

36. Yong Zhang, Zidong Wang, Lei Zou, Huajing Fang. Event-based finite-time filtering for multi-rate systems with fading measurements. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(3): 1431-1441

37. Y Abdel-Aziz and H M. Karara. Direct linear transformation into object space coordinates in close range photogrametry. Urbana-Champaign, 1971: 1-18

38. Yong Zhang, Zidong Wang, Lifeng Ma, Variance-constrained state estimation for networked multi-rate systems with measurement quantization and probabilistic sensor failures. International Journal of Robust and Nonlinear Control, 2016, 26(16), 3507-3523

39. R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal on Robotics and Automation, 1987, 3(4): 323-344

40. T. Moons, L. Van Gool, Proesmans and M. Proesmans. Affine Reconstruction from Perspective Image Pairs with a Relative Object-Camera Translation in Between. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(1): 77-83

41. B. Triggs. Autocalibration and the absolute quadric. Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. IEEE, 1997: 609-614

42. Yajie Liao, Ying Sun, Gongfa Li, Jianyi Kong, Guozhang Jiang, Du Jiang, Haibin Cai, Zhaojie Ju, Hui Yu and Honghai Liu. Simultaneous Calibration: A Joint Optimization Approach for Multiple Kinect and External Cameras. Sensors, 2017, 17(7), 1491. doi: 10.3390/s17071491

43. Zhou Tian. Face recognition from a single image per person using deep architecture neural networks. Cluster Computing-The Journal of Networks Software Tools and Applications, 2016, 19(1): 73-77

44. Ying Sun, Cuiqiao Li, Gongfa Li, Guozhang Jiang, Du Jiang, Honghai Liu, Zhigao Zheng, Wanneng Shu. Gesture

Recognition Based on Kinect and sEMG Signal Fusion. Mobile Networks and Applications, 2018: https://doi.org/10.1007/s11036-018-1008-0

45. Gongfa Li, Peixin Qu, Jianyi Kong, Guozhang Jiang, Liangxi Xie, Po Gao, Zehao Wu, Yuan He. Coke oven intelligent integrated control system. Applied Mathematics & Information Sciences, 2013, 7(3): 1043-1050

46. Gongfa Li, Peixin Qu, Jianyi Kong, Guozhang Jiang, Liangxi Xie, Zehao Wu, Po Gao, Yuan He. Influence of working lining parameters on temperature and stress field of ladle [J]. Applied Mathematics & Information Sciences, 2013, 7(2): 439-448

47. Gongfa Li, Ze Liu, Guozhang Jiang, Hegen Xiong and Honghai Liu. Numerical simulation of the influence factors for rotary kiln in temperature field and stress field and the structure optimization . Advances in Mechanical Engineering, 2015, 7(6):1-15

48. Fei Wang, Kebin Jia and Jinchao Feng. The Real-Time Depth Map Obtainment Based on Stereo Matching. The Euro-China Conference on Intelligent Data Analysis and Applications. Springer International Publishing, 2016: 138-144

49. Qingxiong Yang. Stereo Matching Using Tree Filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(4): 834-846

50. Mikhail G. Mozerov and Joost van de Weijer. Accurate stereo matching by two-step energy minimization. IEEE Transactions on Image Processing, 2015, 24(3): 1153-1163

51. Heiko Hirschmuller and Daniel Scharstein. Evaluation of Stereo Matching Costs on Images with Radiometric Differences. IEEE transactions on pattern analysis and machine intelligence, 2009, 31(9): 1582-1599

52. Jing Li, Wei Huang, Ling Shao, and Nigel Allinson. Building recognition in urban environments: a survey of state-of-the-art and future challenges. Information Sciences, 2014, 277(2): 406-420