Large-field gesture tracking and recognition for augmented reality interaction

Yangshi Ge[a], Qianhong Jiang[b]
[a]School of Computer Science, Beihang University, Beijing, China;[b]Beihang University, Beijing, China

ABSTRACT

In recent years, with the continuous development of computer vision and artificial intelligence technology, gesture recognition is widely used in many fields, such as virtual reality, augmented reality and so on. However, the traditional binocular camera architecture is limited by its limited field of view Angle and depth perception range. Fisheye camera is gradually applied in gesture recognition field because of its advantage of larger field of view Angle. Fisheye cameras offer a wider field of vision than previous binocular cameras, allowing for a greater range of gesture recognition. This gives fisheye cameras a distinct advantage in situations that require a wide field of view. However, because the imaging mode of fisheye camera is different from traditional camera, the image of fisheye camera has a certain degree of distortion, which makes the calculation of gesture recognition more complicated.Our goal is to design a distortion correction processing strategy suitable for fisheye cameras in order to extend the range of gesture recognition and achieve large field of view gesture recognition. Combined with binocular technology, we can use the acquired hand depth information to enrich the means of interaction. By taking advantage of the large viewing Angle of the fisheye camera to expand the range of gesture recognition, make it more extensive and accurate. This will help improve the real-time and precision of gesture recognition, which has important implications for artificial intelligence, virtual reality and augmented reality.

Keywords: wide field of view, gesture recognition, human-computer interaction, augmented reality

1. INTRODUCTION

In recent years, with the continuous development of computer vision and artificial intelligence technology, gesture recognition has been widely used in many fields, such as virtual reality and augmented reality. However, the traditional binocular camera architecture is limited by its narrow field of view and depth perception range, making it difficult to meet the gesture recognition requirements in certain specific scenarios[1].

To solve this problem, fish-eye cameras have gradually been applied to the field of gesture recognition because of their larger field of view advantages. Compared with previous binocular cameras, fish-eye cameras can provide a wider range of vision, thus achieving a larger range of gesture recognition. This makes fish-eye cameras more advantageous in scenarios that require a large field of view. However, due to the imaging method of fish-eye cameras being different from that of traditional cameras, their images have a certain degree of distortion, making gesture recognition calculations more complex[2].

We aim to solve the problems of limited range of traditional binocular camera gesture recognition and fish-eye camera distortion by using a binocular fish-eye camera. Our goal is to design a distortion correction strategy suitable for fish-eye cameras to expand the range of gesture recognition and achieve large field of view gesture recognition. Combined with binocular technology, we can use the obtained depth information of the hands to enrich the interaction methods. By utilizing the advantages of the fish-eye camera's large field of view to expand the range of gesture recognition, it can be made more extensive and accurate. This will help to improve the real-time performance and accuracy of the gesture recognition field, and is of great significance to fields such as artificial intelligence, virtual reality, and augmented reality[3].

The project aims to solve the problems of limited range of traditional binocular camera gesture recognition and fish-eye camera distortion. To address these issues, a distortion correction strategy suitable for fish-eye cameras is designed to expand the range of gesture recognition and achieve large area gesture recognition[4]. Specifically, our work is to utilize the advantages of the binocular fish-eye camera to expand the range of gesture recognition, making it suitable for a wider range of application scenarios.

2. RELATED WORK

Our research contributions can be specifically demonstrated in the following three aspects.

Further development of gesture recognition: The project is expected to promote the further development of gesture recognition, especially in the field of large field of view gesture recognition. It will greatly enrich the application scenarios of gesture recognition, such as virtual reality, human-computer interaction, and other fields.

New binocular fish-eye gesture recognition algorithm: By using the binocular fish-eye camera and the designed distortion correction algorithm, the project will explore and discover a new gesture recognition solution. This solution will effectively solve the problems of limited range of traditional binocular camera gesture recognition and fish-eye camera distortion, thereby expanding the range of gesture recognition and making it more applicable to complex real-world scenarios.

Practical application value: The project has practical application and commercial potential. By solving the problems in the field of gesture recognition, more application scenarios can be explored, and more practical and efficient products can be developed. This will help to improve people's quality of life and contribute to the development of related industries.

2.1 Principle and characteristic of fisheye camera

Fish-eye camera is a type of wide-angle lens that can capture a very wide field of view onto the image sensor of the camera. The characteristic of this type of camera is that its lens has a very short focal length and a very high degree of distortion, allowing it to capture a very large scene at a short distance and present a strong fish-eye effect in the image.

The imaging principle of a fish-eye camera is to refract and reflect the incident light rays through the fish-eye lens and compress the resulting image onto the sensor to capture a wide field of view. However, the fish-eye lens causes distortion in the resulting image because the light rays from different positions converge on different locations on the camera after passing through the lens. Therefore, computer algorithms are needed to correct this distortion and produce more accurate images.

The imaging model of a regular camera approximates the pinhole imaging, where the light rays passing through a small hole propagate in a straight line and eventually hit the imaging plane. As the angle of deviation from the main optical axis increases, the distance from the center of the imaging plane that the light rays hit increases rapidly, causing severe distortion, resulting in a smaller field of view for regular cameras[5].

The main difference between fish-eye cameras and regular cameras is that the light rays are refracted when passing through a small hole, and larger incident angles can also be refracted to a smaller range, hitting a limited imaging plane, which is the source of the fish-eye's wide field of view, as shown in the figure below.

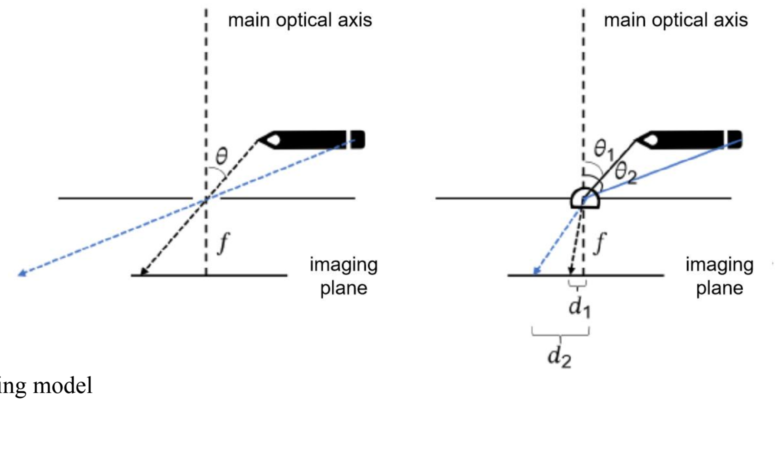

Figure 1. Fish-eye imaging model

Generally, the refractive index of a fisheye lens is designed according to certain principles, which approximately specify the relationship $d = f(\theta)$ between the incident angle of light rays θ and the distance from the center of the imaging plane d when the light rays finally hit the imaging plane. The common relationships are shown in the table below[6].

Table 1. Common projection model

Projection model name	Projection model function	Description
Perspective projection	$d = f\tan\theta$	Pinhole imaging model
Stereopic projection	$d = f\tan\theta/2$	The Angle between the lines doesn't change
Isometric projection	$d = f\theta$	The distance is proportional to the Angle of incidence
Equal-area projection	$d = f\sin\theta/2$	The solid Angle of the object does not change
Orthogonal projection	$d = f\sin\theta$	Maximum distortion, Angle of view less than 180 °

Projection functions of all projection models are odd functions. In order to universally express projection models of all fisheye cameras, Taylor expansion is carried out on them, and there are only odd-degree terms:

$$d = k_1\theta + k_2\theta^3 + k_3\theta^5 + k_4\theta^7 + k_5\theta^9 + ... \quad (1)$$

In reference[7]retains the first five terms as the approximation of the projection model, and takes $(k_1, k_2, k_3, k_4, k_5)$ as the undetermined distortion parameter. OpenCV further simplifies the above formula and fixes the coefficient of the primary term as 1 to obtain the final projection model:

$$d = \theta + k_1\theta^3 + k_2\theta^5 + k_3\theta^7 + k_4\theta^9 \quad (2)$$

Distortion parameter: $(k_1, k_2, k_3, k_4, k_5)$.

2.2 Binocular fisheye calibration

The task of fisheye camera calibration is to solve the imaging model of a specific fisheye camera, that is, to determine various parameters, including the intrinsic matrix and distortion parameters that implement the transformation from the camera coordinate system to the pixel coordinate system. The task of stereo calibration is to further correct the errors in the installation of the two cameras, obtain the rotation matrix and translation vector between the two cameras, as well as the projection matrix in the new coordinate space, and the re-projection matrix in the new coordinate space after applying the above model parameters to align the left and right images. Finally, the stereo vision system is able to measure the distance between objects. The process is shown in the diagram below[8].

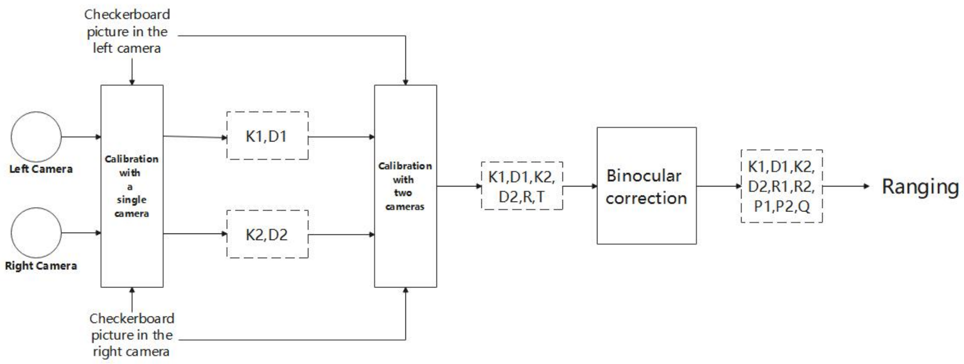

Figure 2. Binocular calibration process

The chessboard calibration method is used for calibration, which has the advantages of convenient use and easy recognition of corner points. The key points used for correction are the corner points formed by the black and white squares in the chessboard. The "world coordinate system" of the corner points is defined with the upper left corner of the chessboard as the origin and the chessboard plane as the x and y axes, without the need for special measurements. After capturing images and identifying corner points, images with incorrectly identified corner points are first filtered out. Then, the floating-point sub-pixel coordinates of each corner point are calculated to improve the accuracy, and finally the calibration is completed using OpenCV API.

3. FISHEYE CORRECTION

Most gesture recognition algorithms are designed for regular cameras. However, in fisheye cameras, existing gesture recognition methods can perform well when the hand is close to the center of the image and the distortion is minimal. However, at the edges of the image, the hand occupies a smaller area of the image, and the distortion at the image edges is larger, which is not conducive to gesture recognition. Specifically, when the hand suddenly appears at the edge of the image, the gesture tracking algorithm may fail to recognize the hand and become invalid. Therefore, it is necessary to perform some correction on fisheye images.

3.1 Corrects to keyhole camera model

The most popular fisheye correction approach is to correct the fisheye image to the image of the pinhole camera model, because the pinhole camera model seems to be the most familiar camera model, and existing open-source and reliable gesture recognition tools are trained on the pinhole camera model. Using the calibration results mentioned above, the remapping matrix that corrects the fisheye image to the imaging space of the pinhole camera model can be obtained.

However, as shown in Figure 1, in the pinhole camera model, the distance from the point where light hits the imaging plane to the center sharply increases as the incident angle increases[5]. Objects at the edge of the fisheye camera will undergo severe stretching distortion after correction, resulting in a significant loss of the viewing angle. At the same time, even parts close to the center of the image will also undergo significant stretching distortion, seriously destroying the shape features of objects. As shown in the figure below, the hand undergoes severe deformation after stretching, which destroys the shape features of the hand and is not conducive to gesture recognition, even worse than directly using the fisheye image for recognition.

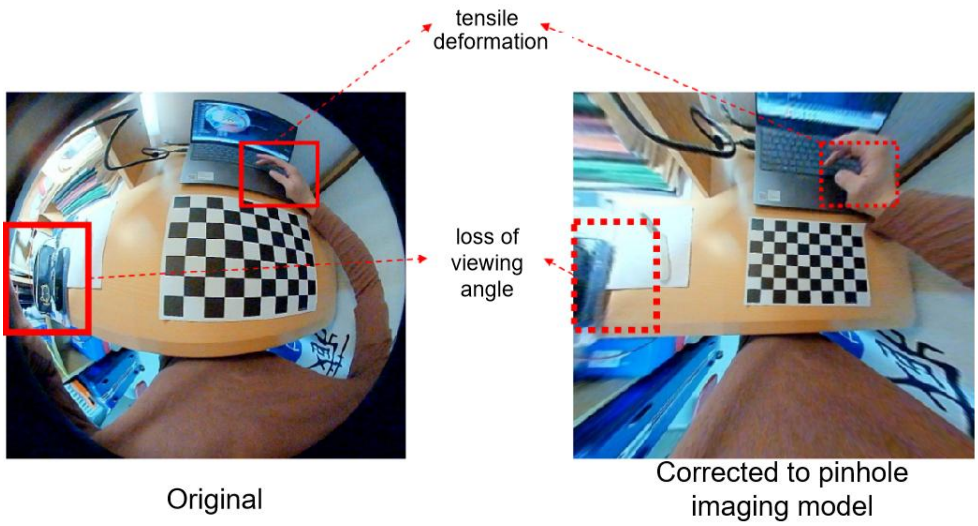

Figure 3. Corrected for pinhole camera model results schematic

3.2 Transverse development correction

3.2.1 Principle and advantage

Another fisheye correction method is similar to unfolding the fisheye photo into a panoramic image. In the fisheye image, the further away from the center, the more stretching there is along the tangential direction. The horizontal unfolding method abandons the fisheye camera's own imaging model and directly "unfolds" the original circular image into a long strip. Obviously, the farther away from the center of the circle, the longer the length after the circle is unfolded. The horizontal unfolding method makes all the circular strips of the original image the same length, which is equivalent to stretching the inner circle to a certain extent, thus compensating for the problem of tangential distortion in the outer circle and making the image look corrected, reducing the distortion for the user [9]. The horizontal unfolding method has a good ability to restore the shape of objects near the outer circle of the fisheye.

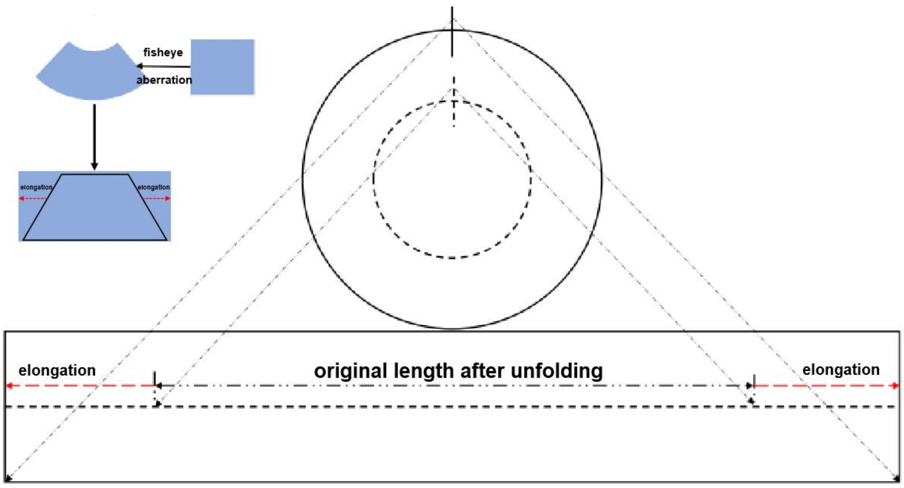

Figure 4. Corrected for pinhole camera model results schematic

Specifically, the radius of the circular fisheye camera image is R, and the width of the expanded panoramic image is $[2\pi R]$ and the height is R. For each pixel in the expanded image, the pixel coordinate (x,y) satisfies the condition $(x,y), 0 \leqslant x < [2\pi R], 0 \leqslant y < R$.

Define the x coordinate of this point in the original image as x_{map} and y coordinate as y_{map}, and establish the mapping between pixel points in the new image and pixel points in the original image according to the following formula:

$$x_{map}(x,y) = R + y \sin \frac{x}{R} \tag{3}$$

and

$$y_{map}(x,y) = R - y \cos \frac{x}{R} \tag{4}$$

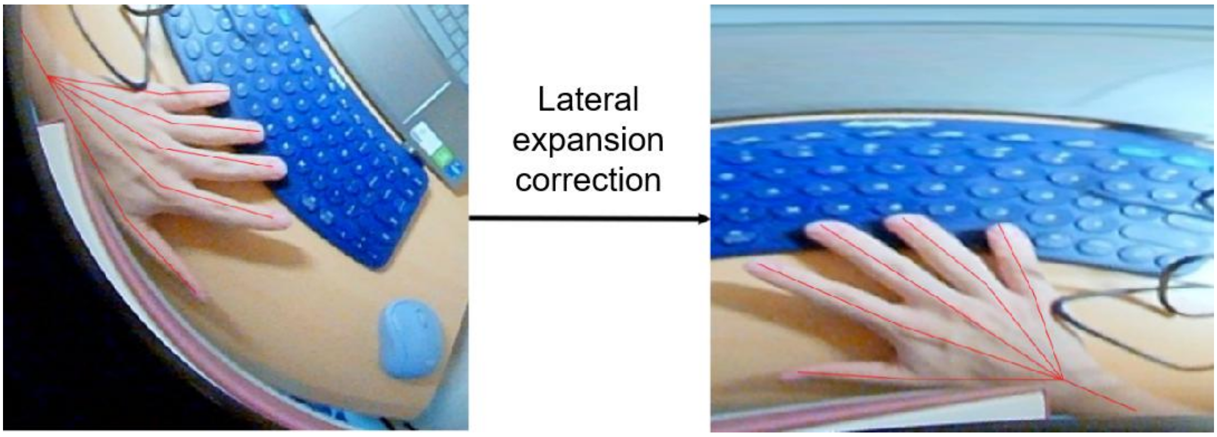

Figure 5. Expand the effect image horizontally

3.2.2 Limitations of the horizontal expansion method

The horizontal expansion method expands all circles with the center of the fisheye image as the center, and stretches them to the same length as the outermost circle, causing severe distortion in the middle of the image near the center, which must be abandoned. In contrast, normal fisheye images generally do not produce distortion near the center of the screen, and the two methods complement each other.

In addition, the size of the unfolded image is much larger than the original, and the hand can only occupy a small portion of the image, which is also unfavorable for gesture recognition. When the proportion of the hand in the image is too small, it is easily ignored by the gesture recognizer and cannot be recognized.

3.2.3 Fisheye edge gesture recognition based on sliding window and horizontal expansion

Hands that cannot be directly recognized by the gesture recognizer are generally located at the edges of the screen, with deformation issues and a small proportion in the image. Therefore, the horizontal expansion method is used to correct the edge fisheye image, and the sliding window method is used to extract part of it in the unfolded image for further and more precise recognition to make up for the shortcomings of the initial recognition, as shown in Figure 6.

Considering that the camera in AR or VR devices is often installed in front of the forehead, and hands are usually held above the head, higher than the camera, and less likely to be located in the upper part of the screen. Therefore, we optimized the search order of the sliding window by using a method that searches from the middle to both sides. When the hand is recognized, the search can be terminated in advance to speed up the search.

Because hands that cannot be directly recognized occupy a small proportion of the screen, the proportion occupied after expansion is even smaller. Combined with the image resolution we use, we use a small window of 251*150 (width * height) and a step of 125, so that at least half of adjacent windows overlap to avoid the hand crossing the boundary of two windows. At the same time, the 150 height removes the severely distorted part near the fisheye center in the unfolded image to avoid meaningless calculations. After experiments, processing one frame in this way takes about 0.1~0.2s, which is a relatively large time cost compared to real-time (more than 30 frames per second). However, considering that real-time is meaningless if the hand cannot be recognized, and the speed can be increased to real-time after successful tracking, such a time cost is acceptable.

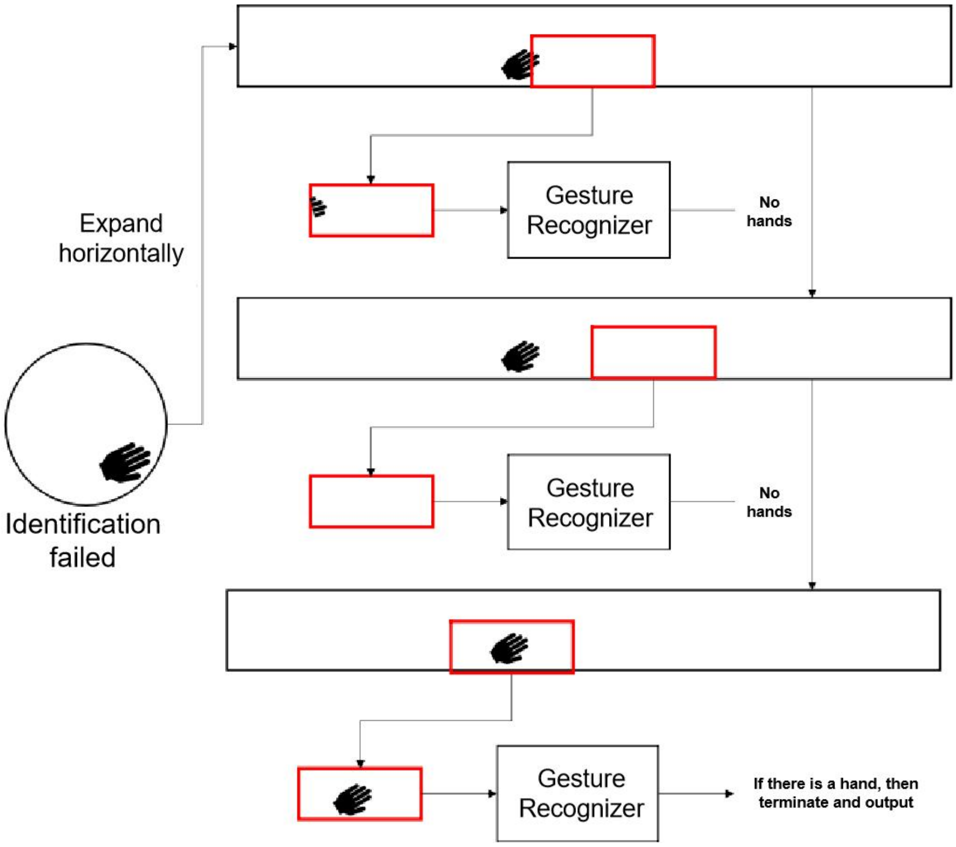

Figure 6. Slide the window to search for the sketch

Through experiments, it was found that when hands are located at the edge of the fisheye and far from the camera, and gesture recognition directly on the original fisheye image fails, hands can still be recognized. This method can retain the advantages of less distortion when hands are located in the center of the fisheye image, and also make up for these shortcomings when hands are located at the edge, producing distortion and occupying a small portion of the screen. This approach fully leverages the advantages of the fisheye camera's wide angle of view.

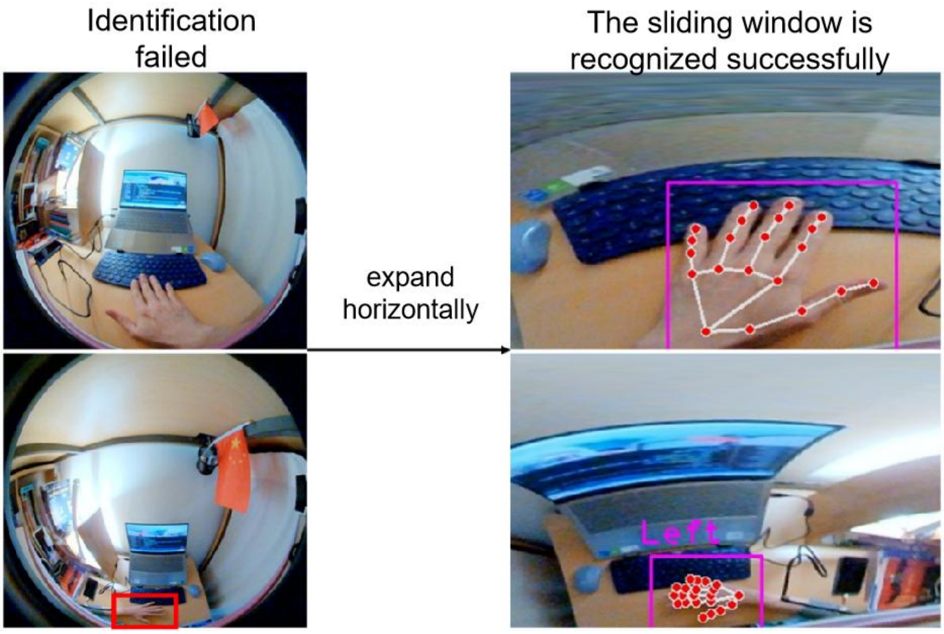

Figure 7. Horizontally expand the effect of the sliding window

4. BINOCULAR FISHEYE RANGING

Generally speaking, the process of binocular fisheye reconstruction is as follows: the binocular calibration results are used to correct the fisheye image, the left and right images are matched, the x pixel coordinates of the matched points are different to get parallax, and finally the reprojection matrix Q is used to calculate the coordinates of the final object in the camera coordinate system[10].

As mentioned above, correcting speech images directly with calibration results is not only bad for hand recognition because of deformation, but also loss of perspective. As for the most important step in binocular ranging method: matching, the coordinates of 21 key points in the left and right figures have been given by the neural network algorithm, and we only care about the positions of these key points, so there is no need to use correction for matching.

Key point recognition of the hand is equivalent to matching. We move the correction step to matching, and directly correct the identified pairs of key points into the dedistorted space to obtain new coordinates, without correcting the direct image, so as to avoid the problem of correcting the deformation.

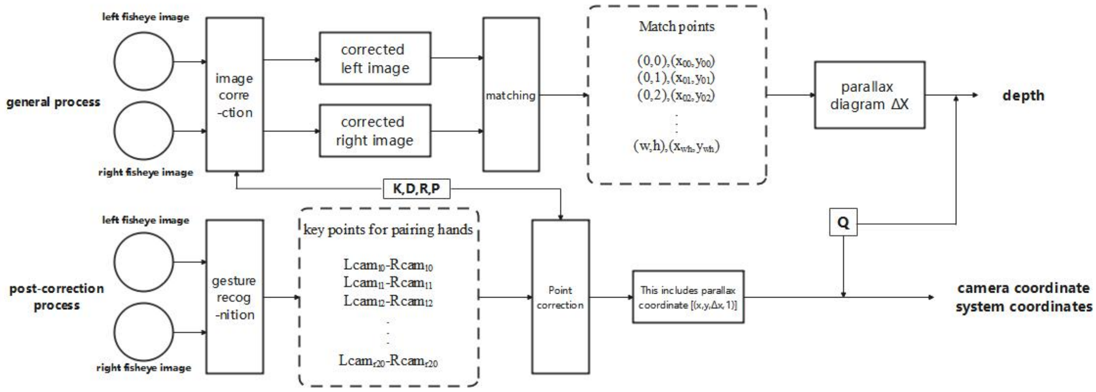

Figure 8. Deep computing process comparison

After the coordinate of the key points of the hand paired in the left and right camera photos is obtained by gesture recognition, these key points will be aligned after correction due to the previous binocular correction calibration, that is, the y coordinate values of the paired points in the new space are basically the same. For all pairs of points, parallax Δx is obtained by using the X-coordinate difference in the corrected space, and parallax homogeneous coordinate $(x, y, \Delta x, 1)^T$ is obtained. Using the reprojection matrix Q obtained by binocular correction calibration, the following formula is used to calculate the coordinates in the camera coordinate system[11]:

$$(x', y', z', w')^T = Q(x, y, \Delta x, 1)^T \tag{5}$$

and

$$(X, Y, Z) = \left(\frac{x'}{w'}, \frac{y'}{w'}, \frac{z'}{w}\right) \tag{6}$$

5. BINOCULAR FISHEYE GESTURE RECOGNITION ALGORITHM

So far, we design a complete binocular fisheye gesture recognition algorithm. This algorithm takes the photos taken by the fisheye camera as the input and the identified key points of the left and right hand in the coordinate system of the left camera in millimeters as the output. The algorithm flow is shown in the figure below.

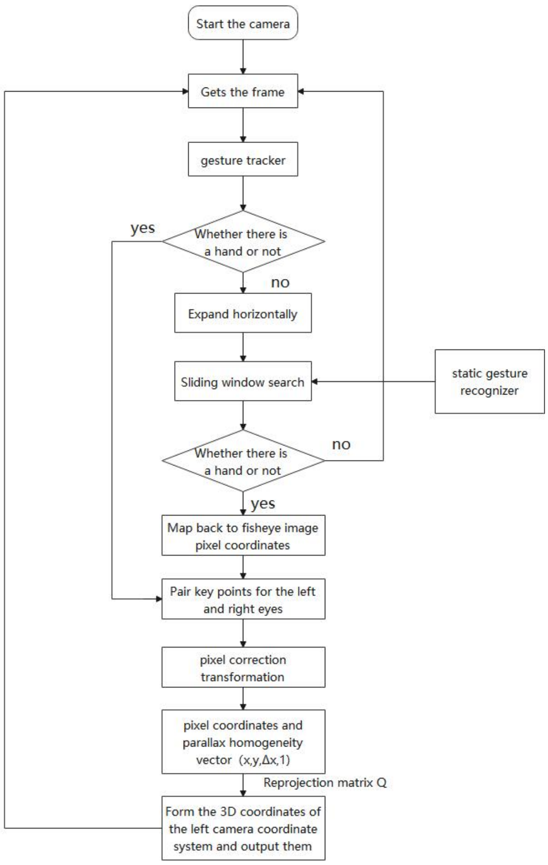

Figure 9. Flow chart of binocular fisheye gesture recognition algorithm

Gesture recognizers are divided into "gesture tracker" and "static gesture recognizer". Gesture tracking in the before and after the input frame is relevant, according to the previous frame of identify the location of the hand has been estimated that the position of the hand, the frame only when lost track position, just call the hand placement of search web search, otherwise directly identify key point, speed up the recognition effectively. The algorithm is used to recognize gestures directly on fisheye images with continuous input. Static gesture recognizer thinks that the input picture is irrelevant before and after, and uses hand view network and key point recognition network to recognize each picture, which is suitable for processing independent pictures. In the algorithm, it is used for sliding window search after horizontal expansion.

After starting the camera, for each frame acquired, the fisheye image is directly processed with the gesture tracker. If the hand is traced, the depth calculation is directly entered. If no hand is found, the current frame is expanded horizontally, and the static gesture recognizer is used to search for the hand by sliding search method. If no hand is found, it is considered that there is no hand, and the next frame is obtained by returning to the beginning. If it is found, the pixel coordinates of hand key points are mapped from the horizontal expansion map back to the position in the original fisheye map, and then the depth calculation is carried out. Finally, output the coordinates of the calculated key points in the left camera coordinate system, and go back to the beginning to obtain the next frame.

The algorithm successfully applied the gesture recognition tool to the fisheye camera, and carried out targeted improvements on the fisheye edge and gesture recognition due to distortion or distance failure by using the horizontal expansion correction and sliding window search method. By taking full advantage of the features of the fisheye camera's large Angle of view, the effective Angle range of gesture recognition was increased to more than 150°. Even close to the maximum viewing Angle of the fisheye camera used (above 180°).

6. SYSTEM FRAMEWORK AND HARDWARE SETUP

In this project, we aim to use two fisheye cameras for real-time object detection and tracking. To achieve this goal, we need a complete system framework and relevant hardware. Below is a detailed description of the system framework and hardware setup we designed.

System framework: Our system framework mainly includes the following steps:

1. Initialization and opening of cameras: Use the Python programming language and OpenCV computer vision library to initialize the camera device and open the camera to start capturing images.

2. Fisheye distortion correction: By performing distortion correction on the fisheye camera image, we can improve the image quality and make it more accurate.

3. Image acquisition: Use two fisheye cameras fixed on the same plane to capture images and process them.

4. Real-time object detection and tracking: Use object detection and tracking algorithms, such as mediapipe, to recognize and track objects in the scene.

5. Send data to the computer or Unity for presentation: Send the processed images to the computer or Unity for presentation.

Hardware setup: To maintain high accuracy, we need to do the following in hardware setup:

1. Two fisheye cameras are fixed on the same plane: We need to use a plane bracket to keep two fisheye cameras on the same horizontal plane.

2. Use computer vision algorithms to calibrate camera positions: To ensure the spatial position accuracy between cameras, computer vision algorithms can be used to calibrate camera positions.

The system framework and hardware setup used in this project can achieve real-time object detection and tracking in dynamic environments. At the same time, we also need to ensure that the computer has sufficient performance to process high-resolution images captured by each camera, and the network bandwidth allows camera images to be transmitted to the computer or Unity in real-time performance. With these technologies and achievements, we can improve the accuracy of object detection and tracking and bring better results to fields such as robotics, autonomous driving, and virtual reality.

7. DEMONSTRATION OF EXPERIMENTAL EFFECT

In the computer side display scenario, we set up a multi-window environment, similar to a workbench or control room. In this environment, the user can control through gestures (controlling the mouse and keyboard), such as using fingers to zoom in on the size of each window or move the window position, etc. Because binocular fisheye cameras provide a wider field of view, users can freely move and manipulate individual Windows across the entire screen through gestures, without the need to switch and adjust Windows frequently. This can improve the user's work efficiency and comfort.

In Unity's demo scene, we've used the hand information we've captured to map it into the animation. This can also set up a variety of virtual environments, such as the living room, bedroom and kitchen. In this environment, users can interact with virtual objects through gestures, such as turning on the TV, adjusting the lights, adjusting the curtains and so on. Because fisheye cameras provide a wider field of view and more precise depth information, users can interact freely across a room using gestures and can more accurately control the position and attitude of virtual objects. This can improve user immersion and experience.

The functions and features shown in these scenes are realized based on the large field of view technology. Through the large field of view technology, we can expand the user's field of view and operation range, so that the user is more free to operate in a wider space. At the same time, the large field of view technology can also provide more accurate depth information, so that the system can recognize the effect of gestures more accurately and reliably. In this way, we can create a more intelligent, convenient and comfortable human-computer interaction system, bringing more convenience and fun to people's life and work.

8. RESULT ANALYSIS AND DISCUSSION

8.1 Analysis of calculation accuracy and practical application effect

Firstly, as for the calculation accuracy, the binocular parallax method is used to calculate the depth information of the hand, so the accuracy will be affected by many factors, such as hardware calibration, image preprocessing, etc. In the project, the accuracy of depth information calculation can be effectively improved through the hardware setting and calibration of the binocular fisheye camera and the distortion correction of the fisheye camera. Meanwhile, the algorithm parameters can be adjusted according to specific application scenarios to further improve the precision of depth information.

Second, with regard to practical application effects, the project can provide more accurate and reliable data support for tasks such as object recognition, environment perception and path planning in robotics, virtual reality and autonomous driving. For example, in the field of autonomous driving, depth information acquisition based on fisheye cameras can help vehicles better sense surrounding roads and obstacles, thus improving driving safety and efficiency. In the field of virtual reality, the large field of view feature of fisheye camera can help users better perceive the surrounding environment and enhance the fidelity and interaction of virtual experience. In addition, the deep information acquisition technology based on this project can also provide more accurate and reliable data support for medical, industrial and other fields.

8.2 Correction algorithm optimization analysis

The algorithm is optimized for scalability and adaptability. Because different fisheye cameras have different imaging characteristics and distortion conditions, it is necessary to adjust and optimize for different camera types and application scenarios. Therefore, learning methods, such as convolutional neural networks, can be considered to predict and process distortion conditions and imaging features of different cameras, so as to improve the universality and adaptability of the algorithm.

The real-time performance of the algorithm is optimized. In practical application, the processing speed and real-time performance of the algorithm are very important. Therefore, image compression, feature extraction and other methods can be considered to reduce the amount of data and computational complexity, so as to improve the processing speed and real-time performance of the algorithm.

8.3 Research deficiency and improvement direction

Firstly, the calculation accuracy of the algorithm is improved. Although the distortion correction strategy suitable for fisheye camera is adopted in this project to improve the calculation accuracy of depth information, the binocular parallax method itself has certain error limits, which may affect the accuracy of depth information. Therefore, more advanced and complex algorithms, such as depth estimation of single image based on deep learning, can be introduced to improve the computational accuracy of depth information and make it more in line with the needs of practical applications.

Secondly, the reliability and stability of the algorithm are improved. In the actual application process, the uncertainty of environment and the interference of noise may lead to the reliability and stability of the algorithm. Therefore, the robustness and stability of the algorithm can be improved by increasing data samples and optimizing algorithm parameters, so as to ensure the reliability and stability of the algorithm in practical application.

In addition, the real-time performance of the algorithm is improved. In some real-time application scenarios, the real-time performance and processing speed of the algorithm are very important. Although some optimization measures have been taken to improve the real-time performance and processing speed of the algorithm, there may still be some bottlenecks and limitations. Therefore, more efficient and fast algorithms, such as GPU parallel computing and FPGA acceleration technologies, can be considered to further improve the real-time performance and processing speed of the algorithm.

Finally, the project can be considered to be combined with more practical application scenarios, such as unmanned aerial vehicle inspection, smart home and other fields, so as to expand the application scope and application scenarios of the project and better meet the needs of society.

In a word, there are some shortcomings and improvement space in this project, which needs to be explored and improved constantly to make it better meet the needs of practical applications and create greater value for the society. In the future, with the continuous development and improvement of the technology, there is a lot of room for further research and application of this project.

9. CONCLUSION

In this project, the binocular fisheye camera was used to obtain the depth information of the hand, and the distortion correction strategy suitable for the fisheye camera was designed, thus improving the calculation accuracy of the depth information. Another important innovation is the use of a fisheye camera with a large field of view, which is of great significance for indoor navigation, human-computer interaction, virtual reality and other applications.

Specifically, the large field of view fisheye camera can provide a wider field of vision and can cover a larger space than traditional cameras, thus improving the efficiency and accuracy of object detection and recognition. The acquisition of hand depth information by using large-field fisheye camera can effectively avoid the problem of not being able to obtain complete hand depth information due to insufficient camera field of view, and also expand application scenarios, such as object recognition, environment perception and path planning tasks in virtual reality, automatic driving, medical and other fields.

In addition, the large field of view feature of fisheye camera can also improve the efficiency of human-computer interaction, which has a wide range of applications in indoor navigation, gesture recognition and other fields. For example, in the field of smart home, users can complete the operation of devices through gesture control, while the use of large field of view fisheye camera can realize more accurate and natural gesture recognition.

This study has profound significance and extraordinary application prospect, which can be expressed in three aspects.

First, the project has made important breakthroughs in the acquisition of depth information about hands. Hand depth information is the basis of gesture recognition, hand posture estimation, hand motion analysis and other tasks, and can provide reliable data support for object recognition, environment perception, path planning and other tasks in virtual reality, automatic driving, medical and other fields. Therefore, the research results of this project have important theoretical and practical significance, which can provide more accurate and comprehensive data support for the research and application in related fields.

Secondly, the project adopts a wide-field fisheye camera to acquire depth information of the hand, which is of great significance for indoor navigation, human-computer interaction, virtual reality and other application fields. The use of large field of view fisheye camera can provide a wider range of visual field, thus improving the efficiency and accuracy of object detection and recognition, and has been widely used in indoor navigation, gesture recognition and other fields. With the continuous development of artificial intelligence technology, the technology of large-field fisheye camera will become more and more mature, providing more reliable technical support for the development of these fields.

Finally, the project also provides new ideas and methods for deep learning and computer vision research. For example, in terms of distortion correction processing strategy, the project designed a distortion correction algorithm suitable for fisheye cameras, and improved the calculation accuracy of depth information by effectively calibrating camera parameters and optimizing algorithm parameters. These algorithms and methods have certain generalization ability and can provide reference for deep learning and computer vision problems in other fields.

In short, the project has made important breakthroughs in theory and practice, and has wide application prospects. In the future, we can further improve and optimize the project to adapt to more complex and diverse application needs, and create greater value for society.

REFERENCES

[1] Furnas, George W. "Generalized fisheye views." Acm Sigchi Bulletin 17.4 (1986): 16-23.

[2] Shouzhang, Xiao, and Wang Fengwen. "Generation of panoramic view from 360 degree fisheye images based on angular fisheye projection." 2011 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. IEEE, 2011.

[3] Sio, Chon Hou, Hong-Han Shuai, and Wen-Huang Cheng. "Multiple fisheye camera tracking via real-time feature clustering." Proceedings of the ACM multimedia asia. 2019. 1-6.

[4] Xu, Zhenjin, et al. "Advanced hand gesture prediction robust to electrode shift with an arbitrary angle." Sensors 20.4 (2020): 1113.

[5] Schneider, Danilo, Ellen Schwalbe, and H-G. Maas. "Validation of geometric models for fisheye lenses." ISPRS Journal of Photogrammetry and Remote Sensing 64.3 (2009): 259-266.

[6] Lee, Sangil, et al. "Exposure correction and image blending for planar panorama stitching." 2016 16th International Conference on Control, Automation and Systems (ICCAS). IEEE, 2016.

[7] Kannala, Juho, and Sami S. Brandt. "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses." IEEE transactions on pattern analysis and machine intelligence 28.8 (2006): 1335-1340.

[8] Yang, Lu, et al. "Analysis on location accuracy for the binocular stereo vision system." IEEE Photonics Journal 10.1 (2017): 1-16.

[9] Zhang, Qieshi, and Sei-Ichiro Kamata. "Fisheye image correction based on straight-line detection and preservation." 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2015.

[10] Xie, Liangliang, Xu Zhang, and Dawei Tu. "Underwater large field of view 3D imaging based on fisheye lens." Optics Communications 511 (2022): 127975.

[11] Lin, Mingxiu, et al. "Cylindrical panoramic image stitching method based on multi-cameras." 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2015.