A Binocular Vision System for Object Distance Detection with SIFT Descriptors

BI Ping

School of Electronics Engineering, Xidian University, Xi'an 710071, China
School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
biping@xupt.edu.cn

Abstract

The purpose of the real-time monitoring of target or incident is to have a timely alarm, to find the invasion of effective goal is the core technology. Distance measuring of intrusive targets is important for indoor monitoring, according to the detected intrusion target image, calculating the distance between the target and the camera, it can provide the basis for the computer to determine the type of invasion, and finally provide information for the security alarm. In order to meet the needs of the indoor environment monitoring, the paper hereby presents a set of binocular vision-based motion detection system. The system is divided into two parts, one part is offline training, and another is online computation. The offline part consists of stereo calibration module and revised distance parameter module. The online part consists of data acquisition and storage module, moving target detection module, the corresponding point matching module and ranging module. In Stereo calibration module, the internal and external parameters of two cameras are obtained. In motion detection module, for dual-channel video capture respectively, the background subtraction algorithm is used for object detection, and a mixed Gaussian model is used as the adaptive background updating method. Then based on static camera and fixed position, objects in binocular images obtained by binocular camera will be coarse matched firstly, then will be fine matched using scale invariant features transform algorithm to find the accurate match points under this system. In the mean time, revised distance parameter module provides a correction parameter for the traditional binocular parallax distance formula. Finally, the distance between the target and the camera is calculated. Experiments show that this system can be very effective in extracting moving targets, getting match points then measuring distance, especially in distance measuring of obstructed targets.

Keywords: *Object detection, Binocular vision, Camera calibration, SIFT feature, Feature matching*

1. Introduction

Areas of artificial intelligence deal with autonomous planning or deliberation for robotical systems to navigate through an environment. A detailed understanding of these environments is required to navigate through them. Information about the environment could be provided by a computer vision system, acting as a vision sensor and providing high-level information about the environment and the robot.

Traditionally, video surveillance, which is a useful tool to assist the Department of Public Safety to fight against crimes, maintain social stability, is widely applied in security protections. In recent years, with the development of IPv6 technology and

information appliances, further optimization of the mobile monitoring equipments, the improvement of image processing technology, and the wide application of Internet of Things and cloud computing, video surveillance technology is more and more affects and influences other various fields, such as digital home, education, government, entertainment, medical, hotels, sports, etc. [1-7].

Indoor monitoring target refers to moving target which invades to the monitoring area under an improper circumstance, such as people, animals, and so on. Moving target detection method based on machine vision technology is that cameras are installed in the indoor fixed area, according to the detected intrusion target image, calculating the distance between the target and the camera. It can provide the basis for the computer to determine the type of invasion, and finally provide information for the security alarm.

Even though, there have been a lot of classical algorithms on video surveillance methods, the results are not as expected during practical applications. In order to obtain a feasible and effective method for distance measuring of indoor intrusive targets, on the basis of analyzing the characteristics of indoor monitoring, the paper proposes a matching algorithm based on combination of regional crude matching and SIFT feature matching, and modifies the distance detection model. The system is divided into two parts, one part is offline training, and the other is online computation. The offline part consists of stereo calibration module and revised distance parameter module. The online part consists of data acquisition and storage module, moving target detection module, the corresponding point matching module and ranging module. Figure 1 is our binocular stereo vision system.

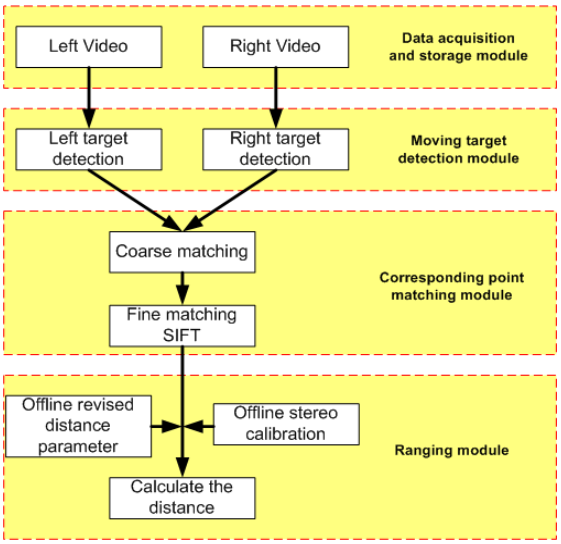


Figure 1. Our Binocular Stereo Vision System

2. Stereo Calibration Module

Camera calibration is a necessary step in 3D computer vision in order to extract metric information from 2D images. Much work has been done, starting in the photogrammetry community and more recently in computer vision. We can classify those techniques roughly into two categories: photogrammetric calibration and self-calibration. In this paper, the camera calibration method based on 2D plane template proposed by Zhang Zheng-You [8] PhD who works in the research institute of

Microsoft is adopted. In our system, internal parameters and external parameters of the two cameras need to be offline calibrated. The calibration procedure is as follows [8]:

1. Print a pattern and attach it to a planar surface.
2. Take a few images of the model plane under different orientations by moving either the plane or the camera.
3. Detect the feature points in the images.
4. Estimate the five intrinsic parameters and all the extrinsic parameters using the closed-form solution.
5. Refine all parameters, including lens distortion parameters.

Refer to the referencs [8], for a more detailed description. Then further we globally optimized the two camera's parameters, so as to minimize the re-projection error of calibration grid of the two cameras. The internal parameters of final offline calibration of the cameras and the system need some external measurement data which are shown in Table 1. Figure 2 shows Three-dimensional display of the binocular stereo vision system parameters.

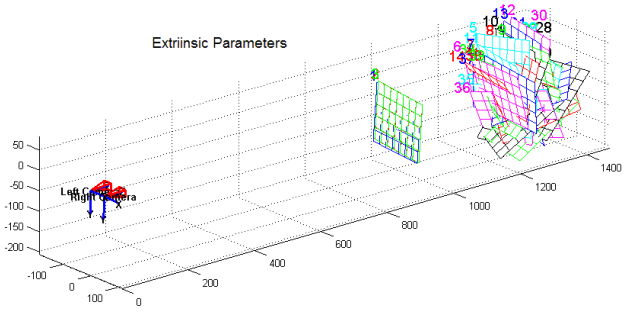


Figure 2. Three-dimensional Display of the Binocular Stereo Vision System Parameters

Table 1. Offline Calibration and System Parameters

f (mm)	fx (mm/pix)	fy (mm/pix)	B (mm)
12	4.8/795	3.6/576	45

3. Ranging Module

Stereo vision is a way to obtain 3D information of objects in space by multiple images. For example, the biological vision systems, almost the entire organism which has visual system has two eyes. And due to the difference between normal pupil distance and version angle, certain horizontal differences would be caused for images on the left and the right eye retinas. When a three-dimensional target is observed, since the two eyes are apart from each other by about 60mm, it will be observed from different angles. This small horizontal aberrations arisen from binocular retinal imaging

process, is known as stereoscopic vision [9]. Parallax is an objective physical phenomenon, which belongs to the depth of information generated by the horizontal parallax, that is, the physiological basis of stereo vision. Binocular stereo vision system is modeled on this principle, which uses two cameras to obtain two images of the same scene from different angles at the same time, then calculates the target point in the disparity in the two images to obtain three-dimensional scene information [10].

Figure 3 is a schematic diagram of the binocular stereo vision system, two cameras with the same performance parameters, and optical axis parallel to each other, the x-axis is collinear, along the x-axis are apart from , camera's optical axis parallels to the z-axis, and two image's plane parallel to each other. In Figure 3, O_1, O_2 are the focus of the left and right camera, I_1, I_2 are the camera's image planes, P_l, P_r are the imaging point of the space point P in the left and right image plane respectively, f is the focal length of the camera. If the stereoscopic vision d is defined as $|P_l - P_r|$, then the distance between the intrusion target and fixed cameras can be calculated as (1):

$$Z = \frac{Bf}{|P_l - P_r|} \quad (1)$$

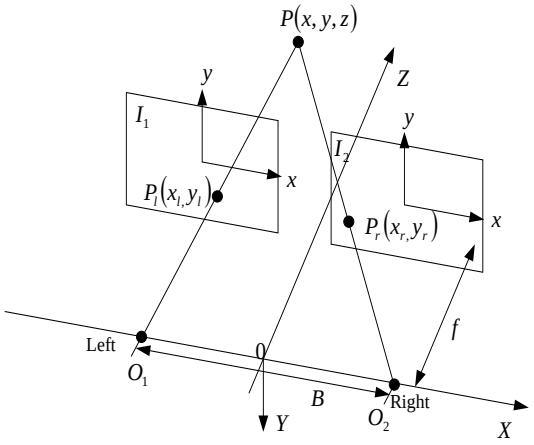


Figure 3. Schematic diagram of distance measuring by binocular stereo vision system: O_1, O_2 are the focus of the left and right camera, I_1, I_2 are the camera's image planes, P_l, P_r are the imaging point of the space point P in the left and right image plane respectively, f is the focal length of the camera, B is the distance between O_1 and O_2.

Equation (1) shows that the parallax of the two images can be applied to calculate the distance, but the calculation of parallax itself is the difficulty in the binocular vision, it requires to match the features of the two images [11]. In addition, this model requires identical calibration parameters of two cameras, and the parallax only exist horizontal displacement but no vertical displacement. However, certain errors exist in practical applications. Therefore, image horizontal corrections and distance model modification are necessary. Figure 4 is the invasion target images obtained by binocular camera in the room with horizontal correction and object detections. The background subtraction algorithm is used for object detection, and a mixture Gaussian model is used as the adaptive background updating method.

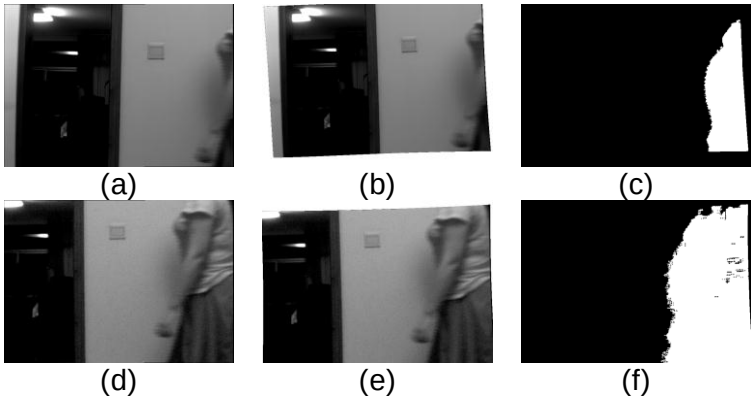


Figure 4. Binocular images and corrected images: The video is collected by ourselves, shown is frames 90 in video, the size of each image are 768×576. (a)Original left video image, (b)After horizontal correction of left video image, (c)object detection result in left video, (d)Original right video image, (e)After horizontal correction of right video image, (f)object detection result in right video.

4. Point Matching Module

The image matching is the alignment of two or more images of the same target in space position. There are basically two kinds of matching method, based on gray level of matching and based on feature matching respectively. Compared with the match method based on gray, feature matching has faster speed, higher accuracy, not sensitive to light and smaller influenced by noises, therefore, more suitable for indoor monitoring.

4.1. Scale Invariant Features Transform Algorithm [12]

SIFT namely scale invariant features transform algorithm, which is proposed by David G. Lowe [12], presents a feature matching algorithm summarized on the basis of the existing detection methods based on invariant features. The method is built on the basis of image scale space, and finds the extreme value points in scale space, then extracts location, scale and rotation invariant features [13-16].

1. The main detection process of extreme value points in scale space in SIFT feature matching algorithm is: in scale space, using the kernel of gauss [12] to build gaussian pyramid, then constructing DOG pyramid, and detecting extreme value points in the DOG gaussian pyramid, in the end, tentatively determine the position of the feature points and their scales.
2. Precisely locate the extreme value point to eliminate the low contrast extreme value point and unstable edge response points, obtain the local feature point.
3. For the selected partial feature points, using characteristics of gradient distribution of neighborhood pixels of feature points to designate their direction parameters, which make SIFT operator has invariant rotation feature.
4. With the feature point as the center of 8×8 window subregion, in every 4×4 sample regions, calculate gradient orientation histogram of eight directions, then sum the values of each gradient orientation and form a keypoint, where each keypoint has eight direction information. Then describe each feature point with 4×4, a total of 16 keypoints. Therefore,

for one single feature point can generate 128 data, which eventually form the 128-dimensional SIFT of feature vector.

5. SIFT matching is mainly to calculate the similarity of SIFT features for two to be matched images, calculating the nearest match from each feature points of one image to all feature points of another image, then use Euclidean distance measure as the similarity of the feature points. In order to exclude the non-matching feature points, caused due to occluded target image and background confusion, Lowe proposed to eliminate the mismatch by comparing the nearest neighbor distance and secondary nearest neighbor distance. When the ratio is less than the distance ratio threshold to determine the correct match otherwise mismatched.

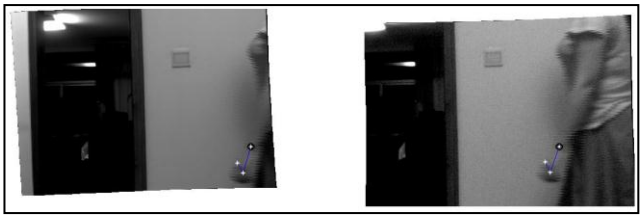
4.2. Rough Match Region Selection

The algorithm steps are as follows:

1. Read the two target images of object detected shot by left and right cameras after respectively.
2. For the two images of target regions, filtering out the noise points and fill tiny regions.
3. Calculated respectively the number of the target areas in the left and right images. One target if there is only one, otherwise multiple targets.
4. If the target numbers are the same and their positions are consistent, go to step 5,otherwise, judging in accordance with the relationship of the position around the target area to identify the simultaneous occurrence of the target area, then roughly matching a wide range of region firstly to find out a couple of matched areas.
5. For a couple of target areas, if the area ratio is less than four, then they can be considered as the same target, if not, then do not process.
6. For a couple of target area dose SIFT matching, then get SIFT features.
7. For left and right target images, remove point-to-multi-points matching points and boundary matching points.
8. Find the corresponding position of the matching point in the original video frame.

Match point selection method could avoid the whole graph traversal and shorten the calculating time. T he strategy of roughly matching first then re-fine matching improves the precision of the matching and effectively eliminates redundancies , and is more convenient for multiple-targets matching as well. Figure 5 is the matching result.

(a)



(b)

Figure 5. Matching Results: (a) Matching Points; (b) Matching Points in the Original Image

5. Experimental Results

Placing a binocular camera in a fixed indoor position, together with Weishi V221 double channel video capture card, they constitute a distance measuring system. The video is collected by ourselves, as shown in Figure 4, the size of each image are 768× 576.

Since the actual camera is not a pinhole camera, errors will be generated according to the formula (1). Based on static camera and fixed position, this paper modifies formula (1) by adding the correction parameter γ in revised distance parameter module. Here, γ is about 10.4. Note that, if the camera position is changed, the parameter values need to re-train. Table 2 shows the distance measurement result of three matching points in Figure 5 using modified formula.

Table 2. Calculation Result and Measurement Result

Position in the left image (pix)	Position in the right image (pix)	Measurement result (mm)	Calculation result (mm)
(688,402)	(555,401)	2578	2678
(666,478)	(529,477)	2576	2524
(650,449)	(514,446)	2575	2558

Averaging the calculated distance of three matching points, derived measurement result is 2587 mm, measuring average is 2576 mm, and the error is 11 mm.

Acknowledgements

This project is supported by Scientific research project of the Education Department of Shaanxi Province. (11JK0929).

References

[1] R. T. Collins, A. J. Lipton, H. Fujiyoshi and T. Kanade, Proceedings of the IEEE, vol. 89, **(2001)**, pp. 1456.
[2] T. Hu and M. Huang, International Journal of Hybrid Information Technology, vol. 3, no. 1, **(2010)**.
[3] I. Haritaoglu, D. Harwood and L. S. Davis, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, **(2000)**, pp. 809.
[4] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C. F. Shu and Y. L. Tian, IEEE Signal Processing Magazine, vol. 22, no. 38, **(2005)**.
[5] J. Ferryman, "AVITRACK: Aircraft surroundings, categorised Vehicles & Individuals Tracking for apRon Activity model interpretation & ChecK", In IEEE International Conference on Computer Vision, **(2005)** October; Beijing, China.

[6] J. Ferryman, "AVITRACK: Aircraft surroundings, categorised Vehicles & Individuals Tracking for apRon Activity model interpretation & ChecK", In IEEE International Conference on Computer Vision and Pattern Recognition, **(2005)** June; San Diego, USA.
[7] P. Dunne and B. J. Matuszewski, International Journal of Grid and Distributed Computing, vol. 4, **(2011)**, pp. 71.
[8] Z. Y. Zhang, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, **(2000)**, pp. 374.
[9] R. Hartley and A. Zisserman, Editor, "Multiple View Geometry in Computer Vision", Cambridge University Press, CUP Cambridge UK **(2003)**.
[10] S. Y. You and G. Y. Xu, China Journal of Image and Graphics, vol. 2, **(1997)**, pp. 15.
[11] J. S. Ku, K. M. Lee and S. U. Lee, Pattern Recognition, vol. 34, **(2001)**, pp. 1701.
[12] D. G. Lowe, International Journal of Computer Vision, vol. 60, **(2004)**, pp. 91.
[13] H. Meng and K. Cheng, Journal of Harbin Engineering University, vol. 30, **(2009)**, pp. 649.
[14] D. R. Kisku, P. Gupta and J. K. Sing, International Journal of Multimedia and Ubiquitous Engineering, vol. 5, no. 1, **(2010)**.
[15] S. W. Ha and Y. H. Moon, International Journal of Smart Home, vol. 5, **(2011)**, pp. 17.
[16] M. Stommel, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 3, **(2009)**, pp. 25.

Authors

BI Ping

Bi Ping received the B.Eng. degree in electronic engineering form Xidian University, China, in 2003, and the M.E. degree in circuit and system form Xidian University, China, in 2006, where she is currently working toward the Ph.D. degree in Pattern Recognition and Intelligent System in the Department of Electronic Engineering, Xidian University.

She was a Lecturer in School of Telecommunication and Information Engineering, Xi ' an University of Posts and Telecommunications, Xi' an, China, in 2006.Her main research interests include image processing, super-resolution reconstruction, and machine learning.