Research on Target Tracking Algorithm Based on Parallel Binocular Camera

Ya'nan Xu, Yue Dong
Information Engineering College
Communication University of China
Beijing, China
d_y@cuc.edu.cn

Jieyu Li
Satellite Telecommunication Department
Space Star Technology Co., Ltd.
Beijing, China
lijieyu@spacestar.com.cn

Xianting Wu
Shenyang No.40 High School
Shenyang, China
787747349@qq.com

***Abstract*—The 3D coordinates have depth information, which improves the accuracy of moving target area acquisition. Therefore, a moving target tracking method based on 3D coordinates is proposed. Firstly, Zhang's calibration algorithm is used to calibrate binocular vision camera; secondly, SAD algorithm is used to extract features of the target, BM algorithm is used to stereo match feature points of the target to determine the 3D coordinates of the center of the moving target; finally, Kalman filter is used to achieve target tracking. Experiments and analysis results show that the algorithm can improve the accuracy of moving object recognition and tracking in left and right views, and has good robustness.**

Keywords—binocular vision; camera calibration; stereo matching; target tracking

I. Introduction

With the development of computer vision, the related research on binocular vision has developed rapidly. The recognition and tracking of moving objects based on binocular vision is one of the hot research fields. The application of binocular stereovision in 3D film shooting has great significance and broad development prospects.

The average distance between the left and right eyes of adults is about 65 centimeters, which results in a slight difference in the angle of view when two eyes look at the same thing, that is parallax, as shown in Fig.1. The reason why human beings can produce a stereoscopic visual effect with spatial stereoscopic effect is that there is a certain difference between the left and right eyes due to the existence of binocular parallax, and thus a stereoscopic effect is produced after synthesis in the brain. Stereoscopic imaging technology and stereoscopic display technology both apply the principle of binocular parallax.

The principle of binocular vision camera shooting is as shown in the Fig.1. The distance between the two camera projection centers is B. The two cameras view the same feature point P of the object at the same time. The left and right lenses of the camera respectively obtain the P point. Images, their coordinates on the image are $P_L = (X_L, Y_L)$ and $P_R = (X_R, Y_R)$.

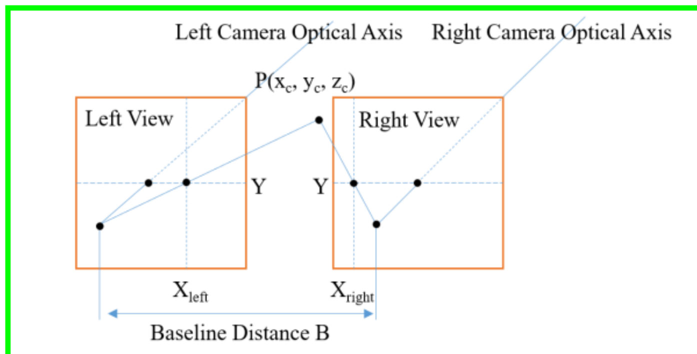


Fig. 1. Parallel camera stereo shooting schematic

In the shooting system, the image plane of the scene is located in the same plane, and the ordinate Y of the feature point P in the image is the same, that is $Y = Y_L = Y_R$, which can be obtained from the triangular geometric relationship:

$$\begin{cases} X_L = f\dfrac{x_c}{z_c} \\ X_R = f\dfrac{x_c - B}{z_c} \\ Y = f\dfrac{y_c}{z_c} \end{cases} \tag{1.1}$$

Parallax can be obtained: $\Delta x = X_L - X_R$. So we can get that the coordinates of the feature point P are:

$$\begin{cases} x_c = \dfrac{B \cdot X_L}{\Delta x} \\ y_c = \dfrac{B \cdot Y}{\Delta x} \\ z_c = \dfrac{B \cdot f}{\Delta x} \end{cases} \tag{1.2}$$

Therefore, the 3D coordinates of a space object can be obtained from the corresponding points on the image planes of the two cameras. This method is a point-to-point operation. As long as there are matching points on the image plane, the corresponding 3D coordinates can be obtained.

978-1-5386-8178-7/19/$31.00 ©2019 IEEE

II. CAMERA CALIBRATION

The position relations of binocular stereo vision system are mainly rotation and translation relations, which are described by rotation matrix column R and translation vector T, respectively. The relative position relations are shown in Fig.2.

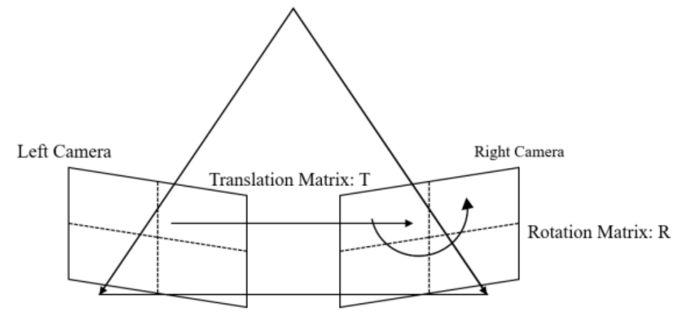


Fig. 2. Position between left and right cameras of parallel binocular camera

Assuming that a point P on the checkerboard used for calibration is X_W in the world coordinate system, and its corresponding points are X_L and X_R in the two cameras coordinate system. The external parameters obtained by single-target calibration of left and right cameras are (R_L, t_L), (R_R, t_R). The corresponding relationship between the camera coordinate system and the world coordinate system can be obtained as follows:

$X_W = (X_L - t_L)/R_L$

$$\begin{cases} X_L = R_L X_W + t_L \\ X_R = R_R X_W + t_R = \dfrac{(X_L - t_L)R_R}{R_L} + t_R \end{cases} \quad (2.1)$$

From the upper formula, we can get the following conversion formulas for left and right cameras:

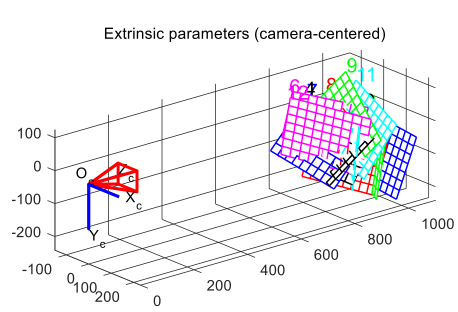
$$X_R = RX_L + T \quad (2.2)$$

From this, the position relationship of binocular camera can be obtained, that is, the external parameters are as follows:

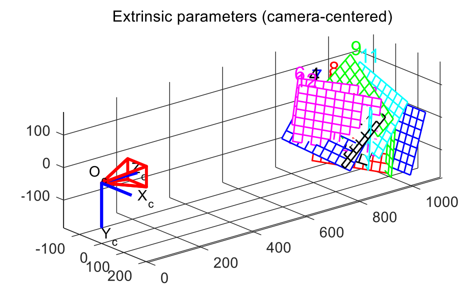
$$R = R_R R_L^{-1}, T = t_R - Rt_L \quad (2.3)$$

Because a large number of chessboard image pairs will be collected during the calibration process, and a set of equations can be obtained for each group of images to obtain a set of external parameters R and T. Therefore, multiple groups of external parameters (R, T) can be optimized non-linearly to obtain the optimal values of external parameters for stereo calibration.

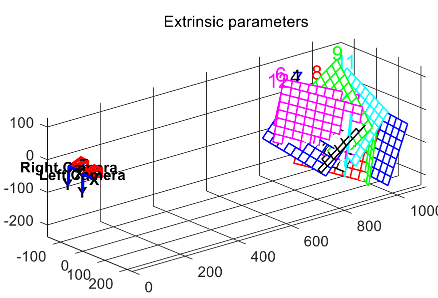
Fig.3 shows the position of the image relative to the camera. Firstly, the external parameters of the left and right cameras are obtained by single target calibration, and then the external parameters R and T of the right camera relative to the left camera are obtained by stereo calibration.



a. Position between calibrated image and left camera



b. Position between calibrated image and right camera



c. Position between binocular image and camera

Fig. 3. Position between calibrated image and parallel binocular camera

III. FEATURE EXTRACTION AND MATCHING

The purpose of feature point matching is to establish point-to-point matching relationship between two images. Based on the consideration of real-time shooting process, SAD algorithm is used to extract feature points:

$$D(i,j) = \sum_{S=1}^{M}\sum_{t=1}^{N} | S(i+s-1, j+t-1) - T(s,t) | \quad (3.1)$$

The fast BM algorithm is used for stereo matching. BM algorithm uses the right-to-left comparison method. At the same time, two heuristic jump rules, bad character algorithm and good suffix algorithm, are introduced to determine the step size of template moving to the right. Its basic process is as follows:

The text string T is the image gray data to be matched and the text string P is the image gray template to be matched, which is called pattern string. First align T with P left, then compare it from right to left, as shown in Fig.4.

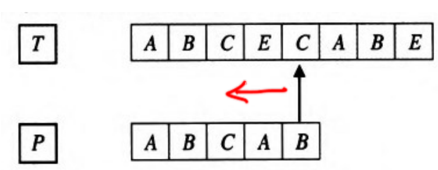


Fig. 4. Text string T and pattern string P matching

If the two do not match, BM algorithm uses two heuristic rules, namely the good suffix algorithm and the bad character algorithm, to calculate the moving step of pattern string P until the whole matching process is over. The definition of bad character algorithm is shown in Fig.5.

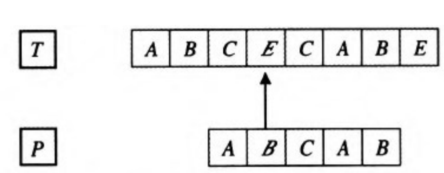


Fig. 5. Bad Characters Match with Good Suffixes

In Fig.5, the first mismatched character arrow connects E and B as bad characters, and the matched part of CAB is a good suffix. BM algorithm is inconsistent in the right-to-left matching in the figure above. At this time, according to the bad suffix algorithm, two cases are needed to deal with:

- If the mismatched character E in T does not appear in pattern P, it can be understood that the m-length string starting from E cannot match P. At this time, P can be jumped directly to E to match the following content.
- If E appears in a field that is not matched in pattern P, the character E is aligned. After two adjustments, the right-to-left matching is performed again. The definition of the suffix algorithm is shown in Fig.6.

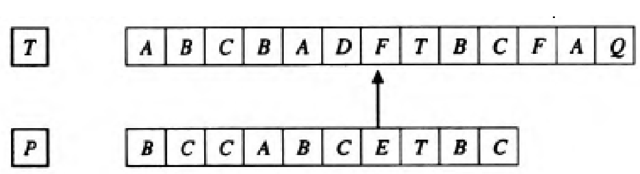


Fig. 6. Application of Suffix Algorithms

In Fig.6, the good suffix is CAB. When a character mismatch is found, some characters have been successfully matched. The following two situations are discussed:

- According to the field T', which is the same as CAB but different from the previous character of CAB in P, the CAB in T is aligned by moving P.
- If there is no T', then the maximum prefix in P is searched, and this prefix is the longest suffix of CAB.

IV. Target Tracking

Kalman filtering algorithm is a linear discrete filtering algorithm. It is a discrete optimal regression data processing method. For a linear system, it can be expressed by a stochastic process equation of motion.

$$X(k) = AX(k-1) + BU(k) + W(k) \tag{4.1}$$

The observational equation of the system is

$$Z(k) = HX(k) + V(k) \tag{4.2}$$

The one-step prediction equation is

$$X(k \mid k-1) = AX(k-1 \mid k-1) + BU(k)$$

The one-step covariance equation is

$$P(k \mid k-1) = AP(k-1 \mid k-1)A^{T} + Q \tag{4.3}$$

$P(k \mid k-1)$ in the formula corresponds to the covariance of the prediction equation, and the optimal estimation equation matrix is

$$X(k \mid k) = X(k \mid k-1) + kg(k)(Z(k) - HX(k \mid k-1)$$

The Kalman gain equation matrix is

$$kg(k) = P(k \mid k-1)H^{T} / (HP(k \mid k-1)H^{T} + R)^{T} \tag{4.4}$$

The covariance updating equation is

$$P(k \mid k) = (I - kg(k)H)P(k \mid k-1)$$

Formula (4.1) is the unit matrix. Kalman filter can realize smoothing, filtering and prediction, and update the autoregressive function under the condition of minimum mean square error. According to the Kalman filtering principle, only the current data need to be processed in real time, but the data of the previous state need not be saved. Therefore, the motion state model of the system must be established before moving target tracking.

The overall algorithm flow is shown in Fig.5.

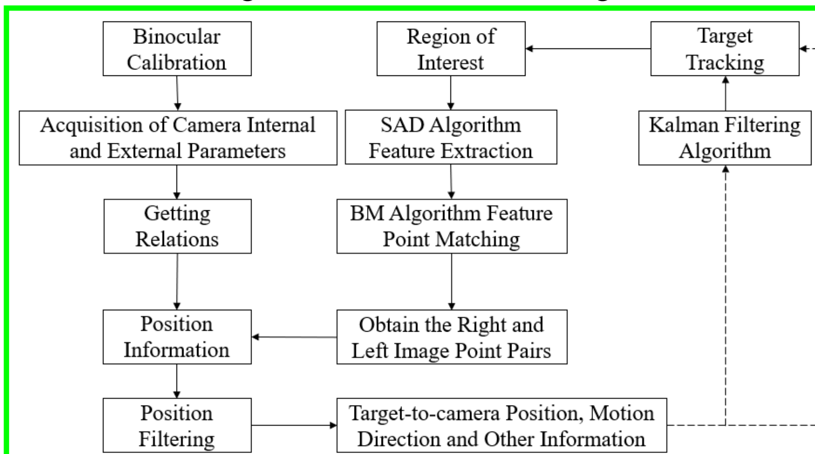


Fig. 7. Overall algorithm flow

V. Conclusion

The binocular camera is used to capture the gray-scale image video sequence with resolution 1200 x 1600. The simulated CPU model is AMD-A10-5750M, the main frequency is 2.5 GHz, the operating system is Windows 10, and the memory is 8 GB, which is realized by OpenCV3.4.1. Fig. 8 is the result of binocular camera calibration. Fig. 9 is the depth information and foreground extraction effect of tracking target. Fig. 10 is the image captured by the left camera of frame 9, frame 25, frame 40, frame 70, frame 80 and frame 135, respectively. The blue rectangular frame is the target tracking frame, and the blue × is the next Central point of the predicted tracking target. It can be seen that the tracking effect is good. In order to test the accuracy of the algorithm better, this paper selects the left view and tests the moving objects in the left view with the classical Kalman filter moving object recognition and tracking algorithm, and then compares it with the proposed algorithm. In Fig.11, a continuous video is tested with a total of 500 frames. From the test results, we find that the accuracy of the proposed algorithm is better than the classical Kalman filtering algorithm in 90% of the video frames. We can draw the following conclusions: The tracking accuracy of the proposed algorithm based on 3D coordinates and Kalman filtering is better than that of the classical Kalman filtering algorithm for moving object recognition and tracking. In binocular vision environment, only the left-right disparity mapping is used to eliminate the mismatching points of left-right view matching, so as to improve the left-right view matching. Corresponding to the accuracy of moving target recognition and tracking, it shows that the proposed algorithm has good robustness and practicability.

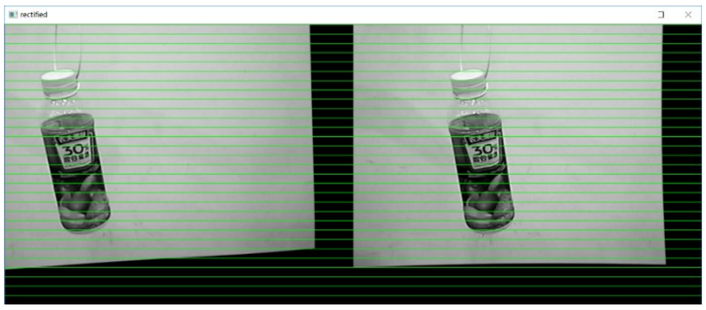


Fig. 8. Calibration results of parallel binocular camera

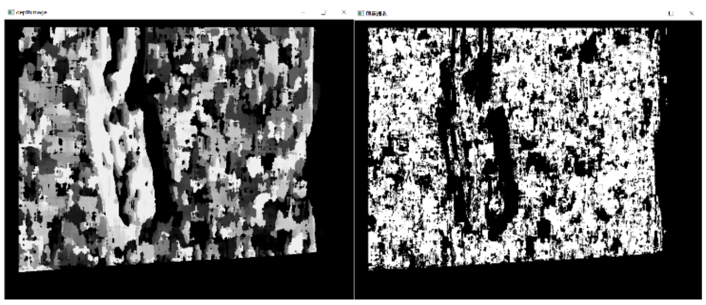


Fig. 9. Tracking target depth information (left) and foreground extraction effect (right)

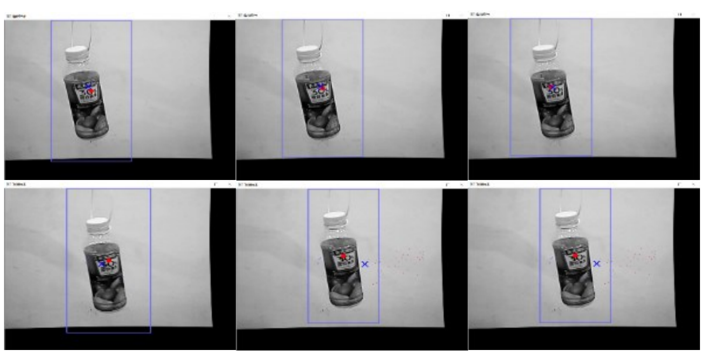


Fig. 10. Target tracking results

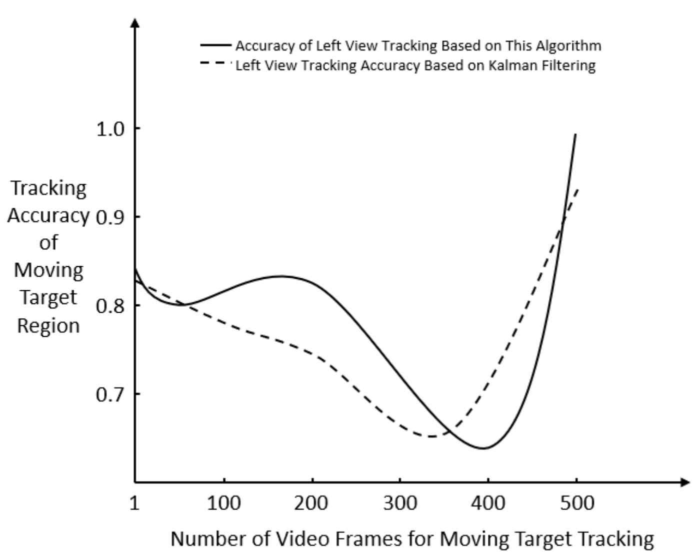


Fig. 11. Comparison of tracking accuracy between left view algorithm and Kalman filter algorithm

This paper presents an algorithm for binocular vision moving object recognition and tracking based on 3D coordinates of parallel binocular camera and Kalman filtering. The biggest difference between the algorithm and the classical Kalman filtering algorithm is that by calculating the disparity mapping between left and right views in the matching process of moving objects in left and right views, the 3D coordinate relationship is adopted. The invalid matching points of left and right views are eliminated, and the disparity mapping is used to increase the accuracy of moving target region acquisition, and then the accuracy of moving target recognition and tracking in left and right views is improved by using the method of moving target region growth based on 3D coordinate centroid of parallel binocular camera. However, there are still some drawbacks in this paper, such as the large influence of illumination conditions on moving object recognition and tracking, the poor effect of multi-target recognition, and the inability to achieve the function of moving object tracking under 3D conditions in depth view. These are the problems that this paper will continue to study in depth.

REFERENCES

[1] Wang Y, Xiao B, Wang Q (2010) Stereo Matching for Face Images Based on Feature Points. Chinese Journal of Electronics 2010(2): 241-244.

[2] Dong Y, Chai J, Du H, Wang C. Stereo Film and Television Shooting Device: CN. Patent 201310315515.6. 2013-07-25.

[3] Dong Y, Xiao H, Yang Q, Wang C, Bao N. A Communication Module for Parallel Stereo Photography: CN. Patent 103369250A. 2013-10-23.

[4] J Kim, JH Jung, C Jang, et al. (2013) Real-time capturing and 3D visualization method based on integral imaging. Optics Express, vol. 16, pp. 18742-18753.

[5] Zhou Z, Xu M, Fu W, Zhao J (2013) Object Tracking and Positioning Based on Stereo Vision. Sensors, Measurement and Intelligent Materials, Pts 1-4, vol. 303-306, pp. 313.

[6] Peng M, Di K, Liu Z (2014) Adaptive Markov random field model for dense matching of deep space stereo images. Journal of Remote Sensing, 2014(1): P77-89.

[7] Zhang L, Xu J (2014) Target tracking and locating for UAV based on binocular stereo vision. Computer Engineering and Applications, vol. 24, pp. 27-31, 53.

[8] D.Martin, KF.Shahbaz, F.Michael, W.Joost (2014) Adaptive Color Attributes for Real-Time Visual Tracking. 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1090-1097.

[9] Men Y, Zhang G, Men C, Li X, Ma N (2015) A Stereo Matching Algorithm Based on Four-Moded Census and Relative Confidence Plane Fitting. Chinese Journal of Electronics, 2015(4): P807-812.

[10] Chuang M; Hwang J; Williams K; Towler R (2015) Tracking Live Fish From Low-Contrast and Low-Frame-Rate Stereo Videos. IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, pp. 167-179.

[11] Dong Y, Ren H, Dong J, Wang L (2015) Study on Wireless Network Communication in Stage Hydraulic Monitoring System Based on Internet of Things. Discrete Dynamics in Nature and Society, vol. 2015, Article ID 652183, 9 pages.

[12] W. Daniel, K.Wilfried, EN.Corinna (2016) Preceding Vehicle Tracking in Stereo Images via 3D Feature Matching. DAAAM International Scientific Book, vol. 1, pp. 997-1003.

[13] Wang B, Dong Y, Dong J, Wang L (2016) Structure design and system integration of 3D camera platform. Microsystem Technologies Sci 22(10): 2455–2462.

[14] A.Schoob, D.Kundrat, LA.Kahrs(2017) Ortmaier TStereo vision-based tracking of soft tissue motion with application to online ablation control in laser microsurgery. Medical Image Analysis, vol. 40, pp. 80-95.

[15] Chen C, Yang Q, Li D (2018) Moving objective tracking based on 3D coordinates. Applied Science and Technology, vol. 45, pp. 23-28.