Binocular Stereoscopic Vision Algorithm Based on Improved SIFT Feature

Jian Liu[1, a] and Yao Lu [2, b *]

[1]Department of Information and Control Engineering, Shenyang Jianzhu University,

Shenyang 110168, Liaoning, China.

[a]644630900@qq.com, [b]471418469@qq.com

*The Corresponding author

Keywords: The improved SIFT algorithm; Binocular stereo; Object location; Feature matching

Abstract. Stereo matching is the most important step in binocular vision, the traditional regional stereo matching to obtain the target three-dimensional information is slow and inaccurate. This paper presents an improved SIFT algorithm. Firstly, making the epipolar constraint on the left and right image; secondly, selecting the ROI of target from the left of binocular images, and reducing running time by reducing the dimension of feature vectors and accelerating the matching speed by using BBF algorithm based on KD tree; finally, removing the false matching by using RANSAC algorithm. The improved SIFT algorithm can get the target's feature points quickly and accurately, so the 3D coordinates can be calculated by the triangulation method speedy.

Introduction

Binocular stereoscopic vision[1-4] is based on the matching point parallax. The parallax error obtained by the traditional region matching method[5-6]is large, so using the feature matching algorithm[7-10]. The SIFT feature matching algorithm[11]proposed by Lowe has some effect in anti-rotation, translation, scale changes, partial illumination and affine changes. Based on the study of SIFT algorithm, this paper proposes an improved SIFT feature matching and target localization algorithm, the algorithm can quickly and accurately obtain the target three-dimensional information.

Binocular Vision Positioning Principle

The binocular vision follows the principle of the human eyes to observe the object, uses two cameras to observe the same target object from different angles, and obtains two images simultaneously, obtains the three-dimensional coordinate information of the target through the relative parallax of the target. As shown in Fig. 1, point P in space, set the left camera coordinate system is $O_1x_1y_1z_1$,the image coordinate system is $O_LX_LY_L$, the effective focal length is f_L, the right camera coordinate system is $O_2x_2y_2z_2$, the image coordinate is $O_RX_RY_R$, the effective focal length is f_R.The points P on the left and right image coordinate systems are P_1 and P_2.
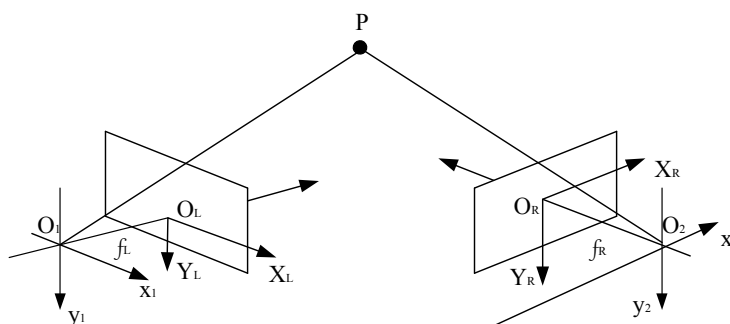

Figure 1.   Finite Schematic diagram of binocular vision

In the world coordinates, the coordinates of the point P (X, Y, Z) can be converted through the

image coordinate system, Let the coordinates of point P in the image coordinates be (u, v), that is, the point of feature extraction, the conversion formula is:

$$
\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R_{3\times3} & T_{3\times1} \\ O^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = M_1 M_2 \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}
\tag{1}
$$

Where M_1 and M_2 are the internal and external parameters of the camera, M is the projection matrix, R is the rotation matrix between the left and right cameras: $R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$, T is the translation matrix between the left and right cameras, $T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$, The internal and external parameters of the camera are obtained by the Zhang's calibration method[12-13].

Feature Point Extraction

Improved SIFT Algorithm. The traditional SIFT algorithm generates a large number of feature points, and needs to generate 128-dimensional feature vector, so the calculation is relatively large, real-time poor. In order to improve the real-time of SIFT algorithm, the following improvements are used.

To calculate the parallax of the target point on the left and right views, the two corresponding points on the left and right views should be matched first. However, it is very time-consuming. In order to reduce the matching search range, pole line constraints are used to make the matching of the corresponding points reduce from two-dimensional search to one-dimensional search. The effect of pole line constraints is to make the corresponding lines of the two images just on the same level. Thus, any point on an image must have the same line number as its corresponding point on another image, only one-dimensional search in the line is required can match the corresponding point.

The target is selected by manually selecting the target ROI, thereby reducing the matching of other non-relevant areas.

The dimension of the feature point descriptor is changed to 32 dimension without reducing the matching effect. As shown in Fig. 2 (a), the center of the 8x8 window is the key point, the middle black point is the location of the key point, each small cell represents a pixel in the scale space where the key point area is located, the arrow direction represents the gradient direction of the pixel, the arrow length represents the gradient modulus, the blue circle represents the range of Gaussian weighting, and then in each 4×4 window to calculate the eight directions histogram, draw a cumulative value in each gradient direction to form a seed point. As shown in Fig. 2 (b), one feature point is composed of four seed dots, each with eight directional vector information, and a 32-dimensional feature descriptor is formed.

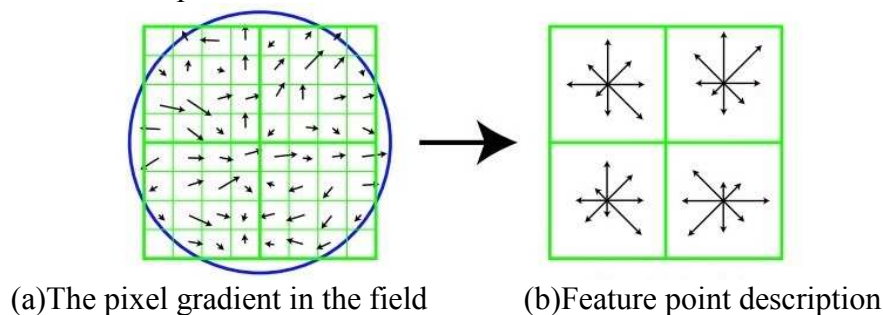(a)The pixel gradient in the field   (b)Feature point description

Figure 2. Finite Generate the SIFT feature point descriptor

On the basis of KD tree, BBF (Best Bin First) algorithm is adopted to reduce the complexity of matching search, BBF is an improved KD tree nearest neighbor query algorithm. The "backtracking" in the search process is determined by the "query path" and does not take into account the nature of the data points on the query path. BBF query idea is to sort the node on the "query path" in the light of separate the hyperplane(called Bin)and the distance of the inquiry point. Backtracking always begins with the tree node with the highest priority（Best Bin）.In addition, the BBF also sets a run timeout limit. When all nodes in the priority queue are checked or out of time, the algorithm returns the best result currently found as an approximate nearest neighbor. Using the BBF method can extend the KD tree to a high-dimensional dataset, thus further speed up the matching speed.

Using RANSAC algorithm to remove the wrong match point. RANSAC can estimate the parameters of the mathematical model iteratively from a set of observations that contain "off-site" observations. RANSAC algorithm to eliminate the wrong match point method is: A transformation matrix is found so that the feature points of the obtained left and right images coincide with this transformation relation, called the projective matrix H. The matrix H is a 3×3 matrix containing 8 degrees of freedom, it is calculated by at least four pairs of matching points in two planes, but the same plane of the three points cannot be coplanar.

Improved SIFT Algorithm Comparison. As shown in Fig. 3(a), for the SIFT algorithm that has not been improved, Since the desired target is not selected, resulting in matching some meaningless match point. Since the dimension of the generated eigenvector is 128 degrees, the matching speed is very slow and the RANSAC algorithm is not added to remove the mismatch, and it can be seen that some points are mismatched. Through the above four methods to improve the algorithm matching results shown in Fig. 3(b),The feature points found at this time are targeted, will not find the feature points which has nothing to do with the target, So that the desired target ROI region is accurately matched, And after adding the RANSAC algorithm, the mismatched points are almost completely eliminated, The improved SIFT match speed is significantly improved, to ensure the late three-dimensional positioning accuracy and real-time.

(a)SIFT algorithm   (b)Improved SIFT algorithm

Figure 3.   Finite Comparison of the algorithm

Target location

Experiment with two tea pots for measuring targets verify the accuracy of the three-dimensional coordinates (X,Y,Z), respectively. Suppose there is a reference point P (X,Y,Z) in the space, The left and right image coordinates of P are (u_L, v_L), (u_R, v_R),Their projection matrices are M_L and M_R, respectively. Then, according to equation (1) we can see

$$Z_{C_1}\begin{bmatrix} u_L \\ v_L \\ 1 \end{bmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = M_L\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (2)$$

$$Z_{C2}\begin{pmatrix} u_R \\ v_R \\ 1 \end{pmatrix} = \begin{pmatrix} m'_{11} & m'_{12} & m'_{13} & m'_{14} \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = M_R\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3}$$

By (2) and (3) eliminate Z_{C1}, Z_{C2} get

$$\begin{pmatrix} u_L m_{31} - m_{11} & u_L m_{32} - m_{12} & u_L m_{33} - m_{13} \\ v_L m_{31} - m_{21} & v_L m_{32} - m_{22} & v_L m_{33} - m_{23} \\ u_R m'_{31} - m'_{11} & u_R m'_{32} - m'_{12} & u_R m'_{33} - m'_{13} \\ u_R m'_{31} - m'_{21} & u_R m'_{32} - m'_{22} & u_R m'_{33} - m'_{23} \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} m_{14} - u_L m_{34} \\ m_{24} - v_L m_{34} \\ m'_{14} - u_R m'_{34} \\ m'_{24} - v_R m'_{34} \end{pmatrix} \tag{4}$$

(X, Y, Z) of the formula (4) is obtained by the least squares method, Whereby the coordinates of the object matching point on the image can be obtained, the spatial coordinates of the target can be obtained to achieve the target location. In order to verify the three-dimensional coordinate information (X, Y, Z), respectively, when the two tea cans and cameras away from the 400mm, 600mm, 900mm, 1200mm, multiple measurements to obtain (X, Y, Z) average.

two tea cans placed parallel to the left and right, select the target center position for feature matching, The actual distance between the horizontal X coordinates of the two tea pots is 150 mm. When verifying the difference between the X coordinates, to ensure that the two tea cans Y coordinate values are consistent or similar. The test results are shown in Table 1.

two tea cans placed up and down, select the center position pf the target for feature matching, The actual distance between the vertical Y coordinates of the two tea pots is 110 mm. When verifying the difference between Y coordinates, to ensure that the X coordinate values of the two tea pots are the same or similar. The test results are shown in Table 2.

It can be seen from Table 1 that when the distance is 400 mm, the measured distance Z 'between the camera and the tea pot is close to the actual distance Z, and the X-coordinate difference between the two tea cans is closest to the actual distance of 150 mm. With the tea pot and camera farther and farther, the error is getting bigger and bigger. Similarly, from Table 2 we can see that when the distance between the camera and the tea cans is closer to Z, the difference between the two tea cans is closest to the actual distance of 110 mm. With the tea pot and camera farther and farther, the error is also growing. The cause of the error is the inaccurate calibration of the internal and external parameters, the camera's own pixel limits and the camera distortion are not completely eliminated.

Table 1　Distance between the X coordinates of the tea pot

Actual distance Z/mm	Left tea can coordinates	Right tea can coordinates	Test distance Z'/mm	Test distance X'/mm	Z error/ %	X error/ %
400	(140,60.5,406)	(-11.7,60.4,403)	403	151.7	0.75	1.13
600	(113.1,-35.5,605.7)	(-40.19,-36.4,610.1)	605	153.3	0.83	2.15
900	(185.2,-84.4,918)	(40.12,-85.3,920.8)	918	145	2	-3.33
1200	(198.5,-107.9,1220)	(58.15,-107,1219.3)	1220	130.3	1.6	-6.7

Table 2　Distance between the Y coordinates of the tea pot

Actual distance Z/mm	Left tea can coordinates	Right tea can coordinates	Test distance Z'/mm	Test distance X'/mm	Z error/ %	X error/ %
400	(12.3,57,403)	(12.9,-54,406)	403	111	0.75	0.9
600	(38,47.6,608)	(38,-60,610.1)	608	107.6	1.3	-2.1
900	(101,40.3,902.8)	(101,-73.9,898.3)	902	114.2	0.22	3.8
1200	(129.1,5.3,1220)	(126,-99,1223)	1220	104.3	1.6	-5.7

Conclusion

The higher the accuracy of the feature point extraction and matching, the higher the accuracy of the binocular vision measurement system, which requires a better matching algorithm, And real-time of matching algorithm must be better. As the traditional SIFT algorithm in the image matching in the high complexity, real-time relatively low, and there are more false matching, so the algorithm is difficult to have advantages in binocular positioning. In this paper, we propose a SIFT image matching algorithm that combines the target region selection, firstly, making the epipolar constraint on the image, and then reduce the feature point descriptor dimension, BBF algorithm to speed up the matching speed and RANSAC to remove the wrong match, thereby reducing the complexity of feature extraction and matching and improve the accuracy of matching, and ultimately improve the target positioning real-time and accuracy.

Acknowledgements

References

[1] H.Meng, K.Cheng. Object location technique for binocular stereo vision based on scale invariant feature transform feature points [J]. Journal of Harbin Engineering University, 2009,30(6): 649-652.

[2] Y.Y.Li, M.Z.Jia. Object location technique for binocular stereo vision based on harris-SIFT algorithm [J]. Video Engineering, 2013, 37(9): 23-26.

[3] X.H.Wang, W.P.Fu. Research on recognition and location system of work-piece based on binocular vision[J]. Manufacturing Automation, 2010, 32(11): 129-131.

[4] X.F.Li, Y.N.Song and R.H.Xu. Tracking and positioning of ship based on binocular vision[J]. Journal of Nanjing University of Information Science & Technology, 2015,7(1): 46-52.

[5] Y.H.Wang, B.Jin. Depth information measurement based on binocular stereo vision[J]. Mechanical & Electrical Engineering Technology, 2015, 44(3): 34-38.

[6] T.T.Wang,G.Li, and J.ZHAO. Detection, tracking and positioning of moving target based on binocular vision[J].Machinery & Electronics, 2015, 6: 73-76.

[7] W.Sun, Y.C.Zhong and Z.Tan. Research on multi-featured fusion for indoor scene recognition[J]. Journal of Guangdong University of Technology, 2015, 32(1): 75-79.

[8] Y.Lei, J.F.He. SAR image target matching based on the improved SURF algorithm[J]. Journal of Guangdong University of Technology, 2014, 31(1): 65-69.

[9] J.B.Yang, F.Y.Hu and G.Zhu. Remote sensing image registration based on improved SURF[J]. Electronic Measurement Technology, 2012, 35(3): 69-72.

[10] H.Bay, A.Ess and T.Tuytelaars. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[11] D.G.Lowe. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision,2004, 60(2): 91-110.

[12] Z.Zhang. A flexible new technique for camera calibration[J].Transactions on Pattern Analysis and Machine Intelligence,2000, 20(11): 1330-1334.

[13] Z.Zhang. Flexible camera calibration by viewing a plane from unknown orientations[C] ∥ Proceeding of IEEE Inter-national Conference on Computer Vision. [S.l.]: IEEE, 1999,11(1): 666-673.