2009 International Conference on Measuring Technology and Mechatronics Automation

Stereo Cameras Self-calibration Based on SIFT

Liu Ran, Zhang Hua, Liu Manlu, Xia Xianfeng, Hu Tianlian
Robotics Laboratory, School of Information Engineering, Southwest University of Science and Technology
MianYang, SiChuan, China
E-mail: 2003liuran@163.com

***Abstract*—At present, a new algorithm of feature matching-SIFT has become a hot topic in the feature matching field, whose matching ability is strong, and could process the matching problems with translation, rotation and affine transformation among different images, and to a certain extent is with more stable feature matching ability for images which are captured from random different angles. In this paper, single camera is first calibrated using plane chessboard based on OpenCV, in order to overcome shortcomings in traditional and previous self-calibration methods, SIFT algorithm is proposed to calibrate stereo cameras after two cameras intrinsic parameters are calibrated. Fundamental matrix is gained through several matching points in two images using SIFT feature matching method, combined with intrinsic parameters, we can compute essential matrix. Translation matrix and the rotation matrix of stereo cameras can be resolved through SVD of essential matrix known as Huang- Faugeras' constrains. Experiment results show that our method can calibrate relationship of stereo cameras accurately, and be able to calibrate two cameras in any circumstances. The algorithm has strong adaptability and robustness, but the expense time needs to be further improved.**

***Keywords*—SIFT, Self-calibration, OpenCV, SVD , Huang-Faugeras' constrains**

I. INTRODUCTION

Image acquisition process can be seen as a process carried out an objective world scene into projection transformation, and the projection can be described by imaging transformation. Imaging transformation relate to transformation of different coordinate. We can get different imaging model under mutual relations of these different coordinate system. Imaging models establish expressions that calculate image plane coordinates or image coordinates according to the real world point.

Camera calibration technology occupies a very important position in the study of stereo vision. Accurate calibration of the camera intrinsic and extrinsic parameters not only can directly improve the measurement accuracy, but also lay a good foundation for the subsequent three-dimensional match and three-dimensional reconstruction[7, 10]. Traditional calibration methods need a known structure, high precision calibration object as a space reference, and establish restriction of camera model parameters through the corresponding relations between the space point the and image point, and use certain algorithm to get inside and outside parameters of model camera. This calibration method can not be used in the occasions that do not use calibration object.

*This work is partially supported by National Defence Key Subject Laboratory of SWUST Grant #07JGZB10 and NSF Grant #60574076.

Self-Calibration overcomes the shortcomings of the traditional method. It does not need calibration object, and calibrate directly relying on the relationship of corresponding points of the number of images solely. So flexibility is huge, but accuracy and robustness is low.

The well known linear method is developed by Longuet-Higgins[2] and Tsai and Huang[13] independently. It extracts the R and t from the so-called Essential Matrix (in short E) between two calibrated perspective images. Another linear method[12] is proposed by Hartley using the Singular Value Decomposition (in short SVD) on E to recover the motion parameters. In Hartley's paper, the uncalibrated case is also considered, however, the answer is negative since it is impossible to recover the camera intrinsic parameters and motion parameters from uncalibrated views except for some special simple camera models or some special motions.

In the paper, OpenCV is used to calibrate a single camera. In order to solve shortcomings of the traditional calibration method using specific calibration block, SIFT feature matching principle is used in stereo camera self-calibration after getting two cameras internal parameters. The method gets rotation and translation matrix of two cameras. 3D information of the point that is in the world coordinate is come back according to the corresponding match point in the two cameras. It is the basis for three-dimensional reconstruction.

II. PRINCIPLE OF STEREO CAMERAS SELF-CALIBRATION

Proper model of image should be selected for calibrating camera, then equal the intrinsic parameters of camera. Calibration arithmetic of OpenCV is based on pinhole model. That is, a scene view is formed by projecting 3D points into the image plane using perspective transformation. The formulations is:

$$s \times p = A \times [R|t] \times P \quad \text{or}$$

$$s \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_u & 0 & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Where *(X,Y,Z)* are coordinates of a 3D point in the world coordinate space; *(u, v)* are coordinates of point projection in pixels. A is called a camera matrix, or matrix of intrinsic parameters; (u_0,v_0) is a principal point(that is usually at the image center), a_u,a_v are focal lengths expressed in pixel-related units. The matrix of intrinsic parameters does not depend on the scene viewed and, once estimated, can be re-used (as long

978-0-7695-3583-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ICMTMA.2009.338

as the focal length is fixed). The joint rotation-translation matrix $[R|t]$ is called a matrix of extrinsic parameters. It is used to describe the camera motion around a static scene, or vice versa, rigid motion of an object in front of still camera.

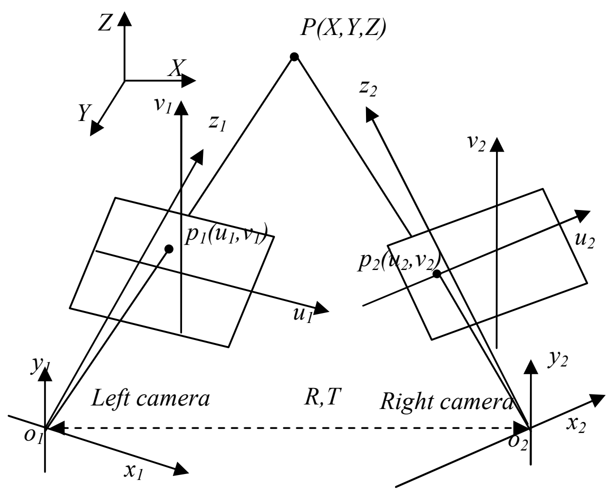

Fig. 1 Stereo Camera Calibration

As shown in Fig.1, A_1 and A_2 are intrinsic matrix of left and right camera, the camera coordinate of left camera is regarded as the world coordinate, so the translation matrix of left camera is:

$$\boldsymbol{M_1} = \boldsymbol{A_1}(\boldsymbol{I} \mid \boldsymbol{O^T}) = (\boldsymbol{A_1} \mid \boldsymbol{O^T}) \tag{6}$$

Suppose that the structure parameters between left camera and right camera should be R and T, thus the translation matrix of right camera is:

$$\boldsymbol{M}_2 = \boldsymbol{A}_2 \cdot (\boldsymbol{R} \mid -\boldsymbol{R} \cdot \boldsymbol{T}) = (\boldsymbol{A}_2 \cdot \boldsymbol{R} \mid -\boldsymbol{A}_2 \cdot \boldsymbol{R} \cdot \boldsymbol{T}) \tag{7}$$

Given skew symmetric matrix of T:

$$S = [T]_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{8}$$

Given F as the fundamental matrix which is from left camera to right camera, and E as essential matrix, thus [8, 10, 11],

$$\boldsymbol{F} = (\boldsymbol{A_2^T})^{-1} \boldsymbol{R S A_1^{-1}} \tag{9}$$

For essential matrix $\boldsymbol{E} = \boldsymbol{A}_2^{\boldsymbol{T}} \cdot \boldsymbol{F} \cdot \boldsymbol{A}_1$, so

$$E = R \cdot S \tag{10}$$

It is well known that a 3×3 real matrix E can be factored as the product of a rotation R and a nonzero skew symmetric matrix S if and only if E has two equal nonzero singular values and one singular values equal to 0.

Suppose the 3×3 real matrix E can be factored into a product RS where R is orthogonal and S is skew symmetric. Let $E=UDV^T$ where $D=diag(k,k,0)$, k is the value of equal nonzero singular value, Then up to a scale the factorization is one of the following:

$$\begin{cases} \boldsymbol{S \approx VZV^T} \\ \boldsymbol{R \approx UWV^T} \ \text{or} \ \boldsymbol{R \approx UW^T V^T} \\ \boldsymbol{E \approx RS} \end{cases} \tag{11}$$

Where the ≈ indicates equality up to a scale factor,

$$W = \begin{vmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}, Z = \begin{vmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

The value of translation vector can't be obtained from the decomposition of essential matrix E. The translation vector from the decomposition is scale with the fact translation vector. If the translation vector from the decomposition is T, well then fact translation vector can be given by $\boldsymbol{T'}=\lambda \boldsymbol{T}$.

$$\text{Let } A_1 = \begin{vmatrix} a_{u1} & 0 & u_{01} \\ 0 & a_{v1} & v_{01} \\ 0 & 0 & 1 \end{vmatrix}, A_2 = \begin{vmatrix} a_{u2} & 0 & u_{02} \\ 0 & a_{v2} & v_{02} \\ 0 & 0 & 1 \end{vmatrix}, R = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix},$$

$$-RT = (T_x, T_y, T_z)^T, -RT' = -lRT = (lT_x, lT_y, lT_z)^T,$$

$$\text{Then } M_1 = (A_1 \mid O^T) = \begin{vmatrix} a_{u1} & 0 & u_{01} & 0 \\ 0 & a_{v1} & v_{01} & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} \tag{12}$$

$$M_2 = A_2(R \mid -RT) = \begin{vmatrix} a_{u2} & 0 & u_{02} \\ 0 & a_{v2} & v_{02} \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} r_{11} & r_{12} & r_{13} & \lambda T_x \\ r_{21} & r_{22} & r_{23} & \lambda T_y \\ r_{31} & r_{32} & r_{33} & \lambda T_z \end{vmatrix} \tag{13}$$

Homogeneous coordinate of one point, which coordination is (X,Y,Z) in the world coordination , is $(u_1,v_1,1)$,$(u_2,v_2,1)$ in two image view of stereo cameras. Let

$$d_x = [a_{u2} r_{11} + (u_{02} - u_2) r_{31}](u_1 - u_{01}) / a_u,$$

$$d_y = [a_{u2} r_{12} + (u_{02} - u_2) r_{32}](v_1 - v_{01}) / a_v,$$

$$d_z = a_{u2} r_{13} + (u_{02} - u_2) r_{33},$$

$$d_T = a_{u2} T_x + (u_{02} - u_2) T_z,$$

$$c_x = [a_{v2} r_{21} + (v_{02} - v_2) r_{31}](u_1 - u_{01}) / a_u, so:$$

$$c_y = [a_{u2} r_{22} + (u_{02} - u_2) r_{32}](v_1 - v_{01}) / a_v,$$

$$c_z = d_z = a_{u2} r_{13} + (u_{02} - u_2) r_{33},$$

$$c_z = a_{v2} r_{23} + (v_{02} - v_2) r_{33},$$

$$c_T = a_{v2} T_y + (v_{02} - v_2) T_z,$$

$$\begin{cases} X=\dfrac{-(u-u_{01})(d_T+c_T)\lambda}{a_{u1}[(d_x+d_y+d_z)+(c_x+c_y+c_z)]} & (14) \\ Y=\dfrac{-(u-v_{01})(d_T+c_T)\lambda}{a_{v1}[(d_x+d_y+d_z)+(c_x+c_y+c_z)]} & (15) \\ Z=\dfrac{-(d_T+c_T)\lambda}{(d_x+d_y+d_z)+(c_x+c_y+c_z)} & (16) \end{cases}$$

From expression (14, 15, 16), we know that the influence of 3D reconstruction, caused by scalar λ of translation vector, is linear. So based on fundamental matrix F, we can gain essential matrix according to

$$E = A_2^T F A_1$$

Once E is known the camera motion, i.e., the relative rotation R and translation t between the cameras can be obtained from the decomposition of E. If we can find two fixed points in the real word and the corresponding image coordinate in the two images, the scale constant can be calculated, and the 3D coordinate of a point in the real word is gained.

III. STEREO CAMERA CALIBRATION PROCEDURE

Step 1: Calibrate two cameras' intrinsic matrix A_1 and A_2. Based on the calibration function in OpenCV, this paper uses plane chessboard as calibration board. In order to improve the success ratio of corner finding, surrounding calibration diamond, there needs one blank diamond area, as shown in Fig. 2 is the 10×10 size plane board used in this paper, where the side of every diamond is 40 mm. The camera only need to capture several images form different directions, and the calibration can be finished with the help of OpenCV. Obviously, more images the camera captures, more accurate calibration results we will gain for adopting Least-squares method. As shown in Fig.2(a) is the original image, and Fig.2(b) is the image when the corners are found successfully.

a b

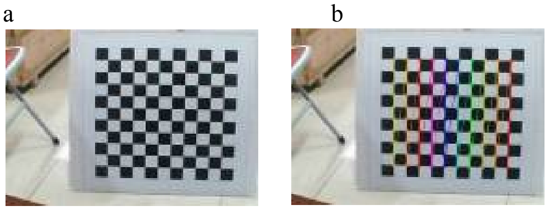

Fig. 2(a) is the original image, (b) is a the result image of (a) where corners are founded successfully.

Step2: Find fundamental matrix. According SIFT keypoints matching principle [14, 15], the corresponding matching keypoints are founded, and using these points fundamental matrix F is solve.

There are two stages of SIFT algorithm. Firstly, generate of sift features, which are foreign to scale, rotation, translation, and affine distortion between images. Secondly, keypoint matching. As shown in Fig. 3 are two images captured from different directions, Fig. 3(a) and Fig. 3(b) is keypoints localization in two images, and Fig. 3(c) is SIFT matching results.

After gain matching points, we use FindFundamentalMat() in OpenCV to find Fundamental Matrix F, here we adopt RANSAC arithmetic to resolve F.

a. c.

b.

Fig. 3 (a) is SIFT Keypoints in left camera image, (b) is SIFT keypoints in right camera image, (c) is SIFT matching results

Step3: Compute essential matrix E. Based on the calibration parameters A_1 and A_2, we can gain E according to $E=A_1FA_2$.

Step4: the relative rotation R and translation t between the cameras can be obtained from the *SVD* decomposition of E. Define Normalization projection matrix of the stereo cameras are M_{cam1} and M_{cam2}:

$$\begin{aligned} M_{cam1} &= A_1^{-1} M_1 = (I \mid O^T) \\ M_{cam2} &= A_2^{-1} M_2 = (R \mid -RT) \end{aligned} \quad (17)$$

Do *SVD* decomposition to E [16]:

$$E = USV^T$$

$$\begin{cases} R=UWV^T & or \quad R=UW^TV^T \\ -RT=u_3 & or \quad -RT=-u_3 \end{cases} \quad (18)$$

Where u_3 is last column of U. So the normalized projection transform matrix M_{cam2} has four possible solutions.

Step5: Get the right solution of normalized projection transform matrix M_{cam2} from these four possible solutions.

$$\begin{cases} w_1p_1=A_1M_{cam1}P \\ w_2p_2=A_2M_{cam2}P \end{cases} \quad (19)$$

Step6: Compute the actual translation matrix:

$$\begin{aligned} M_1 &= A_1 M_{cam1} \\ M_2 &= A_2 M_{cam2} \end{aligned} \quad (20)$$

Step7: Calculate the scale constant according to two fixed distance points in the images.

IV. EXPERIMENTAL ANALYSIS

Based on above calibration principle, we have used Visual studio 6.0 to develop a stereo camera calibration program based on OpenCV1.0, as shown in Fig.4. Calibration program can calibrate single and stereo camera, and the entire calibration process does not require the participation of people.

Fig.4 The interface of stereo camera self-calibration

We calibrate left and right camera separately, and get the intrinsic parameters of two cameras.

We do self-calibration of stereo cameras to determine the cameras relationship after intrinsic parameters of two cameras is calibrated. In order to verify the results of calibration, we use two cameras whose location is fixed to capture images. The resolution ratio of the image is 360×240. We use SIFT feature matching algorithm to find the matching points. When the number of matching points is more than 7, we use these points to calculate the corresponding fundamental matrix, and can get the essential matrix with the help of intrinsic parameters, and decompose essential matrix to get rotation matrix and translation matrix. In order to verify the calibration results of the calibration program, we use SIFT algorithm of this paper and Camera Calibration Toolbox For Matlab of Matlab toolbox in literature [2] to calibrate, and compare results, as shown in table 1.

Camera Parameters	First Experiment		Second Experiment	
	SIFT	Matlab	SIFT	Matlab
Om (R)	[0.063266 -0.18083 0.044896]	[0.04986 -0.19346 0.0626]	[0.001943 -0.085108 -0.12535]	[0.01863 -0.14684 -0.10653]
T	[-0.75455 0.12616 -0.379]		[0.04815 0.35862 -0.27067]	
λ	207.54		221.95	
T'	[-156.599 26.18372 -78.6810]	[-149.97 17.6987 -85.156]	[10.6860 79.59621 -60.0758]	[14.2367 68.36722 -50.4543]

Table. 1 Stereo camera self-calibration results

We can see both the calibration results of two methods are very similar from the table. The calibration results are greatly influenced by essential matrix. The accuracy of SIFT matching points influent the results of calibration. As RANSAC method is used in computing fundamental matrix in this paper, the more matching points, the more accurate the results.

V. COCLUDING REMARKS

In this paper, we get the rotation and translation matrix by SVD of essential matrix. In order to use our calibration algorithm in any environment, the algorithm of feature matching-SIFT is used to search matching points of two images. The experimental results show that it can find the correct solution of *R* and *T* in any environment. In this paper, the wrong matching points is found sometimes when SIFT feature matching is used. The effect will be better, If others algorithm can be combined.

REFERENCES

[1] Zhang Y J. Image Engineering(Second Edition) (M). Beijing: Tsinghua University Press. 2007.5.

[2] Jean-Yves Bouguet. Camera calibration toolbox for Matlab [DB/OL]. 2004. http://www.vision.caltech.edu/bosuguetj/calib_doc.

[3] Zhang Z. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22 (11):1330-1334.

[4] Jean-Yves Bouguet. Visual methods for three-dimensional modeling[D]. Pasadena, California: California Institute of Technology, 1999.

[5] A. Mitiche and etc. A computation analysis of time-varying images. Handbook of Patter Recognition and Image Processing, New York: Academic, 1986.

[6] H.C.Longuet-Higgins. A computer program for reconstructing a scene from two projections. Nature, 293:133–135, Sept. 1981.

[7] J.Q.Fang and etc. Some experiments on estimating the 3d motion parameters of a rigid body from two consecutive image frames. IEEE Trans. PAMI, 6:547–554, 1984.

[8] J.W.Roach and etc. Determining the movement of objects from a sequence of images. IEEE Trans. PAMI, 2(6):554–562, 1980.

[9] J.Y.Weng and etc. Motion and structure from two perspective views: algorithms, error analysis and error estimation. IEEE Trans. PAMI, 11(5):451–476, 1989.

[10] R.I.Hartley. Estimation of relative camera positions for uncalibrated cameras. Proc. of ECCV, LNCS-Series Vol. 588:579–587, 1992.

[11] R.I.Hartley. In defence of 8-point algorithm. Proc. of 5th ICCV, pages 1064–1070, 1995.

[12] R.Y.Tsai and etc. Uniqueness and estimation of 3d motion parameters of rigid bodies with curved surfaces. IEEE Trans. PAMI, 6:13–27, 1984.

[13] David G Lowe. Distinctive Image Features from Scale - Invariant Interest Points[J]. International Journal of Computer Vision, 2004, 60 (2) : 91 - 110.

[14] Babaud J , Witkin A P, BaudinMetal. Uniqueness of the Gaussian kernel for scale - space filtering[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1996, 8 (1) : 26 - 33.

[15] Wei Wang, Hung Tat Tsui. A SVD Decomposition of Essential Matrix with Eight Solutions for the Relative Position of Two Perspective Cameras [J]. Proceeding of the International Conference on Pattern Recognition., Vol 1: 362 – 365, 2000.