Camera Self-Calibration Based on Multiple View Images

Zhiyi Zhang*
College of Information Engineering
Northwest A&F University
Yangling 712100, Shaanxi, China
Email: 815802490@qq.com

Qiuhu Tang
College of Information Engineering
Northwest A&F University
Yangling 712100, Shaanxi, China
Email: tangqiuhu@qq.com

***Abstract*—Camera calibration is an essential part of 3D reconstruction. Conventional calibration methods require high precision equipment and sophisticated operations. Compared with that, Camera self-calibration is quite simple but with low precision, which leads to significant performance degradation of 3D reconstruction. Therefore, a high precision self-calibration method with simple operations is necessary. In this paper, by using the bundle adjustment algorithm and SIFT points matching relationship, a local-global hybrid iterative optimization method is proposed. Considering large number of matching features, a neighborhood image matching method is proposed that can significantly reduce the matching time under the premise of maintaining accuracy. Experimental results show that the proposed method is an effective method for high accuracy, and it can reduce the time consumption of image matching. Based on the relationship between corresponded matching points in multiply view images, our method makes full use of the local-global hybrid idea to compute the parameters of camera. Compared with existed other methods, it is more robust with higher precision.**

***Keywords*-camera self-calibration; multiple view images; image matching; bundle adjustment algorithm**

I. Introduction

Camera calibration is an important technology in the field of computer vision and considered as an essential step to extract the three-dimensional spatial information from two-dimensional images, which has been widely used in the field of 3D structural reconstruction, navigation, visual surveillance [1]. Conventional calibration methods require high precision equipment and sophisticated operations. The plane calibration method proposed by Zhang [2] requires a checkerboard as auxiliary equipment. As a result, this calibration method cannot achieve our aim of automation and simplification. It has become one of the important components in research to take 3D reconstruction by using uncalibrated image sequences [3]. Whereas, for multiple images, a common approach used to estimate the projection matrices of each two images and then the obtained projection matrices are optimized respectively. Finally, the matrices are processed with a global optimization to get camera parameters. The most used algorithm is LM (Levenberg-Marquardt Algorithm) proposed by Moré [4], which uses non-linear least square algorithm retaining the advantages of gradient method and Newton method. Hu [5] etc. and Nguyen [6] etc. have recommended their algorithms to realize 3D reconstruction by using unconstrained and uncalibrated images took from a handheld camera. In both studies they used SIFT algorithm proposed by Lowe [7] for feature point detection and matching, then estimated the projection matrix to realize camera calibration. Accordingly, the space 3D points, corresponding to the detected feature points, could be used to describe the geometric positions of the detected feature points. Considering each two images, this type of calibration algorithms implement the local and global optimization in turn. The camera parameters which obtained with this type of algorithms are not highly accurate enough and the 3D reconstruction results are just satisfactory. Moreover, to match each of two images is very time-consuming for large number of matching features.

Considering the defects of conventional camera calibration and self-calibration methods, this paper highlights a camera self-calibration method based on multiple view images. It adopts the local-global hybrid optimization algorithm first to realize self-calibration of camera positions and orientations that is we can compute the intrinsic and extrinsic camera parameters by image sequences and then implement global optimization by using bundle adjustment. Moreover, a neighborhood image matching method is proposed in this paper to reduce the time consumption of image matching.

II. Self-calibration algorithm

A. Epipolar geometry

Essentially, the epipolar geometry between two images is the intersection geometry of plane with axis of baseline. The relationship between fundamental matrix and essential matrix $\boldsymbol{E} = \boldsymbol{K^T F K}$, where $\boldsymbol{K}$ is an internal parameter matrix. It is convenient to obtain the essential matrix $\boldsymbol{E}$. In addition, the fundamental matrix $\boldsymbol{F}$ can be used to remove incorrect matches during SIFT feature points matching. Hartley [8] proposed that the camera rotation and translation parameters corresponding to the second image could be obtained with SVD singular value decomposition for the essential matrix $\boldsymbol{E}$. We modified RANSAC (Random Sample Consensus) algorithm optimizing the following 8 point method to obtain the fundamental matrix $\boldsymbol{F}$. The steps of our algorithm are given as **Algorithm 1**.

978-1-5090-2305-9/16 $31.00 © 2016 IEEE
DOI 10.1109/NicoInt.2016.16

Algorithm 1 Estimate the fundamental matrix

Input: n pairs of SIFT feature matching points

1. Set sampling stop condition according to confidence probability. In our study, confidence probability = 0.95.
2. Randomly select 8 pairs of matching points from SIFT matches of the two images to constitute a sampling.
3. On the basis of the random sampling in step 2, compute the fundamental matrix $\boldsymbol{F_{tmp}}$ by using linear least squares method for the 8 points.
4. Test each $\boldsymbol{F_{tmp}}$ with all the matching pairs of the two images, and then obtain the quantity of inliers corresponding to each $\boldsymbol{F_{tmp}}$.
5. Repeat the steps 2 to 4 until it meets the confidence probability.
6. Select the corresponding fundamental matrix $\boldsymbol{F}$ according to the quantity of inliers.
7. Find out all the inliers corresponding to optimal $\boldsymbol{F}$, and eventually obtaining the final fundamental matrix $\boldsymbol{F}$ by using nonlinear least squares method for the inliers.

Output: Fundamental matrix $\boldsymbol{F}$

B. Image matching

For the image matching we use a bidirectional SIFT feature matching algorithm proposed by An [9] to match feature points among images. For all the images, if we match each two images, the computational complexity will increase rapidly with the increase of image amount. When we take multiple images revolving around an object, for one of them, the camera rotation and translation of its adjacent images should be closed to the camera rotation and translation of itself. Besides, those images contain much more information, higher similarity and more feature matching point pairs. However, the images, found away from the main image, have low similarity. As a result, it can save time that we only match for its adjacent images to a certain image.

C. Algorithm Description

A point in space can be captured by multiple cameras in different positions. Bundle adjustment is a process in which we can extract 3D coordinates and relative positions of each camera from multi-view information. The initial values are the parameters of camera projection matrices and 3D points.

EXIF is the abbreviation of Exchangeable Image File, which contains the metadata specially designed for digital camera photos, taking parameters and thumbnails recording digital photos and other attributes. We can obtain initial focal length f_0 of cameras corresponding to each image through extracting EXIF information of digital camera photos. Our algorithm are given as **Algorithm 2**.

Algorithm 2 Estimate the camera parameters

Input: n organized images token by a handle camera

1. Extract SIFT features from all input images.
2. For each image:
   A. Select K candidate matching images as its K nearest neighbors.
   B. Solve for the fundamental matrix between pairs of images using algorithm 1.
   C. Implement bidirectional SIFT feature matching to the image with its K nearest neighbors.
3. Track all the matching feature points. So, the matching feature points will generate a trajectory.
4. Select the initial image pair. The standards of selection:
   A. The two images are K nearest neighbors and have maximum number of matching feature points.
   B. The baseline is required to be wide enough.
5. Compute the camera parameters of the initial pair and their 3D points. Then optimizing the parameters and the 3D point sets by using Bundle Adjustment.
6. Cameras adding:
   A. Every 3D point corresponds to a trajectory. Find out an image as the next calibration camera that has the maximum number of trajectory which correspond to the 3D point sets.
   B. Compute the parameters of this camera using the least square method and then a nonlinear least square method is implemented to the parameters.
   C. Compute new 3D points correspond to the newly joined camera and add them to the 3D point sets.
7. For all the calibrated cameras, implement a global optimization by using Bundle Adjustment.
8. Repeat the steps 6 to 7 until all cameras have been calibrated.

Output: The internal and external parameters of the camera

III. Experiment and Result

The method proposed in published articles [10], [11] is a classical self-calibration algorithm with high precision and robustness and results found were quite well. Therefore, we followed the same image sequences as studies of [10]–[14]. We further download the image sequences with resolution ratio of 512×768 from http://www.robots.ox.ac.uk/~vgg/data/data-mview.html as test images, and implement calibration with a certain camera. Fig.1a shows the example of the Valbonne church image sequence. Comparing our calibration result with the results presented in studies of [10]–[14], we conclude the estimation results shown in Table I, where (u_0, v_0) is the coordinate of principal point, indicating the coordinate of center point of the image. f_x and f_y are the focus values in the direction of x and y respectively.

As proposed in study of [12], we also take the results in studies of [10], [11] as the reference standards. As shown in Table I, our result is closer to the standards than that in literature [12]–[14]. Specifically, we compute the difference

(a) Valbonne church (b) Toy penguin

Figure 1. One image of Valbonne church and toy penguin sequence.

Table I
ESTIMATION RESULTS OF THE PROPOSED METHOD AND OTHER METHODS

	f_x	f_y	u_0	v_0
literature [12]	670.457	679.232	248.648	392.561
literature [10]	682.840	682.843	255.999	383.999
literature [11]	679.285	681.345	258.802	383.188
literature [13]	667	693	241	398
literature [14]	619	699	234	372
Our method	673.285	673.285	256.787	392.182
relative error(1)	1.399	1.399	-0.308	-2.131
relative error(2)	0.883	1.183	0.779	-2.347

between present results and that in literature [10], [11], and then divide it by the results in literature [10], [11]. It is clear that the relative error of the focal length f is about 1%, the relative error of u_0 is about $\pm0.5\%$, and the relative error of v_0 is about $\pm2.2\%$. Therefore, the calibration results are quite satisfactory, which proves the precision and validity of the proposed algorithm. We also use the patch-based multi-view stereo algorithm proposed by Furukawa [15] to 3D visualize the Valbonne church image sequences with the camera parameters given in the literature [10]–[14]. The comparing results are shown as Fig.2.

In order to reduce the matching time as described in section II-B and considering different number of multiple view images would influence on the velocity and precision of calibration, we take 70 pictures surrounding a toy penguin taken with a SONY DSC-W310 digital camera and a sample shown in Fig.1b. The image resolution is 2592×1944.

The experiment group the image sequence into three categories, corresponding to 30 images, 50 images, and 70 images respectively, we realize camera calibration and then test the calibration results with 3D point cloud visualization. For each category, there are three kinds of K nearest-neighbors image matching, corresponding to 4 nearest-neighbors, 8 nearest-neighbors and full matching. Table II shows the estimation results by using different number of images with different neighbors. Fig.3 shows the 3D point cloud visualization. It is clear that the matching images of one image are almost in its neighborhood.

Table II
ESTIMATION RESULTS OF THE PROPOSED METHOD WITH DIFFERENT NEIGHBORS

		f	u_0	v_0	matches(pairs)	time(s)
30	4	2044.460	1306.900	973.983	4224	4.859
	8	2044.640	1300.080	969.168	4311	9.914
	all	2043.260	1304.450	972.866	4340	31.184
50	4	2045.490	1271.010	975.916	15235	8.863
	8	2043.010	1301.710	969.753	17024	17.894
	all	2045.670	1301.050	970.449	17244	103.785
70	4	2042.000	1310.230	970.209	29890	12.959
	8	2043.630	1292.640	970.792	36310	25.527
	all	2043.480	1301.320	970.666	37409	205.830

IV. CONCLUSION

We have proposed a camera self-calibration method based on multiple view image sequence which is unconstrained and uncalibrated. We use bundle adjustment algorithm to estimate the internal and external parameters of the camera. During the image adding step, for the newly joined camera, we calculate the camera parameters by using the least square method optimized by RANSAC algorithm and it can be regarded as a local optimization. Then, a global optimization is done to all the calibrated cameras by bundle adjustment after the 3D space points corresponding to the newly joined camera added to the 3D point sets. It is a local-global hybrid iterative optimization process. Compared with the existed other methods, it is more robust with higher precision.

Furthermore, for large number of matching features, we use a neighborhood image matching method to reduce the matching time under the premise of maintaining accuracy. For the image sequence is organized, to one of the images, i.e. camera rotation and translation of its adjacent images is closed to the camera rotation and translation of itself and it is confirmed by the experiment we conducted.

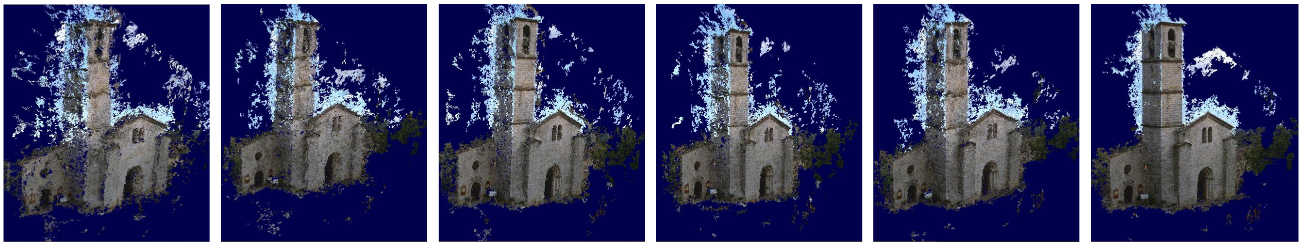
ACKNOWLEDGMENT

This work was supported by the National High Technology Research and Development (863) Program of China (No.2013AA10230402), the National Nature Science Foundation of China (No.61303124), Chinese Universities Fundamental Scientific Research Fund from Northwest A&F University (No.QN2013054 and No.QN2013056) and China Scholarship Council (No.201306305043).

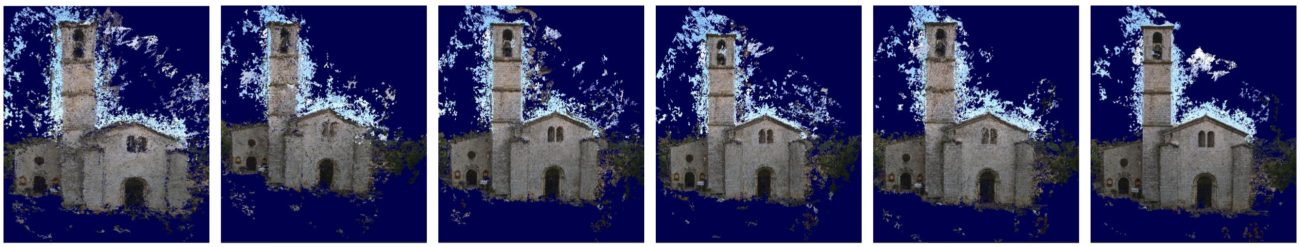
REFERENCES

[1] T. Mi, P. An, S. Liu, and Z. Zhang, “Multi camera calibration based on multi view registration algorithm,” *Journal of Image and Graphics*, vol. 13, no. 10, pp. 1922–1925, 2008.

[2] Z. Zhang, “A flexible new technique for camera calibration,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.



(a) literature [14](Flank) (b) literature [13](Flank) (c) literature [11](Flank) (d) literature [10](Flank) (e) literature [12](Flank) (f) our method(Flank)



(g) literature [14](Front) (h) literature [13](Front) (i) literature [11](Front) (j) literature [10](Front) (k) literature [12](Front) (l) our method(Front)

Figure 2. 3D visualization of Valbonne church with the proposed method and other methods.

(a) (30, 4) (b) (30, 8) (c) (30, all) (d) (50, 4) (e) (50, 8) (f) (50, all) (g) (70, 4) (h) (70 ,8) (i) (70, all)

Figure 3. 3D visualization of penguin sequence with different neighbors.

[3] Q. Wu, Z. Fu, and Q. Meng, "New camera self-calibration method based on image sequences," *Journal of Computer Applications*, vol. 11, p. 037, 2011.

[4] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[5] S. Hu, J. Qiao, A. Zhang, and Q. Huang, "3d reconstruction from image sequence taken with a handheld camera," in *International Society for Photogrammetry and Remote Sensing, Congress Beijing*. Citeseer, 2008, pp. 559–562.

[6] M. H. Nguyen, B. Wünsche, P. Delmas, and C. Lutteroth, "Modelling of 3d objects using unconstrained and uncalibrated images taken with a handheld camera," in *Computer Vision, Imaging and Computer Graphics. Theory and Applications*. Springer, 2013, pp. 86–101.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[8] R. I. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Computer VisionłECCV'92*. Springer, 1992, pp. 579–587.

[9] T. An, "An improved bidirectional sift feature matching algorithm," *Computer Engineering & Science*, vol. 38, no. 1, pp. 138–143, 2016.

[10] L. Lourakis and R. Deriehe, "Camera self-calibration using the svd of the fundamental matrix: From point correspondences to 3d measurements," Tech. Rep. 3748. INRIA, Tech. Rep., 1999.

[11] C. Zeller and O. Faugeras, "Camera self-calibration from video sequences: the kruppa equations revisited," 1996.

[12] Z. Jiang and W. Wu, "An essential matrix-based camera self-calibration method [j]," *Journal of Image and Graphics*, vol. 15, no. 4, pp. 565–569, 2010.

[13] A. Habed and B. Boufama, "Camera self-calibration from bivariate polynomials derived from kruppa's equations," *Pattern Recognition*, vol. 41, no. 8, pp. 2484–2492, 2008.

[14] A. Fusiello, A. Benedetti, M. Farenzena, and A. Busti, "Globally convergent autocalibration using interval analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 12, pp. 1633–1638, 2004.

[15] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 8, pp. 1362–1376, 2010.