Optik 123 (2012) 1923–1927

Contents lists available at SciVerse ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A new approach for the bundle adjustment problem with fixed constraints in stereo vision

Junpeng Xue, Xianyu Su*

Department of Opto-Electronics, Sichuan University, Chengdu, Sichuan 610064, China

ARTICLE INFO

Article history:
Received 4 May 2011
Accepted 2 September 2011

Keywords:
Stereo vision
3D reconstruction
Constrained bundle adjustment
Traditional bundle adjustment
SIFT

ABSTRACT

A new approach for the bundle adjustment problem with fixed constraints in stereo vision is described in this paper. Since the direct application of traditional bundle adjustment fails to use the inner constraints completely which are maintained by fixating the orientation and the baseline between the left and right cameras. However, if the fixed constraints are applied to the traditional bundle adjustment, we refine only the left camera extrinsic parameters and 3D points for simplification in stereo pairs. The new method using the fixed constraints has superior theoretical 3D accuracy, and it can reduce the matrix dimension of the covariance matrix so that the total computation time is decreased. Experiments results using synthetic and real data have shown that our method is better than the traditional bundle adjustment algorithm in the 3D accuracy and the convergence rate.

1. Introduction

3D measuring technique has many practical applications in reverse engineering, online monitoring, machinery processing, and quality testing [1–4]. Stereo vision is a typical system in 3D measurement [5,6]. In this method, two cameras, viewing from different angles, capture the images. Corner feature points are detected in these images of each stereo pairs. These feature points are then triangulated at each frame based on stereo correspondences [7,8]. Using stereo vision technique, 3D scanners that can capture the three-dimensional shape of objects accurately [7] are widely used in reverse engineering. However, it remains challenging for measuring large objects. When comes to large objects, it needs to reconstruct the framework of tie points using bundle adjustment method, and then match every sub-shape to the framework by 3D data registration method [9]. But, if the framework of the reconstruction accuracy is not high, it may fail or do not accurately match. Now, we study how to improve the reconstruction precision of the framework using bundle adjustment in stereo vision.

Bundle adjustment (BA) is the method of choice for many photogrammetry applications [10–12]. It has also come to take a prominent role in computer vision applications as the last step of many feature-based 3D reconstruction algorithms, for example [13–15], for a few representative approaches. BA refines a visual reconstruction to produce jointly optimal 3D structure and viewing parameters (camera pose and possibly intrinsic calibration) estimates. BA boils down to minimizing the re-projection error between the observed and predicted image points, which is expressed as the sum of squares of a large number of non-linear, real-valued functions [10]. Thus, the minimization is achieved using non-linear least squares algorithms, of which most of these articles call on the Levenberg–Marquardt's (LM) algorithm [16].

In this paper we present the discussion for studying how reconstruction accuracy could be improved by employing fixed constraints inherent in the two cameras in stereo vision. Since the fixed constraints are not applied into covariance matrix in traditional BA, more optimized parameters are needed and a long calculated time is taken. By incorporating these constraints into the traditional BA, we obtain the superior 3D accuracy and the convergence rate. Organization of the paper is as follows. Section 2 describes the theory and algorithm used in our approach. In Section 3, we provide an experimental comparison between BA and "constrained" BA, which clearly demonstrates the superiority of the latter in terms of 3D accuracy and convergence rate. Finally, in Section 4, we conclude the work and with a brief discussion.

2. Theoretical background and algorithm

2.1. Projection model

A camera model is a mathematical formation which approximates the behavior of any physical device by using a set of mathematical equations [17]. A 2D point is denoted by $m = [u, v]^T$. A 3D point is denoted by $M = [X, Y, Z]^T$. We use homogeneous coordinates to denote the augmented vector by adding 1 as the last

* Corresponding author. Tel.: +86 28 85466722; fax: +86 28 85464568.
E-mail addresses: junpengxue@163.com (J. Xue), xianyusu@mail.sc.cninfo.net (X. Su).

0030-4026/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ijleo.2011.09.025

element: $\tilde{m} = [u, v, 1]^T$ and $\tilde{M} = [X, Y, Z, 1]^T$. A camera is modeled by the usual pinhole: the relationship between a 3D point M and its image projection m is approximated by means of a transformation matrix, as shown in the equation.

$$s\tilde{m} = KP \times \tilde{M} \tag{1}$$

where s is a scale factor. K, called the camera intrinsic matrix, is given by

$$K = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where f_u and f_v represents the focal length in image u and v axes, (u_0, v_0) is the coordinates of the principal point, and γ is the aspect ratio, which is often ignored and set to zero. The parameters s, f_u, f_v, u_0 and v_0 are the intrinsic parameters of the pinhole camera model. P is a 4 by 4 matrix describing the mapping from world coordinate to camera coordinate. It is decomposed as follows:

$$P = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \tag{3}$$

where $T = [T_x, T_y, T_z]^T$ describes the translation between the two coordinates, and R is a 3 by 3 orthonormal rotation matrix which can be defined by the three Euler angles ω, φ and k, we have

$$R = \begin{bmatrix} \cos\varphi\cos k & \sin\omega\sin\varphi\cos k - \cos\omega\sin k & \cos\omega\sin\varphi\cos k + \sin\omega\sin k \\ \cos\varphi\sin k & \sin\omega\sin\varphi\sin k + \cos\omega\cos k & \cos\omega\sin\varphi\sin k - \sin\omega\cos k \\ -\sin\varphi & \sin\omega\cos\varphi & \cos\omega\cos\varphi \end{bmatrix} \tag{4}$$

The parameters $T_x, T_y, T_z, \omega, \varphi$ and k are called extrinsic parameters.

From Eq. (1), we have

$$\begin{cases} u = f_u \dfrac{r_{11}X + r_{12}Y + r_{13}Z + T_x}{r_{31}X + r_{32}Y + r_{33}Z + T_z} + u_0 \\ v = f_v \dfrac{r_{21}X + r_{22}Y + r_{23}Z + T_y}{r_{31}X + r_{32}Y + r_{33}Z + T_z} + v_0 \end{cases} \tag{5}$$

where r_{ij} is the ith row and the jth column of the rotation matrix R. Eq. (5) is called collinear equation, which are based on "R–T" model, in photogrammetry.

2.2. The fixed constraints in stereo vision

As shown in Fig. 1, inner constraints are maintained by fixating the orientation and the baseline between the left and right cameras. We denote left and right camera by subscripts l and r. We have

$$\begin{cases} R_r = R_c R_l \\ T_r = T_c + R_c T_l \end{cases} \tag{6}$$

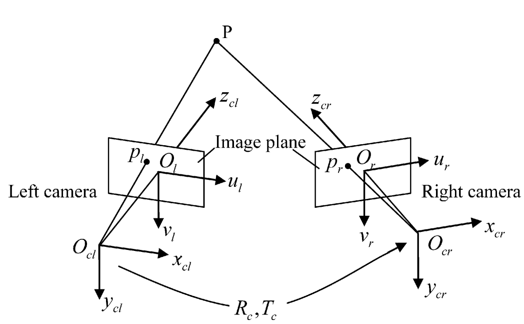

Fig. 1. Stereo vision position.

where R_c and T_c are the relationship between the left and right cameras respectively represent the rotation and translation, this is obtained from camera calibration [17].

Inserting Eq. (6) into Eq. (1), we get right camera transformation matrix represented by the left camera extrinsic parameters, we have

$$s\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = K_r \cdot \begin{bmatrix} R_c R_l & T_c + R_c T \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{7}$$

2.3. The constrained bundle adjustment method

The direct application of traditional bundle adjustment does not make full use of the constraints available. The optimized 3D points and extrinsic parameters of the left images might not be well re-projected on the right images in stereo vision. If the relationship of the two cameras in stereo vision is fixed in Eq. (6), two images, taken from the same point in space, can be gotten, then we have covariance matrix:

$$\begin{bmatrix} B^T B & B^T C \\ C^T B & C^T C \end{bmatrix} \cdot \begin{bmatrix} \delta x \\ \delta t \end{bmatrix} = \begin{bmatrix} B^T L \\ C^T L \end{bmatrix} \tag{8}$$

where $B = \begin{bmatrix} B_l & B_r \end{bmatrix}^T$, $C = \begin{bmatrix} C_l & C_r \end{bmatrix}^T$, $L = \begin{bmatrix} L_l & L_r \end{bmatrix}^T$, $\delta x = \delta x_l$, $\delta t = \delta t_l$.

L_l and L_r are the observed error in two images respectively.

$$L_l = \begin{bmatrix} u_l - \tilde{u}_l \\ v_l - \tilde{v}_l \end{bmatrix}, \quad L_r = \begin{bmatrix} u_r - \tilde{u}_r \\ v_r - \tilde{v}_r \end{bmatrix}$$

where (u, v) are the measured data, $(\tilde{u}, \tilde{v})$ are the back projected coordinates using collinear equation in Eq. (5).

(B_l, C_l) and (B_r, C_r) respectively represent the partial derivative of (u, v) in Eqs. (5) and (7) which are just for the extrinsic parameters of the left cameras $(\omega, \varphi, k, T_x, T_y, T_z)$ and 3D points. We have

$$B_l = \begin{bmatrix} \dfrac{\partial u_l}{\partial \omega_l} & \dfrac{\partial u_l}{\partial \varphi_l} & \dfrac{\partial u_l}{\partial k_l} & \dfrac{\partial u_l}{\partial T_{lx}} & \dfrac{\partial u_l}{\partial T_{ly}} & \dfrac{\partial u_l}{\partial T_{lz}} \\ \dfrac{\partial v_l}{\partial \omega_l} & \dfrac{\partial v_l}{\partial \varphi_l} & \dfrac{\partial v_l}{\partial k_l} & \dfrac{\partial v_l}{\partial T_{lx}} & \dfrac{\partial v_l}{\partial T_{ly}} & \dfrac{\partial v_l}{\partial T_{lz}} \end{bmatrix}$$

$$B_r = \begin{bmatrix} \dfrac{\partial u_r}{\partial \omega_l} & \dfrac{\partial u_r}{\partial \varphi_l} & \dfrac{\partial u_r}{\partial k_l} & \dfrac{\partial u_r}{\partial T_{lx}} & \dfrac{\partial u_r}{\partial T_{ly}} & \dfrac{\partial u_r}{\partial T_{lz}} \\ \dfrac{\partial v_r}{\partial \omega_l} & \dfrac{\partial v_r}{\partial \varphi_l} & \dfrac{\partial v_r}{\partial k_l} & \dfrac{\partial v_r}{\partial T_{lx}} & \dfrac{\partial v_r}{\partial T_{ly}} & \dfrac{\partial v_r}{\partial T_{lz}} \end{bmatrix}$$

$$C_l = \begin{bmatrix} \dfrac{\partial u_l}{\partial X} & \dfrac{\partial u_l}{\partial Y} & \dfrac{\partial u_l}{\partial Z} \\ \dfrac{\partial v_l}{\partial X} & \dfrac{\partial v_l}{\partial Y} & \dfrac{\partial v_l}{\partial Z} \end{bmatrix}, \quad C_r = \begin{bmatrix} \dfrac{\partial u_r}{\partial X} & \dfrac{\partial u_r}{\partial Y} & \dfrac{\partial u_r}{\partial Z} \\ \dfrac{\partial v_r}{\partial X} & \dfrac{\partial v_r}{\partial Y} & \dfrac{\partial v_r}{\partial Z} \end{bmatrix}$$

When all of the 3D points are projected to the sequence pairs, we have the covariance matrix

Table 1
The intrinsic parameters of simulated camera.

f_u (pixels)	f_v (pixels)	u_0 (pixels)	v_0 (pixels)	s (radian)
2500	2500	640	512	0

$$\begin{bmatrix} U_1 & & & & W_{11} & W_{21} & \cdots & W_{n1} \\ & U_2 & & & W_{12} & W_{22} & \cdots & W_{n2} \\ & & \ddots & & \vdots & \vdots & \ddots & \vdots \\ & & & U_m & W_{1m} & W_{2m} & \cdots & W_{nm} \\ W_{11}^T & W_{12}^T & \cdots & W_{1m}^T & V_1 & & & \\ W_{21}^T & W_{22}^T & \cdots & W_{2m}^T & & V_2 & & \\ \vdots & \vdots & \ddots & \vdots & & & \ddots & \\ W_{n1}^T & W_{n2}^T & \cdots & W_{nm}^T & & & & V_n \end{bmatrix} \cdot \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_m \\ \delta t_1 \\ \delta t_2 \\ \vdots \\ \delta t_n \end{bmatrix} = \begin{bmatrix} \varepsilon x_1 \\ \varepsilon x_2 \\ \vdots \\ \varepsilon x_m \\ \varepsilon t_1 \\ \varepsilon t_2 \\ \vdots \\ \varepsilon t_n \end{bmatrix} \quad (9)$$

where n is the number of 3D points, m is the number of stereo pairs.

$$U_j = \sum_{i=1}^{n} B_{ij}^T B_{ij}, \quad j = 1, 2, \ldots, m; \quad V_i = \sum_{j=1}^{m} C_{ij}^T C_{ij}, \quad i = 1, 2, \ldots, n;$$

$$W_{ij} = B_{ij}^T C_{ij}, \quad \varepsilon x_j = \sum_{i=1}^{n} B_{ij}^T L_{ij}, \quad \varepsilon t_j = \sum_{j=1}^{m} C_{ij}^T L_{ij}$$

The update unknown variables $\begin{bmatrix} \delta x_1 & \delta x_2 & \cdots & \delta x_m & \delta t_1 \\ & \delta t_2 & \cdots & \delta t_n & \end{bmatrix}^T$ can be solved using the least squares method from Eq. (9). Then, update the extrinsic parameters of left camera and 3D points. The procedure is repeated until the residual error cannot be reduced significantly.

It is clearly that the matrix in the left hand side of Eq. (9) is a $(6m+3n)$ order symmetric matrix. If we use traditional bundle adjustment method, the order of symmetric matrix is $(6 \times 2m+3n)$. Obviously, the constrained bundle adjustment method requires less computation time to solve Eq. (9) in each of iteration compared to the traditional bundle adjustment algorithm.

3. Experimental results

In this section, we compare the performance of the constrained and unconstrained traditional bundle adjustments by synthetic and real image sequences. Both methods are implemented in C, using LAPACK for linear algebra numerical operations. All experiments are conducted on a Celeron(R) CPU 3.06 GHz running Windows XP Professional and un-optimized BLAS.

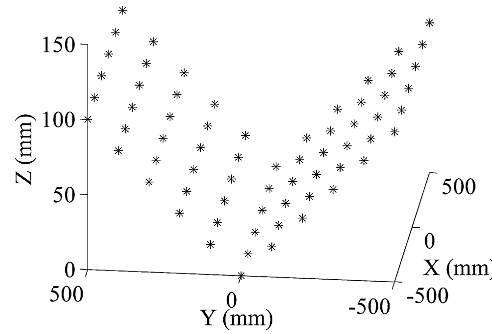

Fig. 2. Scene geometry illustration, the true 3D points can be reconstructed.

3.1. Experiments on synthetic data

We use Matlab to generate a synthetic scene consisting of 66 3D points. As shown in Fig. 2, these points are generated in $1000 \times 1000 \times 100$ "V" shape. The 3D points are viewed by stereo cameras with identical perspective model. The vision system is placed uniformly along a 180°, direct towards the scene, at 10 stations bundle network [18]. The vertical distance for the stereo systems to the XY plane is 2200 mm. The intrinsic parameters of the simulated camera are shown in Table 1. The baseline between the left and right cameras is 80 cm, so $T_c = [80, 0, 0]^T$. The two cameras optical axis are parallel, so the relative orientation angles ω, φ and k are 0 respectively. A sequence of 20 images has been generated from the model under perspective projection.

The constrained bundle adjustment method and the traditional bundle adjustment algorithm were given the same measurements and starting approximations. In each simulation, Gaussian noise was added to the projected points, with a noise level was varied from 0 to 2 pixels, simulating measurement errors. As it is evident from the final squared re-projection error, both approaches converge to almost identical solutions. In each noise level, 100 independent runs were performed, and the average results of the re-projection error and the computation time are reported here.

Table 2 shows the objects real values and the values computed respectively from "constrained" bundle adjustment method and the traditional bundle adjustment method at a noise level of 1 pixel added to the image points. The experimental results show that both methods converge reliably. The value with "constrained" BA method is more accurate than the value with traditional BA.

Fig. 3(a) compares the average 3D relative errors between the two methods. The results for two methods are different. The "constrained" BA method is clearly the better one. The constrained method can produce a smaller residual error than the unconstrained method at all noise level. The constraints help to model the object 3D points more realistically, thereby suppress over-fitting, and produce a more accurate model.

Table 2
The 3D points real values and the values computed respectively from constrained BA and traditional BA method at a noise level of 1 pixel, only show 11 points (unit: mm).

No.	Real value			Constrained BA				Traditional BA			
	x	y	z	x	y	z	Error	x	y	z	Error
1	−500	−500	100	−500.002	−500.074	99.771	0.240	−500.002	−500.162	99.483	0.541
2	−500	0	0	−499.987	−0.062	−0.357	0.362	−499.957	−0.139	−0.621	0.637
3	−500	500	100	−500.024	500.007	100.264	0.265	−499.857	500.010	100.628	0.644
4	100	−500	100	100.037	−499.916	100.035	0.098	100.670	−499.844	100.066	0.691
5	100	−200	40	100.082	−200.096	40.104	0.163	100.103	−200.172	40.198	0.282
6	100	0	0	100.072	0.041	0.240	0.253	100.068	0.530	0.573	0.783
7	100	200	40	99.906	200.095	40.094	0.163	100.166	200.187	40.669	0.714
8	100	500	100	100.027	500.061	99.927	0.099	100.236	500.247	99.874	0.364
9	500	−500	100	500.009	−500.082	99.814	0.203	500.209	−500.610	99.726	0.700
10	500	0	0	499.971	−0.053	−0.126	0.139	499.881	−0.085	−0.199	0.247
11	500	500	100	499.990	499.941	100.184	0.193	499.979	499.895	100.710	0.718

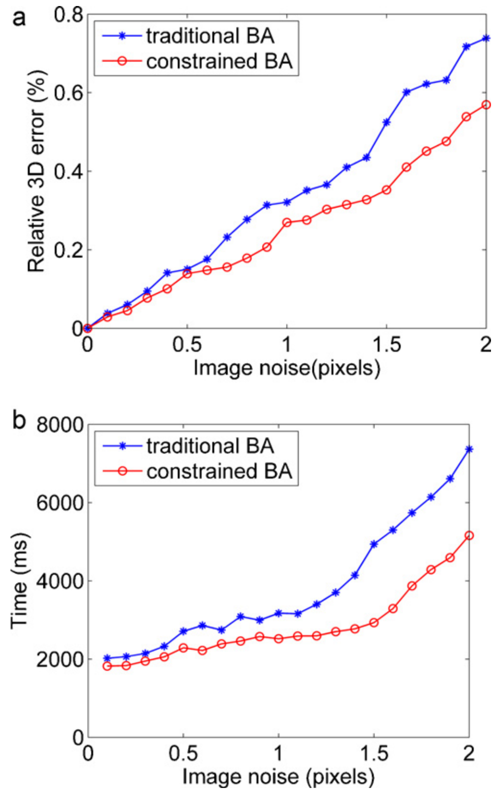

Fig. 3. Comparison of the accuracy and efficiency of the different methods. (a) The residual 3D errors; (b) the computation time.

The computation time for every noise level are shown in Fig. 3(b). It is clearly to see that the constrained method is faster than the unconstrained method at every noise level. As the covariance matrix order of the constrained method is smaller, the computation time are needed less in each iteration. The number of iterations will increase with the increasing of noise level, so the total computation time increases with the number of iterations.

3.2. Experiments on real condition

The constrained bundle adjustment method has also been tested on real images. We have chosen to work on the 312 images sequence "temple" of the online database [19]. In this database we know all the intrinsic and extrinsic parameters. All the intrinsic camera parameters are the same, and any two adjacent cameras relative rotation and translation are the same. Therefore, we select 14 adjacent image pairs as 14 stereo pairs of images. The feature points of interest are extracted using SIFT algorithm [20]. Over 4000 features are successfully matched with subpixel accuracy. The RANSAC algorithm [21], using homography as an underlying model, is applied to eliminate rare mismatches. Fig. 4 shows one of the image pairs of the temple with selected features indicated by small stars.

As mentioned before, the actual state values are not known to us, so it is not possible to display the errors in the reconstruction 3D. Fig. 5 shows the reconstructed view of the set of 4196 points from front views. Clearly visible seems to confirm that the constrained bundle adjustment method is doing a reasonably good job of the reconstructed framework.

Fig. 6 shows the actual processing time of the constrained bundle adjustment method compared with that of the traditional bundle adjustment method. The number of features as 228, 1374, 2675 and 4196 select from all 3D points, while the number of frame

Fig. 4. One of the image pairs of the temple. The superimposed stars indicate the selected features.

Fig. 5. Front view of reconstructed point set.

pairs is fixed at 14. The processing time for tracking features and reconstruction are not included. Clearly to see that the constrained bundle adjustment method has a real advantage of the faster computation.

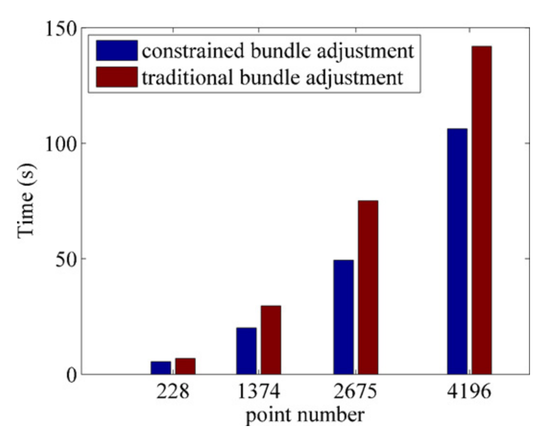

Fig. 6. The processing time of the constrained BA method compared with that of the traditional BA method.

4. Conclusions

In this paper, we have made a comparison for the traditional bundle adjustment and the constrained bundle adjustment in the 3D accuracy and computation time. The experiment results show that our method has a better 3D accuracy than the traditional bundle adjustment. Hence, in the cases when it is not possible to improve the data, e.g. due to restricted access to the object or poor spatial distribution of markers, the improvement in reliability introduced by the constrained bundle adjustment method will be useful. The constrained bundle adjustment method also can reduce the matrix dimension of the covariance matrix. So the total computation time should also been decreased. This increases the convergence rate and provides a very viable option even for real time applications.

According to the results of synthetic and real image sequences, we can affirm that our method is accurate and fast. Our method should be good for the proper implementation.

Acknowledgements

The authors are thankful to the anonymous reviewers for their useful suggestions and comments. This work is supported by the National Natural Science Foundation of China (Grant No. 60838002).

References

[1] C. Lee, Y. Lim, S. Kwon, J. Lee, Stereo vision-based vehicle detection using a road feature and disparity histogram, Opt. Eng. 50 (2) (2011) 027004–27023.
[2] X. Llad, A. Del Bue, A. Oliver, J. Salvi, L. Agapito, Reconstruction of non-rigid 3D shapes from stereo-motion, Pattern Recogn. Lett. 32 (7) (2011) 1020–1028.
[3] M. Agrawal, K. Konolige, Real-time localization in outdoor environments using stereo vision and inexpensive GPS, in: Proc. of the 18th IEEE International Conference on Pattern Recognition (ICPR'06), 2006, pp. 1063–1068.
[4] H. Hirschmüller, P.R. Innocent, J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, Int. J. Comput. Vision 47 (1) (2002) 229–246.
[5] X. Su, Q. Zhang, L. Xiang, Optical 3D shape measurement for dynamic process, Optoelectron. Lett. 4 (1) (2008) 55–58.
[6] X. Bian, X. Su, W. Chen, Analysis on 3D object measurement based on fringe projection, Optik 122 (6) (2011) 471–474.
[7] D.J. Kriegman, E. Triendl, T.O. Binford, Stereo vision and navigation in buildings for mobile robots, IEEE Trans. Robotics Automation 5 (6) (1989) 792–803.
[8] O. Faugeras, Stratification of three-dimensional vision: projective, affine, and metric representations, J. Opt. Soc. Am. A 12 (3) (1995) 465–484.
[9] J. Han, N. Lu, et al., 3D data registration method based on optical location tracking technology, Opt. Precision Eng. 17 (1) (2009) 45–51.
[10] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment – a modern synthesis, Vision Algorithms: Theory Practice 1883 (2000) 298–372.
[11] M.I.A. Lourakis, A.A. Argyros, SBA: a software package for generic sparse bundle adjustment, ACM Trans. Math. Softw. 36 (1) (2009) 1–30.
[12] T. Luhmann, Close range photogrammetry for industrial applications, ISPRS J. Photogramm. Rem. Sens. 65 (6) (2010) 558–569.
[13] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, P. Sayd, Generic and real-time structure from motion using local bundle adjustment, Image Vision Comput. 27 (8) (2009) 1178–1193.
[14] S. Agarwal, N. Snavely, S. Seitz, R. Szeliski, Bundle adjustment in the large, in: ECCV 2010, Part II, vol. 6312, 2010, pp. 29–42.
[15] Z. Ji, M. Boutin, D.G. Aliaga, Robust bundle adjustment for structure from motion, in: IEEE International Conference on Image Processing, 2006, pp. 2185–2188.
[16] D. Marquardt, An algorithm for the least-squares estimation of nonlinear parameters, SIAM J. Appl. Math. 11 (2) (1963) 431–441.
[17] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334.
[18] S. Mason, Expert system-based design of close-range photogrammetric networks, ISPRS J. Photogramm. Rem. Sens. 50 (5) (1995) 13–24.
[19] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, Multiview Stereo Valuation, http://vision.middlebury.edu/mview/data/.
[20] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[21] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.