A 3D Reconstruction System of Indoor Scenes with Rotating Platform

Feng Zhang, Limin Shi , Zhenhui Xu, Zhanyi Hu
Institute of Automation, Chinese Academy of Sciences
{fzhang, lmshi, zhxu, huzy}@nlpr.ia.ac.cnl

Abstract

3D reconstruction has a long history, few workable reconstruction systems are available although many magnificent theories and algorithms are reported in literature. In this paper, a 3D reconstruction system for indoor scenes with a rotating platform is reported, including its key components such as system calibration, feature matching and 3D reconstruction. Experimental results show that this system is not only convenient and flexible, but it can also obtain satisfactory reconstruction accuracy.

1. Introduction

Many methods exist for acquiring 3D information of indoor scenes, such as stereo vision [2], structure light [4][5], scanning laser and so on. No matter what method, as soon as the positions of sensor and object remain unchanged, we can only get the local reconstruction under a single view. For the reconstruction of the whole scene, the relative position between sensor and object must be changed. Hence, the fusion from different reconstructed patches is a prerequisite, but a difficult problem. To this end, we propose a new reconstruction system for indoor scenes with a rotating platform (Fig 1) and some of its key issues are investigated.

At present, there are many methods for camera calibration [6] [7]. In [6], a planar chessboard pattern is used to calibrate the pinhole camera model. Due to the flexibility and convenience, it is perhaps the most widely used method. In [7], the author proposes an accurate method for calibrating wide-angle or fish-eye lenses, using planar pattern. Besides, estimating the transformation between the rotating platform with a mechanical link and a sensor mounted on the link, is a special Hand-Eye calibration [8] [9] [10], which comes down to solving the equation like $R_A R = RR_B$. Because of the speciality of rotating platform, by properly choosing the world system, we can avoid solving the above equation and simplify the calibration.

It is well known that correspondence is the key art for image-based reconstruction. Existing local region descriptors such as SIFT [11] or GLOH [12] have been designed for robustness to perspective and lighting changes and have been proved successful for sparse feature matching. However, knowing some sparse points are not sufficient for scene reconstruction, a quasi-dense approach [13] [14] is at least needed for shape preserving.

The triangulation problem [1] [15] [16] [17] is important and is being studied hotly at present. In [1], a simple algebraic method is used for solving the problem in n-view case. In [15], a geometric method is adopted for the case of two views, involving the solution of a sixth-degree polynomial. And the use of L_∞ optimization is introduced in [16] [17] to ensure a global minimization. To refine the initial reconstruction from multi-view, the bundle adjustment [18] [19] is often necessary.

This paper is organized as follows. In the next section, a brief introduction to our system is proposed. Then, a method for system calibration (Section 3) and that for multi-view reconstruction (Section 4) are discussed. Finally, experimental results are given in Section 5 followed by a short conclusion.

2. A brief introduction to the system

As shown on Fig 1, the system is composed of a camera, a rotating platform and a control unit. The camera is mounted on the rotating platform with a free handle, which can adjust the view direction. The rotating platform causes the camera making pure planar motion, and the angle of rotation can be read from the control unit. For every rotation, the error is less than 0.02° and accumulated error does not exist.

Fig 1(a) System of rotating platform

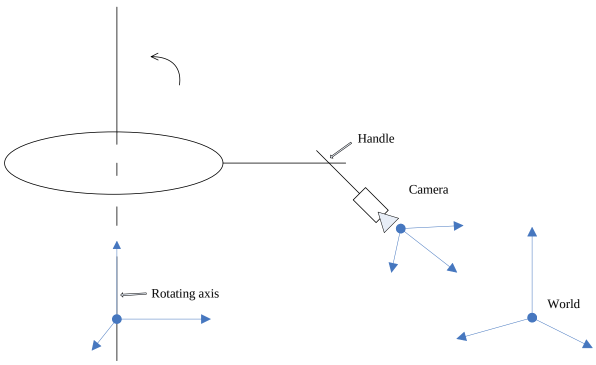


Fig 1(b)Configuration of system

3. System calibration

At first the intrinsic parameters of the camera is calibrated using the method mentioned in [6]. Aiming at reducing the coupling effect of the camera's intrinsic parameters to the extrinsic parameters, both radial distortion and tangent distortion are taken into account in our work.

A method based on transformation of coordinate system [3] is employed to calibrate the relative position between the camera and rotating platform. In the following derivations, the planar calibration pattern, which is placed on the floor, is chosen as the x-y plane of the world coordinate system, so the plane of the calibration pattern is $Z = 0$. By calculating the homography between the calibration pattern plane and the image plane, we have

$$H_1 \approx K_1 [r_1 \quad r_2 \quad t_1]$$

$$X_c^{\ 1} = R_1 X_w + t_1 \qquad (1)$$

where r_i is the ith column of R_1.

Assume that the angle of rotation is θ, and that the world coordinate system remains unchanged. The following transformation between the world coordinate system and the camera coordinate system after rotation is obtained in the same way.

$$X_c^{\ 2} = R_2 X_w + t_2 \qquad (2)$$

Since the direction of rotational axis is orthogonal to the floor, if both the x axis and the y axis of the coordinate system of the rotating platform are on the planar pattern, the following relations are obtained

$$X_p = X_w + t_p \qquad (3)$$

$$t_p = \begin{pmatrix} x_0 \\ y_0 \\ 0 \end{pmatrix}$$

where $(x_0 \quad y_0 \quad 0)^T$ is the origin of the rotating platform, represented in the world coordinate system. From equation (3) we obtain

$$X_w^{\ 1} = X_p^{\ 1} - t_p \qquad (4)$$

Because the world coordinate system remains unchanged, we can write

$$X_w^{\ 2} = X_w^{\ 1} = X_p^{\ 1} - t_p = R_\theta^{\ T} X_p^{\ 2} - t_p \quad (5)$$

where R_θ is the rotation matrix of the rotating platform. Using the above equations, we obtain

$$X_c^{\ 1} = R_1 \left(X_p^{\ 1} - t_p \right) + t_1$$

$$X_c^{\ 2} = R_2 \left(R_\theta^{\ T} X_p^{\ 2} - t_p \right) + t_2$$

Since transformation between the camera coordinate system and the coordinate system of rotating platform are unchanged, the following equations can be obtained

$$R = R_1 = R_2 R_\theta^{\ T} \qquad (6)$$

$$t = t_1 - R_1 t_p = t_2 - R_2 t_p \qquad (7)$$

where R, t denote the rotation transformation and translation transformation between the camera coordinate system and the coordinate system of rotating platform. From equations (6) (7) we obtain

$$R_1^{\ T} R_2 = R_\theta \qquad (8)$$

$$R_1^{\ T} (t_2 - t_1) = (R_\theta - I) t_p \qquad (9)$$

Notice that the rank of the matrix $I - R_\theta$ is 2, except when the R_θ is the identity matrix. Hence, there are two independent constraint for the vector t_p . From equation (9) we can get initial estimation of t_p .Then substituting it in equation (7), we can obtain rotation matrix R .

4. 3D Reconstruction

After completing system calibration, we set the world coordinate system coinciding with the coordinate system of the rotating platform. Thus, the camera projection matrix can be expressed as

$$P_0 = K\begin{bmatrix} R & t \end{bmatrix}$$

Let θ_i be the angle of the ith rotation, we have

$$P_i = K\begin{bmatrix} RR_{\theta i} & t \end{bmatrix}, i = 1, 2, \cdots, n$$

where $R_{\theta i}$ denotes the ith rotation matrix.

The methods in [1] [16] can be used to estimate the 3D structure. Due to the space limit, this part will be skipped over. As we said before, contrary to the system calibration where some sparse correspondences suffice, 3D reconstruction needs at least a quasi-dense correspondence to give a discernible scene shape. In our work, a quasi-dense matching is carried out as: We start from a set of sparse SIFT seed matches, then propagate the matching points to the neighboring pixels. In addition, to remove outliers, the depth knowledge, i.e., the information on possible valid depth range of an indoor scene, is used for outlier removal. That is, we assume:

$$Z_{\min} \le z_j \le Z_{\max}$$

where $Z_{\min}$ and $Z_{\max}$ depend on the specific scene. In addition, to improve the accuracy, sparse bundle adjustment is applied to the multi-view reconstruction.

5.Experiments

After calibrating the system, to assess the reconstruction accuracy, we first reconstruct the calibration grid as shown in Fig2, an object manufactured with high precision. The grid size is of 40mm*40mm, our reconstruction results are shown in Table 1. The used image resolution is of 1600×1200 and the rotational angle of rotating platform is 25° .

(a) (b)

Fig 2. Two indoor scene images , the rotational angle of the rotating platform is 25° .

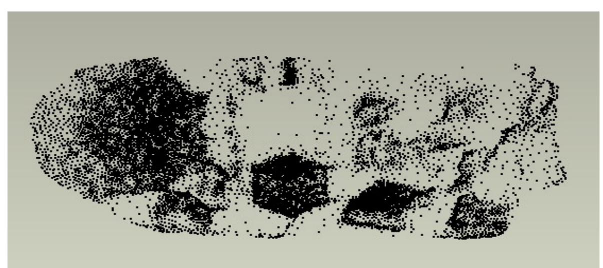
Unit : mm

Real size	Mean estimation	Mean error	Max error	RMS	#
40	39.7872	0.7589	1.7048	0.8841	60

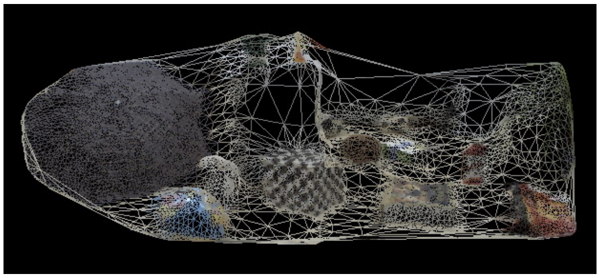
Table1 Comparison of the reconstructed grid size with the ground truth

We have carried out quite a number of indoor scene reconstructions, the results are rather satisfactory. The experiment is the reconstruction of a indoor scene from 9 images. The used image resolution is of 1600×1200 and the rotational angle is 5° per rotation. That is, the scene is covered by a total rotation of 40 degrees.

(a). Three of the 9 indoor scene images



(b) Clouds of reconstructed 3D points



(c) Meshes of 3D points

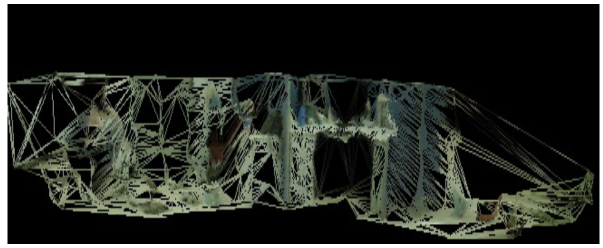
(d) 3D structure with texture

Fig 3

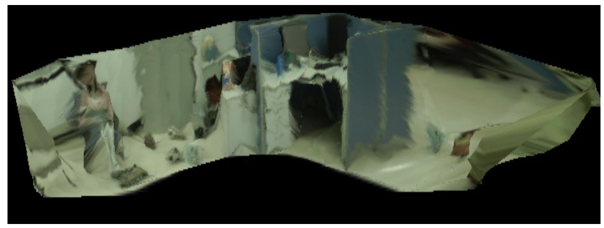
The following experiment is the reconstruction of another indoor scene from 26 images. The used image resolution is of 2048×1536 and the rotational angle is 5° per rotation. That is, the scene is covered by a total rotation of 150 degrees.

(a). Six of the 9 indoor scene images

(b) Clouds of reconstucted 3D points



(c) Meshes of 3D points



(d) 3D structure with texture

Fig 4

Before ending our report, we would list the following remarks:

- From the stand of reconstruction accuracy, our system is not very accurate compared with other reports. However, our system is fully automatic, no human intervention is involved during the reconstruction. For a fully automatic system for a relatively large scene, we thought the results are acceptable and competitive;
- Our goad is to build a fully automatic system for criminal scene reconstruction, here the ability of automatically recovering the scene is a must, but the reconstruction accuracy is not very exigent. That is why we are currently concentrated on automatic reconstruction. Worthy of mention is that automatically reconstructed a complex and large scene as that in Fig 3 is truly a difficult one.
- In this work, we do not report any comparative results. This is because to our knowledge, a comparable system for large and complex scene reconstruction is not available. Besides, for large and complex scene reconstruction, how to evaluate the system's performance is still an open question, still less the comparison among systems.

.

6. Conclusion and future work

We present a new indoor scene reconstruction system with a rotating platform, and report the reconstructed results. The system is fully automatic, flexible and practical. In the future, how to increase the reconstruction accuracy and how to handle textureless parts are two major issues to pursue. Of course, the robustness seems an eternal issue to concern.

7.Acknowledgement

This work was supported by the national 863 program(2007AA01Z341), the national science foundation of china(60673104) and the national science and technology development plan(2006BAK31B04).

8. References

[1] R.Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press,2000.

[2] Y.Jia，*Machine Vision*，Science Press，Beijing，2000

[3] H.Zhang，*Study On 3D Reconstruction From Multiple Views [Doctor Dissertation]*, Institute Of Automation, Chinese Academy Of Sciences，Beijing, 2004

[4] J.Batlle, E.Mouaddib and J.Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey", Pattern Recognition, 1998, pp.963–982.

[5] J.Salvi , J.Pagès and J.Batlle, "Pattern codification strategies in structured light systems", Pattern Recognition , April 2004, pp. 827-849

[6] Z.Zhang, "A Flexible New Technique for Cmera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, Nov.2000, pp.1330-1334,

[7] J.Kannala, S.Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, pp.1335-1340.

[8] Y.Shiu and S.Ahmad, "Calibration of Wrist-Mounted Robotic Sensors by Solbing Homogenous Transform Equations of the From AX = XB". IEEE Trans. Robotics and Automation, 1989, pp16-29.

[9]R.Horaud and F.Dornaika, "Hand-eye calibration. International Journal of Robotics Research",1995,pp.195-210.

[10]K.Daniilidis,"Hand-eye calibration using dual quaternions". International Journal of Robotics Research, 1999, pp.286-298.

[11]D.Lowe," Distinctive image features from scale-invariant keypoints". International Journal of Computer Vision, 2004, pp91-110.

[12]K.Mikolajczyk and C.Schmid, "A Performance Evaluation of Local Descriptors". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, pp.1615-1630.

[13]M.Lhuillier and L.Quan , "Match propagation for image-based modeling and rendering", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, pp.1140-1146.

[14]M.Lhuillier and L.Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, pp.418-433.

[15]R.Hartley and P.Sturm., "Triangulation", Computer Vision and Image Understanding, 1997, pp.146-157.

[16]F.Kahl, "Multiple view geometry and the L_∞ -norm". In Proc. International Conference On Computer Vision, 2005, pp.1002-1009.

[17]Q.Ke and T. Kanade, "Quasiconvex optimization for robust geometric reconstruction". In Proc. International Conference on Computer Vison, 2005, pp.986-963.

[18]B.Triggs." Bundle Adjustment – A Modern Synthesis". In Proc International WorkshopInternational Workshop on Vision Algorithms,1999.

[19]M.Lourakis and A.Argyros, "The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm", FORTH-ICS/TR340, 2004