3D Model Reconstruction by Constrained Bundle Adjustment

Kin Hong Wong[1], Michael Ming Yuen Chang[2]

[1]Department of Computer Science and Engineering,

[2]Department of Information Engineering,

The Chinese University of Hong Kong, Shatin, Hong Kong, SAR China

{khwong@cse.cuhk.edu.hk, mchang@ie.cuhk.edu.hk }

Abstract— **The accuracy of the reconstruction of 3D objects using vision-based techniques depends ultimately on an optimization technique known as bundle adjustment. However, direct application of bundle adjustment fails to make full use of the constraints inherent in the objects' motion. Most of the scanner systems available today use a turntable to rotate an object so as to capture images of the object from different view points. For these systems, the motion of the object is constrained to rotate around a fixed axis. By incorporating this constraint into the traditional bundle adjustment method, we show that a more accurate reconstructed model can be obtained. This "constrained" bundle adjustment method should be useful in the design of low-cost, yet accurate, 3D scanners for use in manufacturing and entertainment industry.**

Index Terms—**3D object reconstruction, bundle adjustment, 3D object scanner, structure from motion.**

I. INTRODUCTION

Scanners that can capture the 3D structure of objects accurately have many practical applications in manufacturing and entertainment industry. At present, most of the 3D scanners are based on laser range sensors [1]. These scanners are robust and accurate, but they are also costly, and have certain restrictions on the size and on the surface properties of the objects. They are also unable to capture the color information of the objects.

The advent of inexpensive digital cameras has made vision-based scanners an increasingly attractive proposition. A typical system requires only a turntable and a camera. By rotating the object on a turntable, multiple views of the object are captured by the camera, and the 3D structure of the object can be reconstructed using structure from motion (SfM) techniques. Tutorials on the reconstruction scheme are described in [2][3]. Apart from recovering the 3D object's structure, SfM can also recover the object's pose in the image sequence. The pose information is useful in reconstruction schemes that are based on silhouette [4][5], and space carving [6]. Some scanning systems require the turntable be controlled by a computer [7]. In our work, we take a minimalist approach so that the turntable is not required to be connected to a computer.

The main contribution in this paper is to study how reconstruction accuracy could be improved by employing constraints inherent in the object's motion. From the literature on SfM, the reconstructed model is frequently subjected to a non-linear optimization technique known as bundle adjustment – a term that is originally used in photogrammetry. Bundle adjustment is basically a steepest-descent algorithm that searches for an optimal model by minimizing the error between the observed 2D feature points and the re-projected feature points from the reconstructed model. We note, however, that the adjustment fails to capture constraints inherent in the object motion – namely that the object is constrained to rotate around a fixed axis. By incorporating this constraint into the optimization process, we expect the accuracy of the reconstruction can be improved substantially.

Organization of the paper is as follows: in section II, we describe the theory and algorithm used in our approach. In III, we describe the experimental results of this work, and finally in IV we conclude the work and discuss the results.

II. THEORY AND ALGORITHM

A. *Problem formulation*

The 3D reconstruction of an object is usually achieved in two stages. The first stage creates an approximate model. The second stage refines the model using an optimization technique known as bundle adjustment [8]. The accuracy of the reconstruction depends critically on the second stage. Here, we assume that an approximate model has already been obtained using techniques such as Extended Kalman Filtering [9]. Our main focus will be on the accuracy of bundle adjustment and how to improve on it.

Consider an object with $\boldsymbol{N}$ feature points. The object is rotated around a fixed axis $\boldsymbol{n}$ and its motion is captured by a stationary camera in a sequence of $\boldsymbol{T}$ images. Let $\boldsymbol{M}_i$ be the 3D coordinate of i^{th} feature point of the model, $\boldsymbol{R}_j$ be the rotational matrix that corresponds to the rotation of the object at the j^{th} frame, and $\boldsymbol{D}$ be the distance from the origin of the camera to the centre of the object, then the 3D position of $\boldsymbol{M}_i$ at the j^{th} frame is give by

$$\boldsymbol{P}_{ij} = \boldsymbol{R}_j \boldsymbol{M}_i + \boldsymbol{D} \tag{1}$$

The point $\boldsymbol{P}_{ij} = [\boldsymbol{X}_{ij}, \boldsymbol{Y}_{ij}, \boldsymbol{Z}_{ij}]^T$ is measured with respect to the camera coordinate. From perspective projection, the observed 2D image point is given by

$$x_{ij} = f\frac{X_{ij}}{Z_{ij}}, \qquad y_{ij} = f\frac{Y_{ij}}{Z_{ij}}, \tag{2}$$

where f is the focal length of the camera. If the position of the observed image point is $[\tilde{x}_{ij}, \tilde{y}_{ij}]^T$, then the re-projection error, defined as the mean-square error between the observed and the re-projected image points, is

$$e = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}[(\tilde{x}_{ij} - x_{ij})^2 + (\tilde{y}_{ij} - y_{ij})^2]} \tag{3}$$

The goal of the model reconstruction is to find the optimal values for M, R and D that minimizes the cost function in (3). From (2), if $P_{ij} = [X_{ij}, Y_{ij}, Z_{ij}]^T$ is a solution that generates the re-projected data $[x_{ij}, y_{ij}]^T$, then so is $P_{ij} = [kX_{ij}, kY_{ij}, kZ_{ij}]^T$, where k is an arbitrary constant. From (1), this follows that if R_j, M_i and D are solutions, then so are R_j, kM_i and kD. Hence we can only recover the model and translation parameters up to a scale factor. The rotation matrix is invariant to scale.

B. Bundle adjustment

From (1) and (2), the observed data is a function of the model and pose parameters. Let $\tilde{M}$ and $\tilde{\theta}$ be the optimal model and pose parameters respectively, where $\tilde{M} = \{\tilde{M}_1, \ldots, \tilde{M}_N\}$ and $\tilde{\theta} = \{\tilde{R}_1, \ldots, \tilde{R}_T, \tilde{D}_1, \ldots, \tilde{D}_T\}$, then $\tilde{x} = f(\tilde{M}, \tilde{\theta})$ and $\tilde{y} = g(\tilde{M}, \tilde{\theta})$. If M and θ are the current estimate of the parameters, so that $\tilde{M} = M + \delta M + \ldots$, and $\tilde{\theta} = \theta + \delta\theta + \ldots$, then by Taylor expansion, we have

$$\tilde{x} = f(M,\theta) + \frac{\partial f}{\partial M}\delta M + \frac{\partial f}{\partial \theta}\delta\theta + \cdots$$

and

$$\tilde{y} = g(M,\theta) + \frac{\partial g}{\partial M}\delta M + \frac{\partial g}{\partial \theta}\delta\theta + \cdots.$$

Keeping only the linear terms, we have

$$\begin{aligned}\delta x = \tilde{x} - x = \frac{\partial f}{\partial M}\delta M + \frac{\partial f}{\partial \theta}\delta\theta, \\ \delta y = \tilde{y} - y = \frac{\partial g}{\partial M}\delta M + \frac{\partial g}{\partial \theta}\delta\theta\end{aligned} \tag{4}$$

so that

$$e = Js \tag{5}$$

where

$$e = \begin{bmatrix}\delta x \\ \delta y\end{bmatrix}, \quad J = \begin{bmatrix}\frac{\partial f}{\partial M} & \frac{\partial f}{\partial \theta} \\ \frac{\partial g}{\partial M} & \frac{\partial g}{\partial \theta}\end{bmatrix}, \quad \text{and} \quad s = \begin{bmatrix}\delta M \\ \delta\theta\end{bmatrix}.$$

The unknown variables $s = [\delta M, \delta\theta]^T$ can be solved using the least square method, as

$$s = (J^T J)^{-1}(J^T e) \tag{6}$$

The estimate is improved iteratively by making $M \to M + \delta M$ and $\theta \to \theta + \delta\theta$. The procedure is repeated until the residual error e cannot be reduced significantly. Given N feature points and T image frames, we have $3N$ unknowns for the model and $6T$ unknowns for the pose parameters (3 unknowns for the angles of rotation per frame, and 3 unknowns for the displacement vector per frame). The total number of unknowns is $3N + 6T$.

C. The constrained bundle adjustment method

The bundle adjustment method described above does not make full use of the constraints available. Since the position of the rotation axis of the turntable is fixed, the displacement vectors $D_1, \ldots, D_T$ are the same for all frames. Furthermore, the rotation matrices $R_1, \ldots, R_T$ of the object should be constrained to model a circular motion around a fixed axis n, where $n = [n_1, n_2, n_3]^T$. To impose the fix position constraint, we make $D_1, \ldots, D_T = D$. To impose the rotational constraint, we use the Rodrigue formula to model the rotational motion, in which

$$R_j = I\cos\phi_j + (1-\cos\phi_j)nn^T + \sin\phi_j \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} \tag{7}$$

where ϕ_j is the rotation angle at the j^{th} frame. To facilitate the differentiation of R with respect to n, we represent the axis n by the spherical coordinate, in which $n = [\cos(\alpha)\cos(\beta), \cos(\alpha)\sin(\beta), \sin(\alpha)]^T$. The motion of the object can be completely described by $\theta = \{D, \alpha, \beta, \phi_1, \ldots, \phi_T\}$ - a total of $T + 5$ unknowns. Comparing with the $6T$ unknown pose parameters in the unconstrained bundle adjustment, the total number of unknowns has been reduced drastically.

With this new set of unknown parameters, we repeat the steps described in equations (5) and (6). The Jacobians $\partial f / \partial M$, $\partial f / \partial \theta$, $\partial g / \partial M$ and $\partial g / \partial \theta$ in (5) are tedious to derive manually. Instead, the symbolic computational package in Matlab was used to obtain the derivatives.

D. Initialization

We use the bundle adjustment method to obtain an initial value of the pose parameters $[R_1, \ldots, R_T]$ and $[D_1, \ldots, D_T]$. To enforce the displacement constraint, we set $D = \frac{1}{T}\sum_{j=1}^{T} D_j$. Since the model's orientation is measured using the first frame as reference, therefore R_1 is a 3×3 identity matrix. To estimate the rotational axis n, note that $\boldsymbol{Rn} = \boldsymbol{n}$, *i.e.*, the rotational axis is not changed by rotation. Hence $(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{n} = 0$. The optimal $\boldsymbol{n}$ for a given $[\boldsymbol{R}_2, \ldots, \boldsymbol{R}_T]$ can be obtained from the equation

$$\boldsymbol{An} = 0, \qquad \text{where } \boldsymbol{A} = \begin{bmatrix} \boldsymbol{R}_2 - \boldsymbol{I} \\ \vdots \\ \boldsymbol{R}_T - \boldsymbol{I} \end{bmatrix} \tag{8}$$

This follows that the vector $\boldsymbol{n}$ is in the null space of $\boldsymbol{A}^T\boldsymbol{A}$, which corresponds to the eigenvector in $\boldsymbol{A}^T\boldsymbol{A}$ that has the zero eigenvalue. In practice, the eigenvector that has the smallest eigenvalue is picked.

To find $\phi_2, \ldots, \phi_T$, rearrange the equation (7) into the form $\boldsymbol{Bx} = \boldsymbol{b}$, where $\boldsymbol{x} = [\cos\phi_j, \sin\phi_j]^T$. Since $\boldsymbol{R}$ and $\boldsymbol{n}$ are known, $\boldsymbol{x}$ can be solved. Depending on the sign of $\cos\phi_j$ and $\sin\phi_j$, the angle ϕ_j obtained is between 0 to 2π .

E. The measurement of 3D residual error

The quality of the reconstruction is usually measured by the residual 2D re-projection error as shown in (3). However, since a small 2D residual error does not necessarily translate into a small 3D residual error, a better measure is to find the absolute difference between the original and reconstructed 3D models.

Let $\tilde{\boldsymbol{M}}_i$ and $\boldsymbol{M}_i$ be the corresponding 3D points in the original and reconstructed models respectively. Both models are measured with respect to the first frame of the sequence, i.e. the rotation matrix $\boldsymbol{R}_1$ is an identity matrix, so that the orientations of both models are aligned to the same direction. The reconstructed model $\boldsymbol{M}$ is differed from the original model $\tilde{\boldsymbol{M}}$ by i) an arbitrary scale factor, and ii) an arbitrary shift between the two models' centers. Let $\boldsymbol{k}$ be the unknown scale factor and $\boldsymbol{d}$ be the shift between the models' centre, then the average 3D residual error can be defined as

$$\boldsymbol{e} = \frac{1}{\boldsymbol{N}} \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{M}}_i - \boldsymbol{k}(\boldsymbol{M}_i - \boldsymbol{d}) \right\|^2 \tag{9}$$

Given the set of model points $\{\tilde{\boldsymbol{M}}_1, \ldots, \tilde{\boldsymbol{M}}_N\}$ and $\{\boldsymbol{M}_1, \ldots, \boldsymbol{M}_N\}$, we can find the optimum scale factor $\boldsymbol{k}$ and shift factor $\boldsymbol{d}$ that minimizes the error in (9) using least square method. Once $\boldsymbol{k}$ and $\boldsymbol{d}$ are known, the residual 3D error can be found. However, since the scale of the residual error is quite arbitrary, it is more meaningful to normalize the residual error by the size of the original model, so that the error can be expressed in percentage, as

$$\boldsymbol{perr} = \frac{\sum_{i=1}^{N} \left| \tilde{\boldsymbol{M}}_i - \boldsymbol{k}(\boldsymbol{M}_i - \boldsymbol{d}) \right|}{\sum_{i=1}^{N} \left| \tilde{\boldsymbol{M}}_i \right|} \times 100\% \tag{10}$$

III. EXPERIMENTAL RESULTS

A. Simulation results

In this section, we compare the performance of the constrained and unconstrained bundle adjustments by simulation. We use Matlab to generate a random 3D model of 200 points. A sequence of sixty images was generated from the model under perspective projection. Different amount of zero-mean Gaussian noise were added to the images in order to model the tracking error. The noise added represents the lower bound of the 2D re-projection error that can be obtained if the model is recovered perfectly. At each noise level, 50 independent runs were made, and the average results are reported here. We had also simulated the effect due to occlusion. On average about half of the points were occluded by the object itself. The occluded points were represented by NaN (Not-a-Number) in the image sequence.

Noise added (unit: pixel)	Unconstrained bundle adjustment	Constrained bundle adjustment
0	00.0	0.0
1	0.97	0.98
2	1.95	1.96
3	2.91	2.94
4	3.89	3.93
5	4.84	4.90
6	5.83	5.89
7	6.79	6.85
8	7.77	7.84
9	8.78	8.86

Table 1. Residual 2D errors of the constrained and unconstrained bundle adjustments

Table 1 compares the average 2D residual error between the two methods. The residual errors obtained by both methods were smaller than the noise added. This shows that both methods had been trying too hard, and had over-fitted the parameters for the sake of minimizing the re-projection error. Secondly, the unconstrained method produced a smaller residual error than the constrained method at every noise level. This is expected because the lack of constraints made it easier for the unconstrained method to over-fit the data, and to produce a smaller residual error.

Fig. 1 compares the 3D residual errors between the two methods. This time the picture is different. The constrained method is clearly the better one. Although the unconstrained bundle adjustment has a smaller 2D residual error, the reconstructed model obtained from the constrained bundle adjustment has a much smaller 3D residual error. The reduction of the error is between 30 to 40%. The constraints help to model the object motion more realistically, thereby suppress over-fitting, and produce a more accurate model.

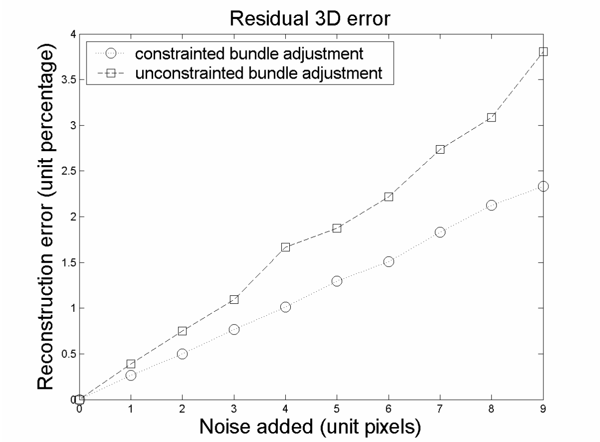

Fig 1. Comparison between the residual 3D errors of the methods

B. Real image results: A model house object on a turn table

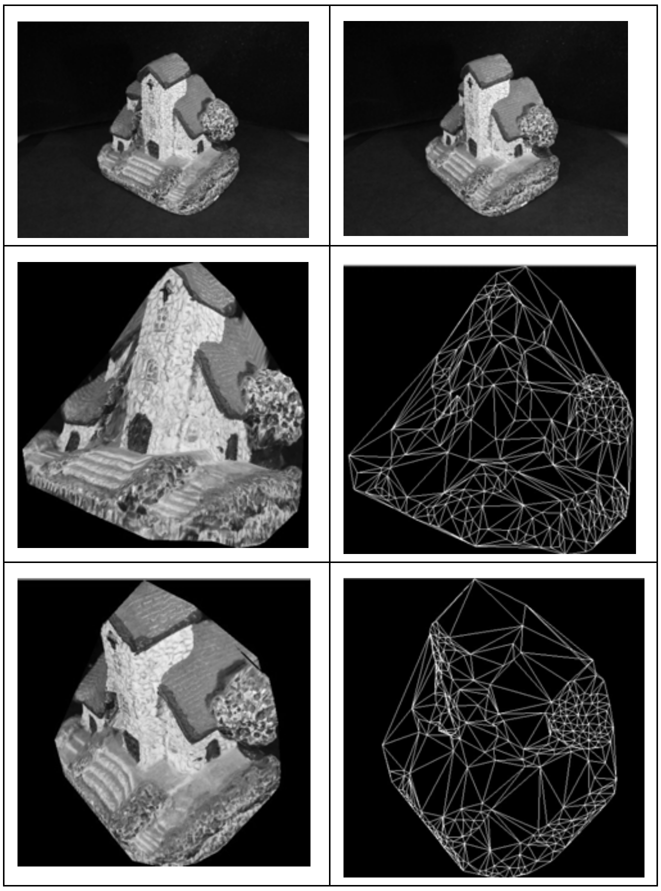

Figure 2: Results for the reconstruction of a model house. The first row contains the first and last (15th) image of the input sequence. The second and third rows contain two reconstructed textured views and their wire-frames.

In Fig.2, features in the image sequence were extracted by the Kanade-Lucas-Tomasi (KLT) Feature Tracker. The features were passed to our algorithm for model reconstruction and saved as a VRML file for display.

IV. Discussion and Conclusion

In this work, we have compared the reconstruction accuracy between classical (unconstrained) bundle adjustment and constrained bundle adjustment. In our experiment, we found that although classical bundle adjustment produced a smaller 2D re-projection error, it actually had a much larger 3D residual error. We believe the 3D residual error is a more meaningful and accurate measure of the reconstruction results. In our experiment, the constrained bundle adjustment method can reduce the 3D error by as much as 30 to 40%, at the cost of only a few seconds in CPU time. A more accurate model reconstruction also leads to a more accurate pose estimation. Hence the constrained bundle adjustment is also useful for reconstruction schemes that depend critically on the accuracy of the pose parameters, such as those that are based on silhouette or space carving.

V. References

[1] B. Curless and M. Levoy, "Better optical triangulation through spacetime analysis," Proceedings., Fifth International Conference on , 20-23 June 1995, pp. 987 -994

[2] M. Pollefeys, Self-calibration and metric 3D reconstruction from uncalibrated image sequences, PhD. thesis, K.U.Leuven, 1999.

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

[4] A. Laurentini, "The visual hull concept for silhouette-based image understanding Laurentini, A.", Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 16 Issue: 2 Feb 1994 Page(s): 150 -162.

[5] K.Y.K Wong and R. Cipolla, "Structure and motion from silhouettes", *ICCV 2001. Proceedings. Eighth IEEE International Conference on Computer Vision*, vol. 2 , pp. 217-222, July 2001.

[6] K.N. Kutulakos, "A Theory of Shape by Space Carving", International Journal of Computer Vision 38(3): 199-218; Jul 2000.

[7] W. Niem and M. Steinmetz, "Camera Viewpoint Control for the Automatic Reconstruction of Object", *Image Processing, 1996. Proceedings., International Conference on* , vol. 3 , pp. 655 -658, Sept. 1996.

[8] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbonm, "Bundle Adjustment - A Modern Synthesis", In Proceedings of the International Workshop on Visual Algorithm: Thoery and Practice, P.P. 298-372, Corfu, Greece, Sept. 1999.

[9] T.J. Broida, S. Chandrashekhar, and R. Chellappa, "Recursive 3-D motion estimation from a monocular image sequence," *IEEE Trans. Aerospace and Electronic Systems*, vol. 26, no. 4 , pp. 639-656, July 1990.