Motion Recovery for Uncalibrated Turntable Sequences Using Silhouettes and a Single Point

Hui Zhang[1], Ling Shao[2] and Kwan-Yee K. Wong[3]

[1] Dept. of Computer Science, United International College, 28, Jinfeng Road, Tangjiawan, Zhuhai, Guangdong, China.
[2] Dept. Video Processing and Analysis, Philips Research Laboratories, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands.
[3] Dept. of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong

Abstract. This paper addresses the problem of self-calibration and motion recovery for turntable sequences. Previous works exploited silhouette correspondences induced by epipolar tangencies to estimate the image invariants under turntable motion and recover the epipolar geometry. These approaches, however, require the camera intrinsics in order to obtain an Euclidean motion, and a dense sequence is required to provide a precise initialization of the image invariants. This paper proposes a novel approach to estimate the camera intrinsics, the image invariants and the rotation angles from a sparse turntable sequence. The silhouettes and a single point correspondence are extracted from the image sequence. The point traces out a conic in the sequence, from which the fixed entities (i.e., the image of the rotation axis, the horizon, the vanishing point of the coordinates, the circular points and a scalar) can be recovered given a simple initialization of the camera intrinsic matrix. The rotation angles are then recovered by estimating the epipoles that minimize the transfer errors of the outer epipolar tangents to the silhouettes for each pair of images. The camera intrinsics can be further refined by the above optimization. Based on a given range of the initial focal length, a robust method is proposed to give the best estimate of the camera intrinsics, the image invariants, the full camera positions and orientations, and hence a Euclidean reconstruction. Experimental results demonstrate the simplicity of this approach and the accuracy in the estimated motion and reconstruction.

1 Introduction

Turntable motion, which also termed as circular motion or single axis motion, has been widely used for generating 3D models in the research of computer vision and graphics. Traditional approaches [19][12] in the modeling process require pre-calibration of the relative pose of the camera as well as the rotation angles. By exploiting the projective geometry of the single axis motion, Fitzgibbon et al. [17] introduced a point-based method to handle uncalibrated camera with unknown rotation angles. Their method for estimating the rotation angles is

based on the computation of trifocal tensors. Jiang et al. [1][2] further extended this approach by making use of at least 2 conic trajectories, and developed an algorithm that avoids the computation of trifocal tensor. However, these techniques cannot be applied if only few point correspondences are available. In this situation, silhouettes become the predominant and stable image feature and offer important clues for determining both the motion and the structure.

Approaches based on silhouettes generally exploit epipolar tangencies [4][11] to induce correspondences. However, such approaches require at least 7 epipolar tangencies and involve a nonlinear optimization with non-trivial initialization. Mendonça et al. [8] proposed a method based on only 2 epipolar tangencies between image pairs. Nonetheless, the recovery of the rotation angles and furthermore the Euclidean reconstruction require the knowledge of the camera intrinsics. Zhang et al. [20] introduced an approach for uncalibrated silhouettes based on a new formulation of the circular point. However, a dense sequence is required to provide a precise initialization of the image invariants. In [6], Wong et al. suggested that this problem can be solved given a good initialization of the image invariants and the camera rotation angles. However, it still requires the knowledge of the camera intrinsics and involves a time-consuming high-dimensional optimization.

In this paper, a novel approach is proposed to solve the above problems. The method exploits the silhouettes and a single point correspondence extracted from the image sequence. The point traces out a conic which in 3D space is actually a circle on a plane orthogonal to the rotation axis. From this conic, the geometry of turntable motion can be fully recovered given an arbitrary initialization of the camera intrinsic matrix. To initialize the camera intrinsics, the principle point can be chosen at the image center. A range is provided from which the initial value of the focal length is selected and a robust method is proposed to give the best estimate of all the parameters. Note since the obtained invariants can only provide 3 independent constraints for the camera intrinsics [20], we only refine 3 parameters of the camera (i.e., the focal length and the 2 coordinates of the principle point). This greatly reduces the search space and thus resulted in high efficiency. The estimated camera intrinsics, along with the image invariants, then allow the recovery of the full camera positions and orientations, and hence a Euclidean reconstruction.

This paper is organized as follows. §2 reviews the fundamentals of turntable motion. Given a point correspondence, §3 proposes a novel method to recover the image invariants based on an assumed camera intrinsic matrix. From these invariants, the camera rotation angles and hence the epipoles can be obtained. §4 presents a method to obtain a good estimate of all the parameters by refining the camera intrinsics. §5 presents the algorithm and the implementations. §6 shows the experimental results and §7 gives the conclusions.

2 Fundamental Theories

Turntable motion refers to the situation where the relative motion between a scene and a camera can be described as a rotation about a fixed axis. Consider

a reference camera $\mathbf{C}_1$ lying on the negative Z-axis of the world coordinate system, and rotating about the Y-axis (see Fig. 1(a)). The relative positions of the camera center describe a circle on a plane Π_h orthogonal to the rotation axis. The image of Π_h is the horizon $\mathbf{l}_h$, whereas the image of the rotation axis is the line $\mathbf{l}_s$ (see Fig. 1(b)). Let the plane defined by the camera center and the rotation axis be Π_s, and we consider 3 orthogonal directions N_x, N_y and N_z, given by the normal direction of Π_s, the Y-axis, and $N_x \times N_y$, respectively. These 3 directions will have vanishing points $\mathbf{v}_x$, $\mathbf{v}_y$ and $\mathbf{v}_z$, respectively. Note that $\mathbf{l}_s$ is also the image of Π_s and hence $\mathbf{v}_x$ and $\mathbf{l}_s$ form a pole-polar relationship w.r.t. the image of the *absolute conic* (IAC) [9]. By construction, N_x is parallel to Π_h, N_y is parallel to Π_s, and N_z is parallel to both Π_h and Π_s. Hence, $\mathbf{v}_x$ must lie on $\mathbf{l}_h$, $\mathbf{v}_y$ must lie on $\mathbf{l}_s$, and $\mathbf{v}_z$ must lie on both $\mathbf{l}_h$ and $\mathbf{l}_s$. The pair of imaged circular points $\mathbf{i}$ and $\mathbf{j}$ for the plane Π_h must also lie on $\mathbf{l}_h$, hence $\mathbf{l}_h = \mathbf{i} \times \mathbf{j}$. If the intrinsic parameters of the camera is kept constant, then all the aforementioned image entities will be fixed.

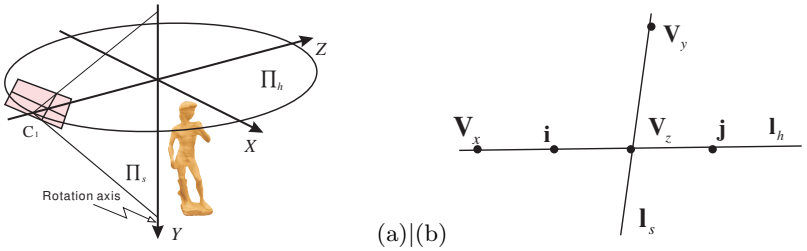

Fig. 1. (a) Geometry of the turntable motion. (b) The 2D Image invariants.

Note that under turntable motion, the fundamental matrix relating any 2 views can be explicitly parameterized in terms of the image invariants, and is given by [17][8]

$$\mathbf{F}(\theta) = [\mathbf{v}_x]_\times + \kappa \tan\frac{\theta}{2}(\mathbf{l}_s\mathbf{l}_h^\mathbf{T} + \mathbf{l}_h\mathbf{l}_s^\mathbf{T}), \tag{1}$$

where θ is the rotation angle between the 2 views. The terms are in homogeneous coordinates, and the fixed scalar κ is therefore necessary to account for the different scales used in the representations. The epipoles $\mathbf{e}_i$ $(i = 1, 2)$ can thus be obtained from the left and right nullspaces of the fundamental matrix, given by [8][20]

$$\mathbf{e}_i \sim \mathbf{v}_x - (-1)^i \kappa \tan\frac{\theta}{2}\mathbf{l}_s \times \mathbf{l}_h, \tag{2}$$

where the left and right hand sides of $\sim$ are equal up to a scalar. Hence if the fixed entities (i.e., $\mathbf{v}_x$, $\mathbf{l}_s$, $\mathbf{l}_h$ and κ) and the rotation angles θ are estimated, $\mathbf{e}_i$ can be recovered.

3 Recovering Epipoles from the Point Correspondence

Consider a scene point under turntable motion being viewed by a fixed camera (see Fig.2(a)). The trajectory of this point, the conic $\mathbf{C}$, is in fact the projection

of the circle $\mathbf{C}_0$ on a plane orthogonal to the rotation axis. $\mathbf{C}$ can be obtained by fitting a conic to the point correspondence in at least five images. From the camera center $\mathbf{O}_c$ and $\mathbf{C}_0$, a general cone $\mathbf{Q}$ can be formed and represented in homogenous coordinates, by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{C}' & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 \end{bmatrix},$$

where $\mathbf{C}' = \mathbf{K}^{\mathbf{T}}\mathbf{C}\mathbf{K}$ is the intersection between $\mathbf{Q}$ and the plane $Z_c = 1$ in the camera coordinates. Here $\mathbf{K}$ is the unknown camera intrinsic matrix, given by

$$\mathbf{K} = \begin{bmatrix} \alpha f & s & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

where f is the focal length, (u_0, v_0) is the principal point, α is the aspect ratio and s is the skew. $\mathbf{K}$ can be initialized by some pre-defined values.

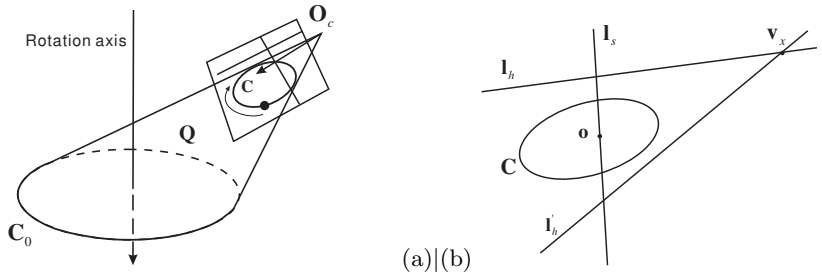

Fig. 2. A point under turntable motion. (a) The trajectory of the point describes a conic $\mathbf{C}$ in the image sequence which is the projection of the circle $\mathbf{C}_0$. (b) The ambiguity of the horizon can be removed by the orientation of $\mathbf{C}_0$.

By diagonalizing $\mathbf{C}'$, a rotation matrix $\mathbf{R}^{\mathbf{T}}_{4\times 4}$ can be recovered to transform $\mathbf{Q}$ to a right elliptic cone

$$\lambda_1 \hat{x}^2 + \lambda_2 \hat{y}^2 + \lambda_3 \hat{z}^2 = 0. \tag{3}$$

Here λ_1, λ_2 and λ_3 are the 3 eigenvalues of $\mathbf{C}'$, with $\lambda_1 > \lambda_2 > \lambda_3$, $\lambda_1, \lambda_2 > 0$ and $\lambda_3 < 0$. Let their corresponding eigenvectors be $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, respectively, the 3×3 upper left sub-matrix $\hat{\mathbf{R}}$ of $\mathbf{R}_{4\times 4}$ can be given by [15]

$$\hat{\mathbf{R}} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]. \tag{4}$$

It can also be derived that the normal vector $\mathbf{n}$ of the supporting plane containing the circle $\mathbf{C}_0$ can be recovered by [15][7]

$$\mathbf{n} = \begin{bmatrix} l \\ m \\ n \end{bmatrix} = \begin{bmatrix} \pm\sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3}} \\ 0 \\ \sqrt{\frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_3}} \end{bmatrix}. \tag{5}$$

Since the normal vector $\mathbf{n}$ of the supporting plane is the unit vector of its vanishing line, the horizon $\mathbf{l}_h$ in the original image can be obtained as

$$\mathbf{l}_h = (\mathbf{KR})^{-\mathbf{T}}\mathbf{n}. \tag{6}$$

Notice that the positive and negative value of l in (5) yield 2 plausible solutions $\mathbf{l}_h$ and $\mathbf{l}'_h$ (see Fig. 2(b)), which correspond to the 2 cases that the camera views the top or bottom of the plane. This ambiguity can be easily removed since it is intuitive to know the facing orientation of the plane normal, by the knowledge of the actual horizon location. Let the correct horizon be $\mathbf{l}_h$, the imaged circle center $\mathbf{o}$ can thus be directly obtained from the pole-polar relationship w.r.t. $\mathbf{C}$, i.e., $\mathbf{o} = \mathbf{C}^{-1}\mathbf{l}_h$. The vanishing point of the rotation axis $\mathbf{v}_y$ can also be obtained from the pole-polar relationship w.r.t. the IAC, i.e., $\mathbf{v}_y = \omega^{-1}\mathbf{l}_h$.

Since $\mathbf{v}_x$ is the vanishing point of the X-axis, it should lie on both $\mathbf{l}_h$ and $\mathbf{l}'_h$, hence $\mathbf{v}_x$ can be obtained as their intersection, i.e.,

$$\mathbf{v}_x = \mathbf{l}_h \times \mathbf{l}'_h. \tag{7}$$

The image of the rotation axis $\mathbf{l}_s$ can then be obtained as

$$\mathbf{l}_s = \mathbf{C}\mathbf{v}_x, \tag{8}$$

from which the vanishing point of the Z-axis can be determined by the intersection of $\mathbf{l}_s$ and $\mathbf{l}_h$, i.e., $\mathbf{v}_z = \mathbf{l}_s \times \mathbf{l}_h$. The pair of circular points $\mathbf{i}$, $\mathbf{j}$ can be obtained as the 2 complex conjugate intersections between $\mathbf{l}_h$ and the conic $\mathbf{C}$. Note under the parametrization of (1), the pair of circular points $\mathbf{i}$, $\mathbf{j}$ of the turntable plane are given by [20]

$$\mathbf{i}, \mathbf{j} \sim \mathbf{v}_x \pm j\,\kappa \mathbf{l}_s \times \mathbf{l}_h, \tag{9}$$

where $j^2 = -1$. Hence the scalar κ can be obtained given the estimated $\mathbf{v}_x$, $\mathbf{l}_s$, $\mathbf{l}_h$ and $\mathbf{i}$, $\mathbf{j}$.

To obtain the epipoles between image pairs, the only missing parameter in (2) is the rotation angle θ. Given the estimated invariants $\mathbf{v}_x$, $\mathbf{l}_s$, $\mathbf{l}_h$, $\mathbf{i}$, $\mathbf{j}$ and κ. θ can be easily estimated with Laguerre's formula [16][1], by

$$\theta = \frac{1}{2j}\log(\{\mathbf{l}_{ox}, \mathbf{l}_{ox'}, \mathbf{i}, \mathbf{j}\}), \tag{10}$$

where $\mathbf{x}$, $\mathbf{x}'$ are the corresponding points between a camera pair with a rotation angle θ and line $\mathbf{l}_{ab}$ is the line joining point $\mathbf{a}$ and $\mathbf{b}$.

4 Refining the Parameters with the Silhouettes

The closed-form solution in the last section is based on a predefined camera intrinsic matrix $\mathbf{K}$. Hence without the knowledge of $\mathbf{K}$, the parameters cannot be obtained correctly if the single point correspondence is the only information provided. In this section, silhouettes are used to solve this problem completely. Consider the upper and lower outer epipolar tangents $\mathbf{l}_i$ and $\mathbf{l}'_i$ $(i = 1, 2)$ of 2 silhouettes (see Fig. 3) passing through the epipole $\mathbf{e}$ and $\mathbf{e}'$, respectively. Note there exists a *harmonic homology* $\mathbf{W}$ [13][8],

$$\mathbf{W} = \mathbf{I} - 2\frac{\mathbf{v}_x \mathbf{l}_s^{\mathbf{T}}}{\mathbf{v}_x^{\mathbf{T}} \mathbf{l}_s}, \tag{11}$$

which relates the 2 corresponding epipolar tangents, i.e., $\mathbf{l}'_i = \mathbf{W}^{-\mathbf{T}}\mathbf{l}_i$. Here $\mathbf{v}_x$ and $\mathbf{l}_s$ are the vanishing point and the imaged rotation axis.

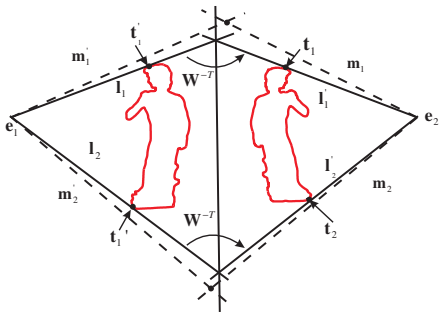

Fig. 3. The overlapping of 2 silhouettes and their epipolar tangents. $\mathbf{l}_1$, $\mathbf{l}'_1$, $\mathbf{l}_2$, $\mathbf{l}'_2$ are the outer epipolar tangent lines. The reprojection errors are given by the geometric distances from the epipolar tangent points to the epipolar lines, and the distances between their intersections with $\mathbf{l}_s$.

Due to the error in $\mathbf{K}$ and the existence of noises, $\mathbf{l}'_i$ and $\mathbf{l}_i$ will not satisfy (11). Let the epipolar tangent points $\mathbf{t}_i$, $\mathbf{t}'_i$ be the intersection points between $\mathbf{l}_i$, $\mathbf{l}'_i$ and the corresponding silhouettes, $\mathbf{s}_i$ be the intersection point between $\mathbf{m}'_i = \mathbf{W}^{-\mathbf{T}}\mathbf{l}_i$, $\mathbf{m}_i = \mathbf{W}^{\mathbf{T}}\mathbf{l}'_i$ (see Fig. 3). The camera intrinsics can be refined by minimizing the cost function

$$Cost(\mathbf{K}) = \sum_{k=1}^{m-1}\{\sum_{i=1}^{2} d_k(\mathbf{t}_i, \mathbf{m}_i) + d_k(\mathbf{t}'_i, \mathbf{m}'_i) + d_k(\mathbf{s}_i, \mathbf{l}_s)\}, \tag{12}$$

where $d_k(\mathbf{p}, \mathbf{l})$ is the distance between the point $\mathbf{p}$ and the line $\mathbf{l}$ for the image pair k, m is the number of images. From the estimated $\mathbf{K}$, the image invariants and the rotation angles can be precisely recovered by the approach in §3.

5 Algorithm and Implementation

The point correspondence can generally be obtained by the Harris point detector [3]. Given a point correspondence in at least five images, a conic $\mathbf{C}$ can be fitted with a least square approach [18]. Another feature, silhouette, is extracted from the image sequence by using of cubic B-spline snake [10]. The cubic B-spline is employed because it can achieve sub-pixel localization accuracy and facilitate the search of epipolar tangents.

As the silhouettes of a turntable sequence provide only 3 independent constraints on the camera intrinsics [20], only 3 parameters (u_0, v_0) and f in $\mathbf{K}$

will be estimated. Under the assumptions of zero skew and unit aspect ratio, we initialize $\mathbf{K}$ with the principal point (u_0, v_0) at image center. Although no information for the focal length is provided, a practical range (f_l, f_u) can be easily given from which the initial focal length $\hat{f}$ is uniformly selected. The image invariants and the camera rotation angles can then be obtained and used to get the epipoles (see §3). According to each initialized $\hat{f}$, the focal length and the principal point can be refined by minimizing the transfer error of the epipolar tangents (see §4). From (f_l, f_u), the best estimate of the refined parameters is selected as the one corresponding to the median of the resulting f. The full camera positions and orientations can then be easily obtained. A bundle-adjustment using Levenberg Marquardt minimization [5] can be applied to refine all the parameters, followed by a Euclidean reconstruction. The complete procedures are summarized in algorithm 1.

Algorithm 1 Generation of 3D model from silhouettes

1: in all images, extract the silhouettes with cubic B-Spline snakes and obtain the single point correspondences;
2: from the point correspondence, obtain a conic $\mathbf{C}$;
3: under the assumptions $s = 0$ and $\alpha = 1$, initialize $\mathbf{K}$ with the principal point $(\hat{u}_0, \hat{v}_0)$ at image center and uniformly sample a value for the focal length $\hat{f}$ from a range (f_l, f_u);
4: for the $(\hat{u}_0, \hat{v}_0)$ and each $\hat{f}$, form $\mathbf{K}$; from $\mathbf{K}$ and $\mathbf{C}$, obtain the invariants $\mathbf{l}_s$, $\mathbf{v}_x$, $\mathbf{l}_h$, the imaged circle center $\mathbf{c}$, circular points $\mathbf{i}$, $\mathbf{j}$, the fixed scalar κ (see §3); from $\mathbf{l}_s$, $\mathbf{v}_x$, obtain the harmomic homology $\mathbf{W}$ by (11);
5: for a pair of image, obtain the camera rotation angles θ with Laguerre's formula (10);
6: obtain the epipoles $\mathbf{e}_i$ by (2), get the epipolar tangents $\mathbf{l}_i$ and $\mathbf{l}'_i$ and transform them by $\mathbf{W}^{-\mathbf{T}}$; minimizing the geometric distance (12);
7: From the different initial values $\hat{f}$, the best estimate of the refined parameters is sampled as the one corresponding to the median of the resulting f;
8: optimize the camera matrix (intrinsic and extrinsic parameters) and all the $n-1$ rotation angles by minimizing the distance between the outer tangent lines and the silhouettes;
9: reconstruct the object model by carving the octree with silhouettes, followed by surface triangulation and texture-mapping.

Note in practise, the corresponding point can be missed in some views along the tracking path. To solve this problem, different strategies can be employed based on the number of missing points. First, when its number p is small compared with the number of images m, the corresponding views C_i $(i = 1, .., p)$ are neglected and only the remaining views in the sequence are used to refine the image invariants and the rotation angles. To search the rotation angle of the view containing the missing point, the epipoles are formed and refined in a 1-dimensional optimization by minimizing the distance from the transformed epipolar tangent line $\mathbf{l}'_i = \mathbf{W}^{-\mathbf{T}}\mathbf{l}_i$ to the silhouette in the second image (see Fig. 3). Second, when the number of missing point p is relatively large, additional

point correspondences can be employed. In this case, we can switch the point correspondence and still obtain the rotation angles with the Laguerre's formula.

6 Experimental Results

Synthetic Data. The synthetic object was composed of 2 intersecting spheres (see Fig. 4(a)) and the point correspondence was obtained by intersecting the 2 edges of a cross in the scene. It rotated about the Y-axis with a constant rotation angle of 20° in a sequence of 18 views. The camera had a focal length $f = 680$, and its principal point was set intentionally to a point $(u_0, v_0) = (345, 265)$ away from the image center. The range of the focal length was set to be $(300, 1100)$ and the initial value $\hat{f}$ was chosen by a equal step length of 100. The corresponding points and the silhouettes of the object were corrupted with Gaussian noise of 15 different levels from 0 to 1.5 pixels. For each level, 50 independent trials were performed using our algorithm. Fig.4(b) shows the average of the percentage errors of f and (u_0, v_0). Fig.4(c) shows the average of the rms error of the rotation angles. It can be seen that the errors increase linearly with the noise level. Under the noise level of 1 pixel, Table 1(a) shows the estimated f converged under different initialization $\hat{f}$. The parameters corresponding to the median of f is selected as the best estimate. Table 1(b) shows the estimated intrinsic parameters and also the percentage errors w.r.t the focal length (Zhang [21]). Fig.4(d) shows the corresponding rotation angles which has a rms error of 0.19°.

(a)

initial $\hat{f}$	300	400	500	600	700	800	900	1000	1100	1200
Estimated f	-	665.3	681.6	694.1	692.6	704.8	698.1	701.3	703.5	684.7

(b)

Approach	f	u_0	v_0
Ground truth	680	345	264
Estimated value	694.06	337.54	239.4
Percentage error	2.07%	1.09%	3.60%

Table 1. Synthetic experiment under noise level of 1. (a) Average of f estimated in 50 experiments from different initial values. (b) Estimated camera parameters from images of turntable motion.

Real Scenes. Real experiments were carried out to test the feasibility of the approach. The first sequence consisted of 18 images of a vase (see Fig. 5(a)). The image had a resolution of 640×480. Each image was taken by sequentially rotating the object by 20° on a manually operated turntable with a resolution of 0.01°. The silhouettes were extracted and the point correspondence was obtained by intersecting 2 fixed edges of the underlying cardboard. In Fig. 5(c), a conic was fitted to the corresponding points. The range of the focal length was set to be $(400, 4200)$ and the initial value $\hat{f}$ was chosen by a equal step length of 200. Table 2 shows the estimated focal length f under the different initial values $\hat{f}$

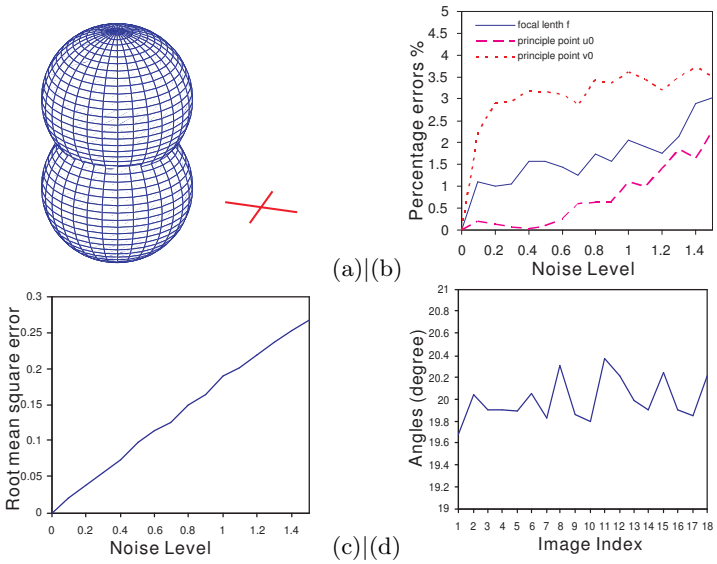

Fig. 4. Synthetic experiment. (a) The synthetic object is composed of 2 intersecting spheres. A cross is given for providing the point correspondence. (b) Average of the relative errors of the camera intrinsic parameters under 15 different noise levels. (c) The rms errors of the rotation angles under 15 different noise levels. (d) The recovered rotation angles under the noise level of 1 pixel.

where '-' indicates the search has a large residual error under that initial value. It can be seen that f is converged. The parameters corresponding to the median of f was selected as the best estimate. Fig.5(e) shows the corresponding rotation angles, with a rms error of only 0.29°, which is about same as the rms error of 0.2° for the same vase sequence in [8]. Under the assumption of a natural camera (zero-skew and unit aspect ratio), columns 2-4 of table 3 compares the estimated camera matrix with that of the ground-truth, which was obtained with a classical L-shape grid pattern [14]. Fig. 5(g) shows 3 views of the 3D model reconstructed from the estimated motion.

initial $\hat{f}$	400	600	800	1000	1200	1400	1600	1800	2000	2200
Estimated f	-	-	2517.4	2365.3	2320.7	2284.3	2311.6	2259.3	2331.3	2334.7
initial $\hat{f}$	2400	2600	2800	3000	3200	3400	3600	3800	4000	4200
Estimated f	2322.5	2387.1	2334.3	2396.7	2496.5	2292.1	-	2325.1	2434.3	-

Table 2. Estimated focal length from different initial values for the vase sequence.

The second sequence consisted of 18 images of a David model with successive rotation angles of 20° (see Fig. 5(b)). The electronic turntable used has a resolution of 0.2°. The initial values of $\hat{f}$ were chosen by a step length of 50 from

	Vase			David		
-	f	u_0	v_0	f	u_0	v_0
Ground-truth	2389.8	342.83	255.32	814.99	414.14	198.70
Estimated value	2396.7	318.73	237.5	821.24	407.13	243.69
Percentage error	0.29%	1.01%	0.75%	0.77%	0.86%	5.52%

Table 3. Estimated and ground-truth values for the intrinsics of the vase sequence and the David sequence.

the range $(350, 1300)$. The point correspondences were picked up by hand. Due to occlusion, some points were missing and 3 point correspondences were used in calculation (see Fig. 5(d)). Table 4 shows that the estimated focal length f converged. The best estimate of the parameters was selected as the one corresponding to the median of f. Fig.5(e) shows the corresponding rotation angles, with a rms error of only 0.23°. Columns 5-7 of table 3 compares the estimated camera matrix with that of the ground-truth, which was obtained with the L-shape pattern. The reconstructed model is shown in Fig. 5(f), which reflects good qualities of our estimated parameters.

Note that in this work, the focal length f and the principal point were both precisely estimated compared with that of [20]. In [20], only the imaged rotation axis $\mathbf{l}_s$ and the vanishing point $\mathbf{v}_x$ are estimated from the profile of the overlapped image envelope and $\mathbf{v}_x$ can only be recovered with high uncertainty. These lead to bigger errors in the calibration. In this work, however, conic(s) can be fitted to the tracked points with much higher accuracy, and 3 image invariants, namely the "pre-transformed" invariants $\mathbf{l}_s$, $\mathbf{v}_x$ and $\mathbf{l}_h$, can be estimated from such conic(s). The remaining motion information are then concisely encoded and optimized by minimizing the reprojection errors of epipolar lines to the silhouettes. These all allow this method to achieve a higher precision.

7 Conclusions

In this paper, we have presented a practical and efficient approach for recovering the full geometry of turntable motion using silhouettes and a single point correspondence. The system input is only uncalibrated images of a sparse turntable sequence. Given an arbitrarily initialized camera matrix, we have proposed a simple approach to recover the geometry from only the point correspondence. By minimizing the geometric distance errors introduced by the tangent lines, the camera intrinsic parameters are refined. Experiments on 2 image sequences have produced convincing 3D models, demonstrating the practicality of our algorithm.

initial $\hat{f}$	350	400	450	500	550	600	650	700	750	800
Estimated f	-	804.0	808.1	808.6	826.7	815.2	824.9	814.7	815.4	823.9
initial $\hat{f}$	850	900	950	1000	1050	1100	1150	1200	1250	1300
Estimated f	823.9	825.7	817.6	821.6	822.8	841.5	821.2	831.4	821.2	836.6

Table 4. Estimated focal length from different initial values for the David sequence.

References

1. Jiang G., Tsui H.T., Quan L., and Zisserman A. Single axis geometry by fitting conics. In *European Conf. on Computer Vision*, pages 482–488, 2002.
2. Jiang G., Quan L., and Tsui H.T. Circular motion geometry by minimal 2 points in 4 images. In *Proc. ICCV*, pages 221–227, 2003.
3. C. Harris and M.J. Stephens. A combined corner and edge detector. In *Alvey88*, pages 147–152, 1988.
4. Porrill J. and Pollard S. B. Curve matching and setreo calibratoin. *IVC*, 9(1):45–50, 2004.
5. More J.J. The levenberg-marquardt algorithm: Implementation and theory. In *Numerical Analysis*, Lecture Notes in Mathematics 630, pages 105–116, 1977.
6. Wong K.-Y. K. and Cipolla R. Structure and motion from silhouettes. In *Proc. 8th IEEE ICCV*, volume II, pages 217–222, Vancouver, Canada, July 2001.
7. K. Kanatani and W. Liu. 3d interpretation of conics and orthogonality. *CVGIP*, 58(3):286–301, November 1993.
8. P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):604–616, June 2001.
9. Faugeras O.D., Luong Q.-T., and Maybank S.J. Camera self-calibration: Theory and experiments. In *ECCV92*, pages 321–334, 1992.
10. Cipolla R. and Blake A. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112, November 1992.
11. Cipolla R., Åström K., and Giblin P. Motion from the frontier of curved surfaces. In *Proc. 5th ICCV*, pages 269–275, Jun 1995.
12. Szeliki R. Shape from rotation. In *Proc. of IEEE CVPR*, pages 625–630, 1991.
13. Hartley R.I. and Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.
14. Tsai R.Y. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(2):323–344, 1987.
15. R. Safaee-Rad, I. Tchoukanov, K.C. Smith, and B. Benhabib. Three-dimensional location estimation of circular features for machine vision. *IEEE Trans. on Robotics and Automation*, 8:624–640, 1992.
16. J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Clarendon Press, Oxford, UK, 1998, Originally published in 1952.
17. Fitzgibbon A. W., Cross G., and Zisserman A. Automatic 3d model construction for turn-table sequences. In *3D SMILSE*, LNCS 1506, pages 155–170. Springer Verlag, June 1998.
18. Fitzgibbon A. W., Pilu M., and Fisher R. B. Direct least-squares fitting of ellipses. *IEEE TPAMI*, 21(5):476–480.
19. Niem W. Robust and fast modelling of 3d natural objects from multiple views. In *Proc. SPIL*, volume 2182, pages 388–397, 1994.
20. H. Zhang, G. Zhang, and K.-Y. K. Wong. Auto-calibration and motion recovery from silhouettes for turntable sequences. In *Proc. BMVC*, volume 1, pages 79–88, Oxford, UK, September 2005.
21. Z.Y. Zhang. Camera calibration with one-dimensional objects. *PAMI*, 26(7):892–899, July 2004.

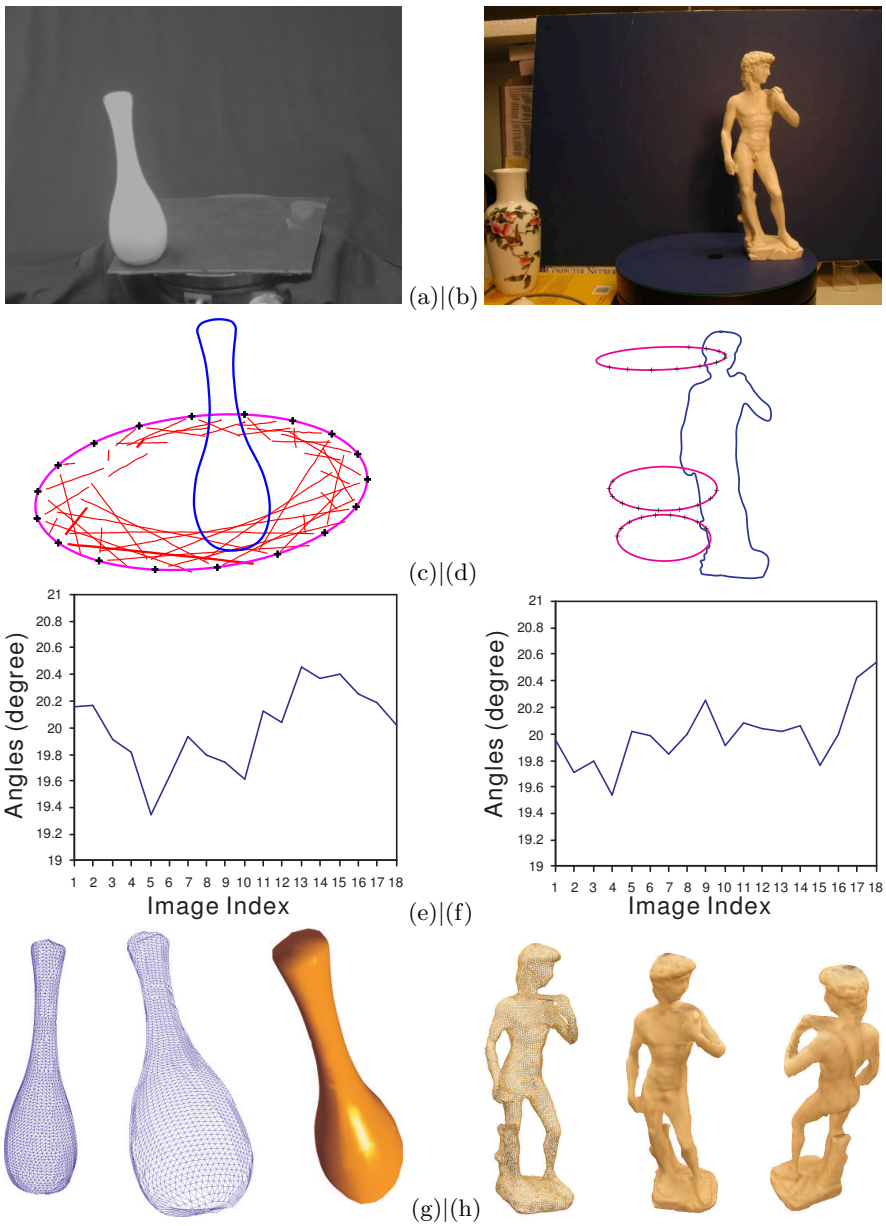

Fig. 5. Real experiments. The 1st column is for the vase sequence and the 2nd is for the David sequence. (a)&(b) A image of the sequence. (c) For the vase sequence, one silhouette and all extracted edges are overlapped. The corresponding point is given by '+' to which a conic is fitted. (d) For the David sequence, 3 points have to be used due to occlusion. (e)&(f) The recovered rotation angles. (g)&(h) Three views of the reconstructed 3D models.