Self-Calibration Using Constant Camera Motion

Xiaochun Cao
University of Central Florida
Computational Imaging Lab
xccao@cs.ucf.edu

Jiangjian Xiao
Sarnoff Corporation
Princeton, NJ 08540, USA
jxiao@sarnoff.com

Hassan Foroosh
University of Central Florida
Computational Imaging Lab
foroosh@cs.ucf.edu

Abstract

This paper investigates using constant inter-frame motion for self-calibration from an image sequence of an object rotating around a single axis with varying camera internal parameters. Our approach is based on the facts that in many commercial systems rotation angles are often controlled by an electromechanical system, and the inter-frame essential matrices are invariant if the rotation angles are constant but not necessarily known. It is shown that recovering camera internal parameters is possible by making use of the equivalence of essential matrices, which relate the unknown calibration matrices to the fundamental matrices computed from the point correspondences. Experimental results on both synthetic and real sequences are presented to determine the accuracy and the robustness of the proposed algorithm.

1 Introduction

Acquiring 3D models from turn-table sequences is widely used due to its simplicity and robustness. Fitzgibbon et. al. [1] recover unknown rotation angles from uncalibrated image sequences based on a projective geometry approach and multi-view geometric constraints. Mendonça [8] recovered the circular motion by using surface profiles. Recently, Jiang et. al. developed new methods to compute single axis motion by either fitting the conic to the locus of the tracked points [6] or computing a plane homography from a minimal of two points in four images [5].

However, most of these methods deal with the case in which the camera has fixed internal parameters, and utilize the fixed image entities of the circular motion. These fixed image entities include two lines: one is the image of the rotation axis, a line of fixed points, while the other one, called the horizon line, is the image of the vanishing line of the horizontal planes. Under the assumption of the fixed camera internal parameters, there are two points located at the intersection of the absolute conic with the horizon line, that remain fixed in all images. Actually, these two fixed points are the images of the two circular points on the horizontal planes [4], and can be found by the intersections of conic loci of corresponding points since the trajectories of space points are circles in 3D space and intersect in the circular points on the plane at infinity. However, these entities are fixed only when the camera has fixed internal parameters.

This paper concentrates on the situation where the camera is free to zoom and focus, and assume rotation angles are often controlled by an electromechanical system [9, 10], i.e. they are constant. It is shown that recovering camera internal parameters is possible by making use of the equivalence of essential matrices. Different from the existing self-calibration methods, our algorithm makes use of constant inter-frame camera motion. In addition, we develop a novel linear algorithm for estimating the relative focal lengths of a camera for different frames. The method needs only a set of fundamental matrices as input, but requires no projective bundle adjustment before self-calibration.

The rest of the paper is organized as follows. In Section 2, a practical calibration method, making use of constant inter-frame motion, is developed. A simple linear solution is also given which can be used as an initialization. The method is then validated through the experiments on both simulated and real data in Section 3. Finally, Section 4 concludes the chapter with discussions on this work.

2 The Method

2.1 Non-linear Method

In a turn-table sequence, the i^{th} camera matrix can be factorized as $\mathbf{P}_i = \mathbf{K}_i[\mathbf{R} \mid \mathbf{t}]$, since our camera is static and thus has the same $\mathbf{R}$ and $\mathbf{t}$ through all views. Let $\mathbf{R}_\theta$ denotes the 3×3 rotation matrix of the object, where the relative rotation angles θ between neighboring views are constant. Therefore, after applying the rotation, the projective transformation of the i^{th} frame becomes $\mathbf{K}_i\ [\mathbf{R}\mathbf{R}_\theta^i \mid \mathbf{t}]$. This means that the new camera center is located at $-(\mathbf{R}\mathbf{R}_\theta^i)^{\mathbf{T}}\mathbf{t}$, with new rotation matrix $\mathbf{R}\mathbf{R}_\theta^i$. Note

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

that the equality is also true for non constant rotations. It is ready to show that the essential matrix has the form

$$\mathbf{E}_{i,i+1} = [(\mathbf{I} - \mathbf{R}\mathbf{R}_\theta\mathbf{R}^T)\mathbf{t}]_\times \mathbf{R}\mathbf{R}_\theta\mathbf{R}^T, \tag{1}$$

where $[\cdot]_\times$ is the notation for the skew symmetric matrix characterizing the cross product. Since $\mathbf{R}$, $\mathbf{t}$ and $\mathbf{R}_\theta$ are all constants, the inter-frame essential matrices are invariant.

The equality of essential matrices can be expressed as:

$$\mathbf{K}_{i+2}^T\mathbf{F}_{i+1,i+2}\mathbf{K}_{i+1} \sim \mathbf{K}_{i+1}^T\mathbf{F}_{i,i+1}\mathbf{K}_i, \tag{2}$$

where $\mathbf{F}_{i,i+1}$ is the fundamental matrix between the i^{th} and $(i+1)^{th}$ views. A solution for the camera matrices may be obtained using a non-linear least squares algorithm. The parameters to be computed are the unknown calibration matrix $\mathbf{K}_i$ and the following criterion should be minimized:

$$\min \sum_{i=1}^{n-2} \|\mathbf{K}_{i+2}^T\mathbf{F}_{i+1,i+2}\mathbf{K}_{i+1} - \mathbf{K}_{i+1}^T\mathbf{F}_{i,i+1}\mathbf{K}_i\|_F^2, \tag{3}$$

where the subscript F indicates the use of the Frobenius norm, and $\mathbf{K}_{i+2}^T\mathbf{F}_{i+1,i+2}\mathbf{K}_{i+1}$ and $\mathbf{K}_{i+1}^T\mathbf{F}_{i,i+1}\mathbf{K}_i$ are both normalized to have unit Frobenius norm. We enforce that two of the essential matrices' singular values are equal and the third one is zero using SVD. In our implementation, we found that the final results are sensitive to errors in the computed fundamental matrices. Therefore, we use the methods [4] that minimize the reprojection errors to compute the fundamental matrices between pairs of images.

2.2 Linear approach

This linear solution can be obtained by assuming zero skew, known aspect ratio and the principal point. For instance, we set the principal point $\mathbf{u}_0$ to $(0,0)$, and the aspect ratio to one. These assumptions yield:

$$\mathbf{E}_{i,i+1} = \begin{bmatrix} f_{i+1}F_i^1 f_i & f_{i+1}F_i^2 f_i & f_{i+1}F_i^3 \\ f_{i+1}F_i^4 f_i & f_{i+1}F_i^5 f_i & f_{i+1}F_i^6 \\ f_iF_i^7 & f_iF_i^8 & F_i^9 \end{bmatrix}, \tag{4}$$

where f_i is the focal length of i^{th} camera, and F_i^k denotes, in a row-major order vector, the components of $\mathbf{F}_{i,i+1}$. It is shown in [8] that whenever the optical axes of the i^{th} and $(i+1)^{th}$ cameras intersect, F_i^9 is equal to zero, since, in this case, the principal points must satisfy the epipolar constraint, i.e. $\mathbf{u}_0^{i+1^T}\mathbf{F}_{i,i+1}\mathbf{u}_0^i = 0$, where $\mathbf{u}_0^{i+1} = \mathbf{u}_0^i = [0\ 0\ 1]^T$. Note that this special case is easy to be detected since all the essential matrices are equal up to an unknown scale. In other words, we only need to check the element, F_i^9, of one inter-frame fundamental matrix $\mathbf{F}$.

Although the case where F_i^9 is equal to zero is a critical motion for Kruppa-based methods [7], the following equations based on the equivalence of the inter-frame essential matrices are still able to provide enough constraints for the computation of the relative focal lengths:

$$F_i^kF_{i+1}^jf_{i+1} - F_i^jF_{i+1}^kf_i = 0, \quad k = 3,6; \tag{5}$$

$$F_i^kF_{i+1}^jf_{i+2} - F_i^jF_{i+1}^kf_{i+1} = 0, \quad k = 7,8, \tag{6}$$

where $j = 1,2,4,5$. Therefore, we can build a matrix $A_{16(n-2)\times n}$, whose null space provides the solution of the focal lengths up to a scale κ. $16 = 4 \times 2 \times 2$ are counted as: the number of options of j (four), the number of options of k (two), and the rest two counters for the two equations (5,6). There are several options to compute κ. In our implementation, we first compensate the effects of varying focal lengths for each image point, $\mathbf{x}_{ij}$ of the i^{th} image, as $\hat{\mathbf{x}}_{ij} = \mathbf{x}_{ij}f_k/f_i$, where f_k could be the focal length of any reference frame k. We then use the existing method [6] to obtain an initial solution of the focal length, f_k, of the reference frame, and therefore those of the other frames.

2.3 Two-stage Optimization

Similar to [6], we first improve the results by enforcing the global constraint that the projected trajectories of 3D points should be conics after compensating for the focusing and zooming effects. Before the conic enforcement, the image points $\mathbf{x}_{i,j}$ are compensated as,

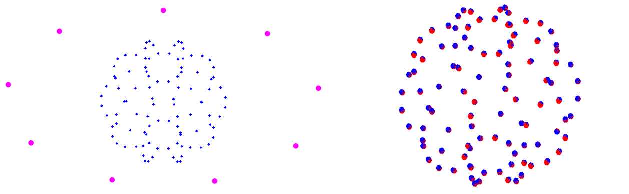
$$\hat{\mathbf{x}}_{ij} = \mathbf{K}_k\mathbf{K}_i^{-1}\mathbf{x}_{ij}, \tag{7}$$

where $\mathbf{K}_k$ is the camera calibration matrix of the reference view. After the compensation, the conic property of the correspondence tracks are fully restored, where the entities related to the conic and plane motion become fixed again.

After the refined camera matrices are obtained, the 3D points or structure can be determined by triangulation from two or more views [3]. In order to minimize the overall reconstruction errors and to further refine the estimated camera parameters, here we use a bundle adjustment approach [11] explicitly enforcing another available constraint: static camera and constant rotation angle. Given n images and m corresponding image points, the maximum likelihood estimate (MLE) can be obtained:

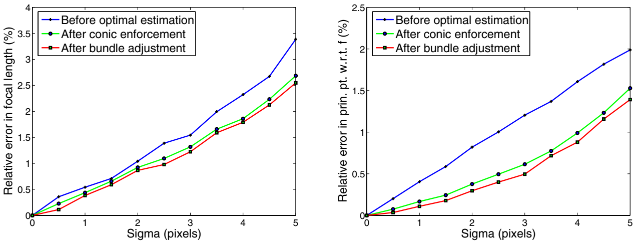
$$\arg\min_{\Theta_2} \sum_{i=1}^{n}\sum_{j=1}^{m} d^2(\mathbf{x}_{ij}, \mathbf{K}_i[\mathbf{R}\mathbf{R}_\theta^i|\mathbf{t}]\mathbf{X}_j), \tag{8}$$

where $\Theta_2 = \{\mathbf{K}_i, \mathbf{R}, \mathbf{R}_\theta, \mathbf{t}, \mathbf{X}_j\}$, and $d(\cdot,\cdot)$ is the distance function between the image measurement $\mathbf{x}_{ij}$ and the projection of the estimated 3D point $\mathbf{X}_j$. However, as shown in [1, 4], the circular motion has the fundamental ambiguity on the vertical apex, $\mathbf{v}$, which causes unknown ratios between the horizontal and vertical direction for the 3D reconstruction. In order to remove this ambiguity, we assume a unit aspect ratio and zero skew for all cameras and specify a reasonable choice of the aspect ratio of the object.



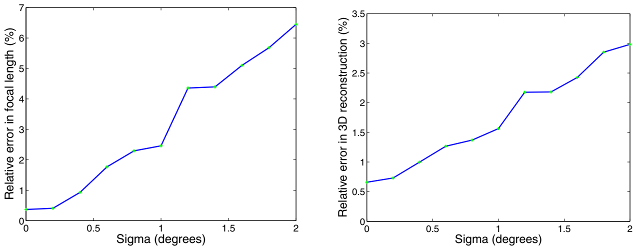
(a) Scene setup (b) Reconstructed points

Figure 1. (a) Magenta points denote the positions of cameras. (b) Blue cubes denote the ground truth while red cubes are reconstructed ones.



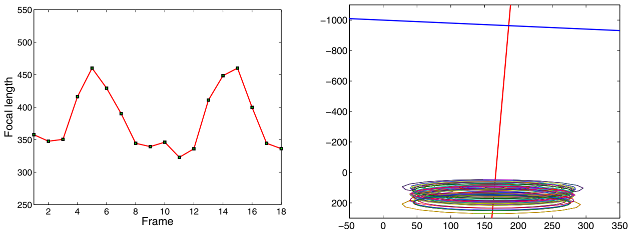
(a) focal length (b) principal point

Figure 2. Performance in a function of noises.



(a) focal length (b) 3D metric error

Figure 3. Performance under errors in rotation angles.



(a) (b)

Figure 4. (a) Computed focal lengths of the Tylenol sequence. (b) The conics, rotation axis (red) and horizontal line (blue) of the compensated frames.

3 Experimental Results

3.1 Computer simulation

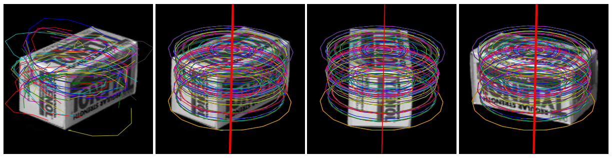
The synthetic scene consists of 100 points uniformly distributed on a sphere with a radius of 200 units and centered at the origin. The synthetic camera is located in front of the scene at a distance of 500 units with three rotation angles (20°, 20° and 15°) between the world coordinate system and the camera coordinate system. In addition to a unit aspect ratio and zero skew, the camera's focal lengths are different for each view, randomly chosen with an expected value of 1000 (in pixels) and a standard deviation of 250. In order to avoid the case that the chosen focal lengths fall outside the reasonable range, e.g. below zero, we limit them to vary between 750 and 1250. The principal point, $\mathbf{u}_0$, had an expected value of $[0\ 0]^T$ with a standard deviation of $20\sqrt{2}$. A view of the equivalent scene, where the camera is moving and the object is stationary, is shown in Fig. 1 (a).

To assess the performance versus noise on the projected image points, nine views are generated. Gaussian noise with zero mean and a standard deviation of $\sigma \leq 5.0$ pixels was added to the projected image points. The estimated camera parameters were then compared with the ground truth. We measured the relative errors of focal lengths and the principal points with respect to the ground truth focal lengths, while varying the noise level from 0.5 pixels to 5.0 pixels. At each noise level, we perform 100 independent trials, and the averaged results of the proposed self-calibration algorithm are shown in Fig. 2. Using our two-stage optimization, the results are refined from a coarse starting point to a fine level for both f and $\mathbf{u}_0$. Fig. 1 (b) shows the 3D reconstructed scene in one trial at the noise $\sigma = 2.5$ pixels.

Another experiment (nine views) is carried out to evaluate how sensitive the algorithm is to noise in the rotation angles. Gaussian noise with zero mean and a standard deviation of $\sigma \leq 2.0$ degrees was added to the rotation angles. Considering the fact that extracted feature points will in practice be affected by noise, we also added a typical noise of $\sigma = 1.0$ pixels to all projected image points. The final results after optimal estimation are shown in Fig. 3. The influence of the orientation noise is larger than that of pixel noise (see Fig. 2), which of course depends on the absolute rotation angle between the views. Note that this coincides with the observation in [2]. Notice also that the errors do not go to zero as noise goes towards zero due to the added noise in image projections.

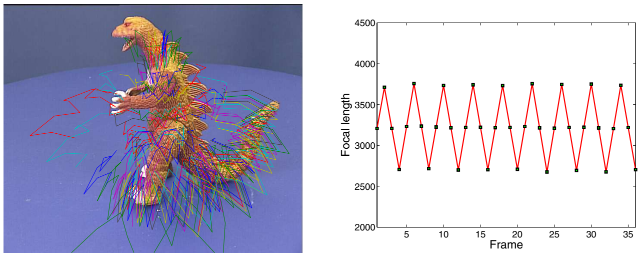
3.2 Real data

The first real sequence (see Fig. 5) is the Tylenol sequence from Columbia Object Image Library. The subset of tracks of the corresponding points estimated using our pre-



(a) (b) (c) (d)

Figure 5. (a) One original frame (frame 7) of the Tylenol sequence and a subset of the point tracks. (b-d) show the estimated conics on compensated frames 7, 5 and 11. The red vertical line is the rotation axis.



(a) (b)

Figure 6. (a) One frame of the dinosaur sequence and a subset of the point tracks. (b) Computed focal lengths.

vious work [12] are shown in Fig. 5 (a), and the computed focal lengths are shown in Fig. 4 (a). In order to evaluate the proposed method, we compensate the frames according to the estimated calibration matrices by using the 7^{th} frame as the reference. The fitted conics and estimated rotation axis are shown in Fig. 5 (b-d) for three compensated frames (frames 7, 5, and 11). We also show the conics, rotation axis and horizontal line of the compensated frames in Fig. 4 (b) (the vertical direction is scaled to show the all entities).

We also tested our approach on the popular dinosaur sequence from the University of Hannover. The focal lengths of the camera is set to change in a zigzag fashion $(0.8 - 1.0 - 1.2)$, by rescaling the original images. When the static camera is free to zoom and focus, the 3D circular trajectory is not projected to a conic anymore (Fig. 6 (a)). The computed focal lengths for the dinosaur sequence is shown in Fig. 6 (b), which is close to the changing pattern in a zigzag fashion. In order to estimate the correctness of our proposed method, the visual hull of the dinosaur can be computed [9] as shown in Fig. 7. The processing of the volume is performed using a resolution in space of 200^3 unit cubes for the bounding box of the dinosaur.

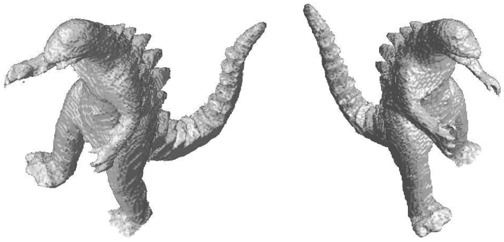


Figure 7. Two views of the 3D reconstruction of dinosaur from silhouettes.

4 Conclusion

Using the invariance property of the inter-frame essential matrices, we present a new solution for camera calibration. Compared to the existing methods, we effectively utilize the prior information, such as constant rotation angle and circular motion, and develop linear algorithm to find an initial solution assuming zero skew and known aspect ratio and the principal point.

References

[1] A. W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3D model construction for turn-table sequences. In *SMILE Wkshp.*, pages 155–170, 1998.

[2] J. Frahm and R. Koch. Camera calibration with known rotation. In *Proc. IEEE ICCV*, pages 1418–1425, 2003.

[3] R. I. Hartley and P. Sturm. Triangulation. In *Proc. CAIP*, pages 190–197, 1995.

[4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[5] G. Jiang, L. Quan, and H. T. Tsui. Circular motion geometry using minimal data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):721–731, 2004.

[6] G. Jiang, H. T. Tsui, L. Quan, and A. Zisserman. Single axis geometry by fitting conics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1343–1348, 2003.

[7] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *J. Math. Imaging Vis.*, 13(2):131–146, 2000.

[8] P. R. S. Mendonça. *Multiview Geometry: Profiles and Self-Calibration*. PhD thesis, University of Cambridge, Cambridge, UK, May 2001.

[9] W. Niem. Robust and fast modelling of 3d natural objects from multiple views. In *Proc. SPIE*, volume 2182, pages 388–397, 1994.

[10] S. Sullivan and J. Ponce. Automatic model construction, pose estimation, and object recognition from photographs using triangular splines. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(10):1091–1097, 1998.

[11] B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle adjustment — a modern synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–373, 1999.

[12] J. Xiao and M. Shah. Two-frame wide baseline matching. In *Proc. IEEE ICCV*, pages 603–609, 2003.