MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

8—9 Auto-calibration of a Rotating and Zooming Camera*

Yongduek Seo† Ki Sang Hong‡
Dept. of EE, Pohang University of Science and Technology (POSTECH), Pohang, Korea
http://cafe.postech.ac.kr

Abstract

We propose a method to calibrate a rotating and zooming camera without 3D pattern, where the internal parameters change frame by frame. First, we show that the calibration is unique up to an orthogonal transformation under the assumption that the skew of the camera is zero. The auto-calibration is possible by analyzing inter-image homographies computed from the matches in the images of the same scene. At least four homographies are needed for auto-calibration in general. When we assume that the aspect ratio is known and the principal point is fixed then one homography will yield camera parameters, and when the aspect ratio is not known with fixed principal point then two homographies are enough. The algorithm is implemented and validated on several sets of synthetic data and real image data.

1 Introduction

Recently, there have been lots of researches for calibrating a camera based only on matches of multiple images. They reported algorithms of auto-calibration for fixed internal camera parameters [5, 2, 9]. Applying the techniques of Projective Geometry, they showed that it is possible to compute the five internal camera parameters when they are fixed for all the views. When camera parameters are varying from image to image, then under the assumption that at least one of five internal parameters is known, auto-calibration is possible[3, 4, 6]. All these auto-calibration methods require that the views be taken at different viewpoints. That is, translation is not zero.

Hartley proposed a self-calibration algorithm given matches of images taken by a rotating camera whose internal parameters are fixed[1]. One limitation of the work is that the algorithm cannot be applied when the images are taken by a zooming or auto-focusing camera, which is common in video images of sports games like soccer or American football. In this paper, we propose a method to auto-calibrate such a rotating and zooming camera so that 3D information can be extracted for future analysis.

*This work is partly supported by IITA (Institute of Information Technology Assessment)

†dragon@indy.postech.ac.kr
‡hongks@vision.postech.ac.kr

2 Self-Calibration is Possible

In this paper we consider a set of rotating cameras with camera matrices $\mathbf{P}_k = \mathbf{K}_k[\mathbf{R}_k|\mathbf{0}]$, where $\mathbf{K}_k$ is the camera calibration matrix of zero skew defined by

$$\mathbf{K}_k = \begin{bmatrix} \alpha_k & 0 & x_k \\ & \beta_k & y_k \\ & & 1 \end{bmatrix}.$$

The parameters in $\mathbf{K}_k$, the *intrinsic parameters*, represent the properties of the image formation system: β_k represents *focal length*, $\gamma_k = \alpha_k/\beta_k$ represents the *aspect ratio* and (x_k, y_k) is called the *principal point*.

Note that given a set of images $\mathcal{I}_0,..., \mathcal{I}_N$ taken from the same location by cameras with the calibration matrices $\mathbf{K}_k$, then there exist 2D projective transformations $\mathbf{H}_k$, taking image $\mathcal{I}_0$ to image $\mathcal{I}_k$, whose matrices are of the form:

$$\mathbf{H}_k = \mathbf{K}_k\mathbf{R}_k\mathbf{K}_0^{-1} \tag{1}$$

where $\mathbf{R}_k$ represents the rotation of the k-th camera with respect to the 0-th. Also, the inter-image homography can be computed from image matches and satisfy the relationship $\mathbf{u}_k = \mathbf{H}_k\mathbf{u}_0$ where $\mathbf{u}_k$ and $\mathbf{u}_0$ are matching points.

Using the inter-image homographies $\mathbf{H}_k$'s computed from image matches, we can find camera matrices $\mathbf{P}_k = \mathbf{K}_k\mathbf{R}_k$, $k = 0, ..., N$, that satisfy the relationship $\mathbf{H}_k = \mathbf{P}_k\mathbf{P}_0^{-1} = \mathbf{K}_k\mathbf{R}_k\mathbf{K}_0^{-1}$. Notice that given one such sequence of camera matrices $\mathbf{P}_k, k = 0, ..., N$, $\mathbf{P}_k\mathbf{Q}$ may be also a possible choice of camera matrices, where $\mathbf{Q}$ is a nonsingular 3×3 matrix, because they also produce the same inter-image homographies. Now we need a lemma to go further [4, 6].

Lemma 1 *A camera matrix* $\mathbf{P} = \mathbf{KR} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3]^{\mathrm{T}}$ *represents a zero-skew camera if and only if*

$$(\mathbf{p}_1 \times \mathbf{p}_3) \cdot (\mathbf{p}_2 \times \mathbf{p}_3) = 0. \tag{2}$$

Due to this lemma, the projective transformation $\mathbf{Q}_{3\times 3}$ can not be arbitrary because every camera matrix should satisfy the constraint equation (2).

Now it remains to show that given a sequence of camera matrices $\mathbf{P}_k$, $k = 1, ..., N$, which 1) solves the inter-image transformation problem and 2) represents zero-skew cameras, then the only possible transformations $\mathbf{Q}_{3\times 3}$ that preserve the zero-skew camera condition (equation (2)) are the orthogonal transformations.

Denote by $\mathcal{M}_\mathbf{P}$ the manifold of all 3×3 camera matrices defined up to scale. Denote by $\mathcal{M}_{zs}$ the manifold of all camera matrices that represent zero-skew cameras. Denote the group of all projective transformations, represented by 3×3 matrices, by $\mathcal{G}_\mathbf{P}$. Finally, denote by $\mathcal{G}_{zs}$ the group of transformations that preserve the property in Lemma 1, and the group of orthogonal transformations by $\mathcal{G}_\mathbf{O}$:

$$\mathcal{G}_{zs} = \{\mathbf{Q} \in \mathcal{G}_\mathbf{P} | (\mathbf{P} \in \mathcal{M}_{zs}) \Rightarrow \mathbf{PQ} \in \mathcal{M}_{zs}\}$$
$$\mathcal{G}_\mathbf{O} = \{\lambda \mathbf{Q} | \mathbf{QQ}^\top = \mathbf{I}, 0 \neq \lambda \in \mathrm{R}\}$$

It is clear that the group of orthogonal transformations is contained in $\mathcal{G}_{zs}$. If $\mathcal{G}_{zs} = \mathcal{G}_\mathbf{O}$ then it is possible to calibrate cameras uniquely up to orthogonal transformation.

Theorem 1 *Let $\mathcal{G}_{zs}$ denote the class of transformations that preserve the zero-skew camera condition and $\mathcal{G}_\mathbf{O}$ the group of orthogonal transformations. Then*

$$\mathcal{G}_{zs} = \mathcal{G}_\mathbf{O}.$$

Proof: It is clear that $\mathcal{G}_\mathbf{O} \subseteq \mathcal{G}_{zs}$. Now we show that $\mathcal{G}_\mathbf{O} \supseteq \mathcal{G}_{zs}$. Assume that $\mathbf{P}$ represents a zero-skew camera, $\mathbf{Q}$ a projective transformation in $\mathcal{G}_{zs}$. Then, from the definition, $\mathbf{PQ} = \mathbf{KRQ}$ can be re-written in the form of $\mathbf{K}'\mathbf{R}'$ where $\mathbf{K}'$ is a zero-skew calibration matrix and $\mathbf{R}'$ is an orthogonal matrix. Also $\mathbf{UQV}$ has this property for every pair of orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$, since

$$\mathbf{KRUQV} = \mathbf{KR}''\mathbf{QV} = \mathbf{K}'\mathbf{R}'''\mathbf{V} = \mathbf{K}'\mathbf{R}'$$

where $\mathbf{R}''$ and $\mathbf{R}'''$ denote orthogonal matrices. Now, using singular value decomposition of $\mathbf{Q}$ we may write

$$\mathbf{D} = \mathbf{UQV} = \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix}.$$

Suppose that the rotation matrix $\mathbf{R}$ is given by θ degree rotation about x-axis and by ϕ degrees about y-axis (note that the rotation $\mathbf{R}$ is arbitrary)

$$\mathbf{R} = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$
$$= \begin{bmatrix} \cos\phi & \sin\theta\sin\phi & -\cos\theta\sin\phi \\ 0 & \cos\phi & \sin\phi \\ \sin\phi & -\sin\theta\cos\phi & \cos\theta\cos\phi \end{bmatrix},$$

we have

$$\mathbf{RD} = \begin{bmatrix} d_1\cos\phi & d_2\sin\theta\sin\phi & -d_3\cos\theta\sin\phi \\ 0 & d_2\cos\phi & d_3\sin\phi \\ d_1\sin\phi & -d_2\sin\theta\cos\phi & d_3\cos\theta\cos\phi \end{bmatrix}.$$

Now, according to Lemma 1, $\mathbf{RD} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^\top$ is a zero-skew calibration matrix if and only if $(\mathbf{r}_1 \times \mathbf{r}_3) \cdot (\mathbf{r}_2 \times \mathbf{r}_3) = 0$. After some calculation we have

$$\mathbf{r}_1 \times \mathbf{r}_3 = [0, -d_1 d_3 \cos\theta, -d_1 d_2 \sin\theta]^\top,$$
$$\mathbf{r}_2 \times \mathbf{r}_3 = [d_2 d_3 \cos\phi, d_1 d_3 \sin\theta\sin\phi, -d_1 d_2 \cos\theta\sin\phi]^\top$$

and

$$\begin{aligned}(\mathbf{r}_1 \times \mathbf{r}_3) \cdot (\mathbf{r}_2 \times \mathbf{r}_3) &= -d_1^2 d_3^2 \cos\theta\sin\theta\sin\phi + d_1^2 d_2^2 \cos\theta\sin\theta\sin\phi \\ &= d_1^2 \cos\theta\sin\theta\sin\phi(-d_3^2 + d_2^2) \\ &= 0.\end{aligned}$$

That is, $\mathbf{RD}$ is a zero-skew calibration matrix if and only if $d_2 = d_3$. Permutation of the singular values yields $d_1 = d_2 = d_3$. Thus, all singular values of $\mathbf{Q}$ are equal, which means that $\mathbf{Q}$ is an orthogonal matrix. ■

Note that the matrix $\mathbf{Q}$ is related to the selection of the camera coordinate system. Therefore, choosing the camera coordinate system with respect to an image determines implicitly the matrix $\mathbf{Q}$.

3 Estimation Method

From the equation (1), we have

$$\mathbf{H}_k \mathbf{K}_0 \mathbf{K}_0^\top \mathbf{H}_k^\top = \mathbf{K}_k \mathbf{K}_k^\top. \quad (3)$$

Then the number of unknowns is $4N + 4$, given N homographies, and the number of equations is $5N$ because $\mathbf{H}_k$ is defined and can be computed only up to scale, which means that at least four homographies are needed to compute the camera parameters.

If we know the principal points (x_k, y_k), other calibration parameters (α_k, β_k) can be computed using a linear equations. Now let's define a 3×3 matrix $\mathbf{C}_k = \begin{bmatrix} 1 & & x_k \\ & 1 & y_k \\ & & 1 \end{bmatrix}$. By multiplying matrices $\mathbf{C}_k^{-1}$ and $\mathbf{C}_0$ on the left and on the right side of $\mathbf{H}_k$, respectively, we have principal-point-free version of the the equation (3), $\bar{\mathbf{H}}_k \bar{\mathbf{K}}_0 \bar{\mathbf{K}}_0^\top \bar{\mathbf{H}}_k^\top = \bar{\mathbf{K}}_k \bar{\mathbf{K}}_k^\top$, from which we have five equations to compute the scale factors:

$$\begin{bmatrix} h_{11}h_{21} & h_{12}h_{22} \\ h_{11}h_{31} & h_{12}h_{32} \\ h_{21}h_{31} & h_{22}h_{32} \end{bmatrix} \begin{bmatrix} \alpha_0^2 \\ \beta_0^2 \end{bmatrix} = \begin{bmatrix} -h_{13}h_{23} \\ -h_{13}h_{33} \\ -h_{23}h_{33} \end{bmatrix} \quad (4)$$

$$\alpha_k^2 = \frac{\alpha_0^2 h_{11}^2 + \beta_0^2 h_{12}^2 + h_{13}^2}{\alpha_0^2 h_{31}^2 + \beta_0^2 h_{32}^2 + h_{33}^2} \quad (5)$$

$$\beta_k^2 = \frac{\alpha_0^2 h_{21}^2 + \beta_0^2 h_{22}^2 + h_{23}^2}{\alpha_0^2 h_{31}^2 + \beta_0^2 h_{32}^2 + h_{33}^2}. \quad (6)$$

It means that the scale parameters (α_k, β_k) may be parameterized by the principal points, and given principal points the scale parameters are linearly computed. Now we define a nonlinear error function to find the optimal calibration parameters including the principal points. Using the relationship

$$\mathbf{R}_k = \frac{\mathbf{K}_k^{-1}\mathbf{H}_k\mathbf{K}_0}{det(\mathbf{K}_k^{-1}\mathbf{H}_k\mathbf{K}_0)^{\frac{1}{3}}} \quad (7)$$

we minimize the following error function:

$$E = \sum_{k=1}^{N} \left(\left\| \mathbf{R}_k \mathbf{R}_k^\top - \mathbf{I} \right\|_F^2 + \left\| \mathbf{R}_k^\top \mathbf{R}_k - \mathbf{I} \right\|_F^2 \right) \quad (8)$$

Notice that E is a function of principal points. Since the principal points are around image center, a search window may be chosen around the image center, and the algorithm proposed is:

1. Set principal points: $x_k \leftarrow \bar{x}_k$ and $y_k \leftarrow \bar{y}_k$ for $k = 0, ..., N$.
2. Compute α_k and β_k, for $k = 0, ..., N$.
3. Compute $\mathbf{R}_k$ using equation (7) and the error E using equation (8).
4. if E is smaller than the previous one, record the calibration parameters.
5. repeat 1 – 4 for searching area
6. The optimal calibration parameters are the recorded ones.

For nonlinear optimization, we have two approaches. One is the area searching method searching in the whole or a part of image space for the principal points that minimize the error function, and the other is the use of a gradient based minimization algorithm like conjugate gradient method or Levenberg-Maqurdt method. In the latter case, initial values may be obtained by assuming that the principal points are image centers and computing the other calibration parameters using the linear algorithm.

As mentioned previously, we need at least four inter-image homographies for computing time-varying calibration parameters. However, the number of homographies can be reduced if we make some restrictions on camera models. If we assume that the principal point does not move at all in zooming or focusing and the aspect ratio is known, only focal lengths will vary. In this case one inter-image homography is enough to calibrate the camera using the area searching method, which will be discussed in Section 4. In Section 5, we generalize the model by assuming the aspect ratio is unknown, in which case two inter-image homographies are required. Assuming that the principal point is fixed, we will use the area searching method to find the global minimum. Finally, all the calibration parameters are assumed to vary except the skew. This case is the most general and we use an iterative optimization method like the Levenberg-Maqurdt method.

4 Fixed principal point with known γ

Provided that the principal point is fixed during the sequence and the aspect ratio γ is known *a priori*, then only one homography is needed to compute camera parameters. The calibration matrix is now $\mathbf{K}_k = \begin{bmatrix} f_k & & x \\ & f_k & y \\ & & 1 \end{bmatrix}$, and the equations (4) - (6) are now of the form:

$$f_0^2 \begin{bmatrix} h_{11}h_{21} + h_{12}h_{22} \\ h_{11}h_{31} + h_{12}h_{32} \\ h_{21}h_{31} + h_{22}h_{32} \end{bmatrix} = \begin{bmatrix} -h_{13}h_{23} \\ -h_{13}h_{33} \\ -h_{23}h_{33} \end{bmatrix}, \quad (9)$$

$$f_k^2 = \frac{f_0^2(h_{11}^2 + h_{12}^2) + h_{13}^2}{f_0^2(h_{31}^2 + h_{32}^2) + h_{33}^2} = \frac{f_0^2(h_{21}^2 + h_{22}^2) + h_{23}^2}{f_0^2(h_{31}^2 + h_{32}^2) + h_{33}^2}. \quad (10)$$

However, notice that we should not use the equation (9) when the rotation is only about the x-axis or y-axis. When the x-axis is the rotation axis, only one equation $f_0^2(h_{21}h_{31} + h_{22}h_{32}) = -h_{23}h_{33}$ among three equations is valid. When the y-axis is the rotation axis, the second equation $f_0^2(h_{11}h_{31} + h_{12}h_{32}) = -h_{13}h_{33}$ is valid. Details can be found in [7] at our web site. This analysis is important because the axis of the rotation is usually the x-axis, y-axis or the composition of the two axes, and the rotation about z-axis is usually small.

5 Fixed principal point

The calibration matrix is now assumed that the principal point is fixed and the aspect ratio is not known. That is, the calibration matrix is of the form $\mathbf{K}_k = \begin{bmatrix} \alpha_k & & x \\ & \beta_k & y \\ & & 1 \end{bmatrix}$, and two homographies or three images are needed to compute the calibration parameters as well as rotation angles. At this time, one should be careful about the rotation axis. When the rotation axis is only the x-axis for all the input images, it is impossible to compute the scale factors α_k's. Also when the rotation is about the y-axis, one cannot compute the β_k's. This is due to the special form of the rotation matrices in these two cases. When the rotations are about the x-axis, the rotation matrices are of the form

$$\mathbf{R}_k^{X,\theta_k} = \begin{bmatrix} 1 & & \\ & c & -s \\ & s & c \end{bmatrix} \quad (11)$$

where $c = \cos\theta_k$, $s = \sin\theta_k$ and θ_k is the rotation angle between the 0-th camera and the k-th camera. Notice that

$$\mathbf{R}_k^{X,\theta_k} = \mathbf{D}(\lambda, 1, 1)\mathbf{R}_k^{X,\theta_k}\mathbf{D}(\lambda^{-1}, 1, 1)$$

where

$$\mathbf{D}(\lambda, 1, 1) = \begin{bmatrix} \lambda & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

That is, we have multiple solutions for the calibration parameters that satisfy the equation $\mathbf{H}_k = \mathbf{K}_k\mathbf{R}_k\mathbf{K}_0^{-1}$, or more specifically we cannot determine exact α_k's, because if $\mathbf{K}_k$ is a solution then $\mathbf{K}_k\mathbf{D}(\lambda, 1, 1)$ is a solution, too. In the case of rotations about the y-axis, it is impossible to compute unique β_k's for the same reason. In conclusion, the rotation axis should not be purely the x-axis nor the y-axis for unique computation of the calibration parameters. Except for those two cases, the rotation axis may be fixed. This problem was investigated previously for rotating camera of fixed internal parameters in [1].

6 Experiments

Here, we show results of our algorithm using two views due to space limit. One can find details in the long version of this

Noise level (σ in pixel)	f_0	f_1	u	v	r_x	r_y	r_z
0	1000	1100	330	230	10°	10°	0°
0.5 (mean)	1001.4	1101.7	328.6	228.3	9.96	9.99	-0.01
(σ)	15.0	16.9	9.0	9.5	0.22	0.19	0.07
0.7 (mean)	997.2	1097.0	331.7	231.6	10.05	10.01	-0.02
(σ)	21.9	23.8	13.4	13.0	0.28	0.25	0.08
1.0 (mean)	1005.1	1106.5	330.3	229.0	9.95	10.06	-0.01
(σ)	44.7	49.5	19.3	22.8	0.43	0.40	0.11

Table 1. Computation results after 100 runs at each noise level. About one hundred matching points are used in the computation of the homography.

paper [7]. Assuming that the aspect ratio is 1 and the principal points are fixed, auto-calibration can be done using only two views. Table 1 shows the calibration result of 100 runs with 2 image matches for various image noise. Since the noise is added to each of image coordinates, the actual RMS error is $\sqrt{2}$ times the indicated value σ. Note that the principal point is the most sensitive to input image noise. On the contrary, rotation angles are less sensitive to input noise.

Figure 1 shows two video frames of a soccer game. Notice that there are scale change due to zooming as well as rotation. Inter-image homography is estimated by direct iterative error minimization method [8] where initial parameters are obtained using matches of lines and points, and the calibration result is: $f_0 = 1145.9$, $f_1 = 1376.5$ and $(x, y) = (324.5, 181.5)$. Computed rotation angles for the three axes are $(-3.78°, -10.27°, -0.57°)$.

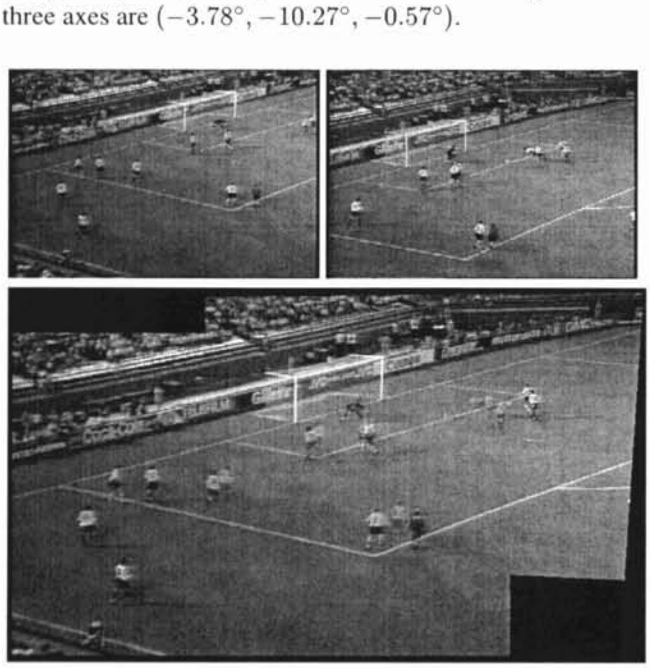

Figure 1. Two images sampled from a soccer game. Mosaic is obtained using direct method of [8]. The result is basically needed in estimating 3D locus of the ball or in virtual view synthesis.

7 Conclusion

We showed that auto-calibration of a rotating and zooming camera without 3D pattern is unique up to orthogonal transformation and implemented and tested the algorithm for synthetic and real data. This algorithm is important for the applications like 3D reasoning from monocular rotating camera in sports games or video re-generation of a scene based on image mosaic.

References

[1] Richard I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1), 1997.

[2] Anders Heyden and Kalle Aström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. ICPR'96*, pages 339–343, 1996.

[3] Anders Heyden and Kalle Aström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. CVPR'97*, 1997.

[4] Anders Heyden and Kalle Aström. Minimal conditions on intrinsic parameters for euclidean reconstruction. In *Proc. Asian Conference on Computer Vision, Hong Kong*, 1998.

[5] Stephen J. Maybank and Olivier D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[6] Marc Pollefeys and Luc Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. Int. Conf. on Computer Vision*, 1998.

[7] Yongduek Seo and Ki Sang Hong. Self-calibration of a rotating and zooming camera. Submitted to IJCV, This paper can be obtained at http://cafe.postech.ac.kr/~dragon/calib.html.

[8] Richard Szeliski and Sing Bing Kang. Direct methods for visual scene reconstruction. In *IEEE Computer Society Workshop: Representation of Visual Scenes*, 1995.

[9] B. Triggs. Autocalibration and the absolute quadric. In *Proc. CVPR'97*, 1997.