A MULTISTAGE CAMERA SELF-CALIBRATION ALGORITHM

Yongying Gao and Hayder Radha
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA
Email: {gaoyongy, radha}@egr.msu.edu

ABSTRACT

We present a new camera self-calibration algorithm that uses a low-complexity multistage approach. We derive a polynomial optimization function with respect to the camera intrinsic parameters, based on the equal singular value property of the essential matrix. In terms of the stability analysis of the intrinsic parameters, we propose a multistage procedure to refine the estimation. Experimental results with both synthetic and real images show the accuracy and robustness of our method.

1. INTRODUCTION

Camera self-calibration has attracted a great deal of attention in the field of computer vision because of its role in automatic 3D reconstruction. Unlike the classical calibration problem, a self-calibration algorithm attempts to find the camera intrinsic parameters from a set of images without the ground truth.

Faugeras et al. [1] proposed a theory of self-calibration expressed by the Kruppa's equations and a numerical method based on the Kruppa's equations. Pollefeys and Gool [2] proposed another self-calibration method using the modulus constraints. Hartley [3] introduced a new self-calibration method based on the equal singular value (ESV) property of the essential matrix [4]. Mendonca and Cipolla [5] extended Hartley's method to the case of a larger sequence of images. Roth and Whitehead [6] provided a stochastic optimization approach to self-calibration.

In this paper, we propose a new multistage self-calibration algorithm based on the ESV property of the essential matrix. Differing from previous approaches [3][5][6], where the optimization function is not explicit with respect to the camera intrinsic parameters, we derive a polynomial optimization function of the intrinsic parameters, and then follow a multistage procedure to refine the results. We applied our method to both synthetic and real images. The experimental results show the accuracy and robustness of our approach when compared with other leading approaches such as the ones proposed in [6][7].

2. SELF-CALIBRATION BASED ON THE ESV PROPERTY OF THE ESSENTIAL MATRIX

2.1 Background of Camera Self-calibration

The camera calibration matrix $\boldsymbol{K}$, which consists of the camera intrinsic parameters, has the following entries:

$$\boldsymbol{K} = \begin{bmatrix} \alpha_u & -\alpha_u \cot(\theta) & u_0 \\ 0 & \alpha_v / \sin(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where α_u and α_v are the focal lengths in pixels along orthogonal axes, u_0 and v_0 represent the coordinates of the principal point, and θ is the skew angle and often considered to be $\pi/2$. The goal of self-calibration is to estimate α_u, α_v, u_0 and v_0.

The essential matrix $\boldsymbol{E}$ represents the epipolar geometry if the camera calibration is already known. Based on the assumption that the camera intrinsic parameters remain unchanged throughout the whole set of images, $\boldsymbol{E}$ is related to the fundamental matrix $\boldsymbol{F}$ [8][9] by

$$\boldsymbol{E} = \boldsymbol{K}^T \boldsymbol{F} \boldsymbol{K} . \quad (2)$$

It is proven in [4] that one of the singular values of $\boldsymbol{E}$ is zero and the other two are equal to each other. The zero singular value condition is automatically satisfied since $\boldsymbol{F}$ is of rank 2 and $\boldsymbol{K}$ is of full rank, while the ESV property establishes a link between the camera relative motion and associated intrinsic parameters.

2.2. Polynomial Optimization Function Based on the ESV Property

We show that the ESV constraint can be expressed as a polynomial with respect to the entries of $\boldsymbol{K}$. In terms of the definition of singular value, the property of the singular values of $\boldsymbol{E}$ corresponds to the property of the eigenvalues of $\boldsymbol{E}^T\boldsymbol{E}$. Let $\boldsymbol{A} = \boldsymbol{E}^T\boldsymbol{E} = (\boldsymbol{K}^T\boldsymbol{F}\boldsymbol{K})^T\boldsymbol{K}^T\boldsymbol{F}\boldsymbol{K}$, then the characteristic equation $\det(\lambda\mathbf{I} - \boldsymbol{A}) = 0$ can be expressed as

$$\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = 0, \quad (3)$$

where l_0, l_1 and l_2 are functions of the entries of $\boldsymbol{A}$ and in turn functions of the entries of $\boldsymbol{K}$ and $\boldsymbol{F}$. The three eigenvalues λ_1, λ_2 and λ_3 of $\boldsymbol{A}$ should satisfy $\lambda_1 = \lambda_2$ and $\lambda_3 = 0$. Substituting $\lambda_3 = 0$ in Eq.(3) leads to an order-reduced equation

$$\lambda^2 + l_2\lambda + l_1 = 0. \quad (4)$$

Furthermore, $\lambda_1 = \lambda_2$ leads to:

$$l_2^2 - 4l_1 = 0. \quad (5)$$

Given $\boldsymbol{F}$, and after some manipulation, we rewrite Eq.(5) in a quartic polynomial explicitly with respect to $x = \alpha_u^2$ and $y = \alpha_v^2$:

$$f(x, y, c_i, i = 1, \cdots, 15) = 0, \quad (6)$$

where $c_i, i = 1, \cdots, 15$ is the coefficient of each item x^4, y^4, $x^3 y$, and so on, and is expressed as a function of u_0, v_0 and the entries of $\boldsymbol{F}$.

Eq.(6) has two advantages. First, $f(x, y, c_i, i = 1, \cdots, 15)$ is a bivariate polynomial with respect to $x = \alpha_u^2$ and $y = \alpha_v^2$, based on the assumption that (u_0, v_0) is fixed. This property enables the computation of exact derivatives, which simplifies the optimization. Second, since the coefficients of x^4 and y^4 are positive numbers, this function monotonously increases as x and y approach infinity. Therefore, the initialization is less critical than it is in a usual optimization problem.

2.3. Weighted Global Optimization Function

From the perspective of numerical analysis, we may achieve better performance if Eq.(6) is weighted. We use a normalized version of Eq.(5) as follows:

$$\frac{l_2^2 - 4l_1}{l_2^2} = 0. \quad (7)$$

Comparing Eq.(7) with Eq.(5), we take $1/l_2^2$ as the weight.

In practice, we have a set of N images so that we can obtain at most $N(N-1)/2$ fundamental matrices. The advantage of using all of the $N(N-1)/2$ fundamental matrices is twofold: first, the redundancy reinforces the numerical robustness; second, it avoids bias towards any given image. Hence we employ the following weighted global optimization function

$$C(x, y) = \sum_{i=1}^{N(N-1)/2} w_i f_i(x, y, c_j^i, j = 1, \cdots, 15), \quad (8)$$

where $f_i(x, y, c_j^i, j = 1, \cdots, 15)$ is the optimization function of the i-th image pair, and the weight $w_i = (1/l_2^2)_i$ is a function of x, y, u_0, v_0 and the entries of $\boldsymbol{F}_i$.

3. MULTISTAGE APPROACH TO CAMERA SELF-CALIBRATION

3.1. Stability Analysis of the Intrinsic Parameters

In practice, we do not directly minimize Eq.(8) with respect to all of the four intrinsic parameters because it is computationally extensive and unstable. In fact, these parameters impact the final 3D reconstruction quite differently. Zhang et al. [10] stated that shifting the principal point from its true position does not cause large distortion of the reconstructed 3D points, based on the assumption that the values of α_u and α_v are correct. In the case that none of the four parameters are known, the offset of the principal point impacts the estimation of α_u and α_v.

However, experiments show that the estimated aspect ratio (α_v/α_u) remains close to its true value while suffering from the offset of the principal point. In [5] it is stated that the estimation of α_v/α_u is very robust to noise. This observation is extended here: the estimation of the aspect ratio is robust to both the noise of the coordinates of the image points and the noise caused by the incorrect location of the principal point.

3.2. A Multistage Algorithm for Self-calibration

Based on the above observation, we formulate our multistage algorithm of self-calibration as follows:

Step 1. Estimate α_u and α_v, assuming that (u_0, v_0) is located at the center of the image. The outcomes are denoted by $\tilde{\alpha}_u^{(1)}$ and $\tilde{\alpha}_v^{(1)}$.

Step 2. Refine the estimation of α_u, u_0 and v_0, assuming $\alpha_v/\alpha_u = \tilde{\alpha}_v^{(1)}/\tilde{\alpha}_u^{(1)}$. The outcomes are $\tilde{\alpha}_u^{(2)}$, $\tilde{u}_0^{(2)}$ and $\tilde{v}_0^{(2)}$.

Step 3. Refine the estimation of α_u and α_v, assuming $(u_0, v_0) = (\tilde{u}_0^{(2)}, \tilde{v}_0^{(2)})$. The outcomes are $\tilde{\alpha}_u^{(3)}$ and $\tilde{\alpha}_v^{(3)}$.

Step 4. Refine α_u, α_v, u_0 and v_0, with the initial conditions $\tilde{\alpha}_u^{(3)}$, $\tilde{\alpha}_v^{(3)}$, $\tilde{u}_0^{(2)}$ and $\tilde{v}_0^{(2)}$. The final outcomes are $\tilde{\alpha}_u$, $\tilde{\alpha}_v$, $\tilde{u}_0$ and $\tilde{v}_0$.

Step 1 and step 3 are accomplished by optimizing the objective function expressed in Eq.(8). In step 2, we need to estimate u_0 and v_0. However, Eq. (8) is not a function of u_0 and v_0. Hence, instead of using Eq.(8), we use the following optimization function [6], which directly computes the singular values of $\boldsymbol{E}$

$$C(x, y) = \sum_{i=1}^{N(N-1)/2} (1 - \frac{\lambda_2^i}{\lambda_1^i}), \quad (9)$$

where λ_1^i and λ_2^i represent the nonzero singular values of $\boldsymbol{E}_i$, in descending order. Eq.(9) is also used in step 4.

4. EXPERIMENTAL RESULTS

4.1. Experimental Results with Synthetic Data

In this experiment, 20 synthetic images (512×512) were generated with 200 points randomly scattered in a cube of edge size 800 centered at $(0,0,2000)$. The intrinsic parameters are chosen as $\alpha_u = 957.8$, $\alpha_v = 891.2$ and

$(u_{0,}v_0) = (279,241)$, to simulate the standard settings of a real camera. We added (to the pixel locations) two types of noise: uniformly distributed noise in [-0.5 pixel, 0.5 pixel], which simulates the quantization error, and Gaussian noise with standard deviation of 1 pixel, which simulates the noise caused from point corresponding match. Fundamental matrices were computed from various numbers (4 to 20) of images using the normalized linear criterion [9].

To evaluate the performance of our method, we compared our algorithm with the one presented in [6], which is equivalent to using only step 4 of our method. This approach, referred to as the RW method, represents an example of an ESV-based state-of-the-art algorithm. Each estimation task based on a certain number of images was repeated $N = 100$ times. We measured the average value of the relative error $\varepsilon_\alpha = \frac{1}{N\alpha}\sum_{i=1}^{N}\left|\tilde{\alpha}_i - \alpha\right|$. Here $\alpha \in \{\alpha_u, \alpha_v\}$ is the true value and $\tilde{\alpha}_i \in \{\alpha_{u,i}, \alpha_{v,i}\}$ is the estimated value from the *i*-th round of experiment. Below, we present only the results for α_u. Similar results were obtained for α_v.

We also measured the average relative error resulted from estimating the coordinate of the principal point:

$$\varepsilon_{pp} = \frac{1}{N\sqrt{u_0^2 + v_0^2}}\sum_{i=1}^{N}\sqrt{(\tilde{u}_{0,i} - u_0)^2 + (\tilde{v}_{0,i} - v_0)^2}\ .$$

Here (u_0, v_0) is the true coordinate of the principal point and $(u_{0,i}, v_{0,i})$ is the estimated value from the *i*-th round of experiment.

From the experimental results, we can make the following conclusions:

- Our method outperformed the RW method for the estimation of both α_u and the principal point under the two noise conditions. The initial values of α_u and α_v are set to 1000 and 1000 respectively in the RW method while they are set to 2000 and 2000 respectively in our method. Hence, although we have selected worse initial values, our estimation results are still better than the results from the RW method. This observation is consistent with the statement in section 2.2 that our optimization method is insensitive to the initialization.
- Our performance improvement over the RW method is greater with respect to the estimation of α_u than the estimation of the principal point. This result is not unsatisfactory because as mentioned in section 3.1, the scaling factors α_u and α_v have more impact on the 3D reconstruction than the principal point does.
- The number of used images influences the estimation. Generally speaking, using more images may improve the estimation. In the case of real images, we may select well-estimated fundamental matrices for the self-calibration.

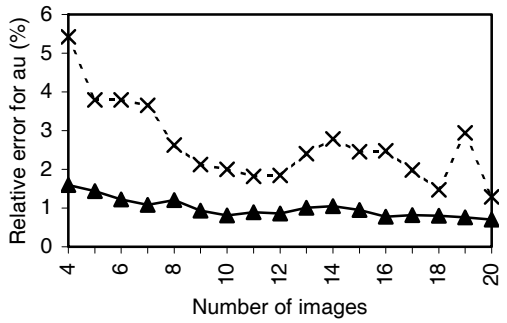

Figure 1. The comparison of the average relative error for α_u when the uniformly distributed noise [-0.5 pixel, 0.5 pixel] is added to the pixel coordinates. Our method: the solid line with triangle mark; RW method: the dotted line with cross mark.

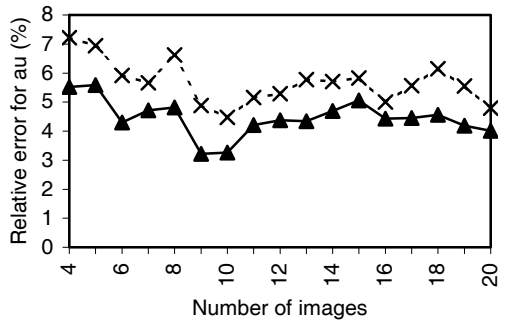

Figure 2. The comparison of the average relative error for α_u when the Gaussian noise with standard deviations of 1 pixel is added to the pixel coordinates. Our method: the solid line with triangle mark; RW method: the dotted line with cross mark.

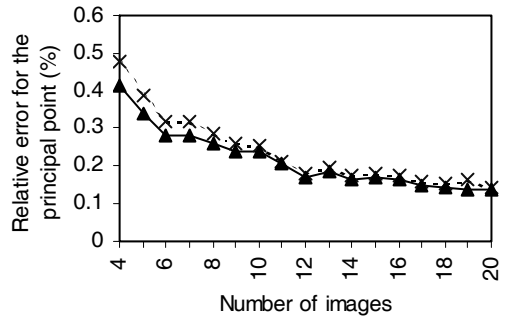

Figure 3. The comparison of the average relative error for the principal point when the uniformly distributed noise [-0.5 pixel, 0.5 pixel] is added to the pixel coordinates. Our method: the solid line with triangle mark; RW method: the dotted line with cross mark.

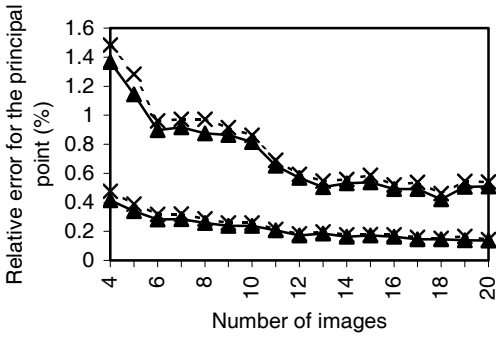

Figure 4. The comparison of the average relative error for the principal point when the Gaussian noise with standard deviation of 1 pixel is added to the pixel coordinates. Our method: the solid line with triangle mark; RW method: the dotted line with cross mark.

4.2. Experimental Results with Real Data

In this section, we show the results of self-calibration for a set of images named "Valbonne Church" of size 768×512. We downloaded these images from the INRIA ftp site.

We selected six images of this set for our experiment. The point correspondences were picked up manually. We computed all of the fifteen fundamental matrices and then selected "well-estimated" fundamental matrices in terms of the error of epipolar distance [8]. In Table 1, we compare our results with those stated in [6][7].

Table 1: Estimation results of our method and other methods on real images

	α_u	α_v	(u_0, v_0)
Kruppa	679.285	681.345	(383.188, 258.802)
RW	605.5	—	—
Prog	658.5	661.6	(406, 238)

In Table 1, the first row labeled "Kruppa" represents the estimation from [7], which is regarded as a precise estimation. The second row labeled "RW" represents the results from [6]. This method estimated only the focal length. The last row labeled "Prog" shows our results. Compared with the results of the RW method, our estimated α_u and α_v are much closer to those obtained by the Kruppa method. But our estimation of the principal point is different from that by Kruppa method. The above estimation results illustrate that our proposed approach does, at minimum, provide a very close performance to other well-established approaches for self-calibration. More importantly, the proposed method provides new advantages such as stability and simplicity due to the polynomial form of our optimization function.

5. CONCLUSIONS AND SUMMARY

In this paper, we proposed a multistage camera self-calibration algorithm based on the ESV property of the essential matrix. Unlike previous ESV-based approaches [3][5][6], we derived a polynomial optimization function, which is an explicit expression of the unknown intrinsic parameters. This makes the optimization simple and insensitive to the initialization.

We also performed a stability analysis of the intrinsic parameters and then proposed a multistage procedure to refine the self-calibration. We compared our method with the one presented in [6] on synthetic image data. The statistical results show that our method performed better than the method in [6]. We also compared our estimation results with the results from [6] and [7] on real image data. In this case, we obtained, at minimum, comparable performance to these well-established methods.

AKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Dr. Zhengyou Zhang from Microsoft Research for his valuable comments on an earlier version of this paper.

6. REFERENCES

[1] O. D. Faugeras, Q.-T. Luong and S. J. Maybank, "Camera self-calibration: theory and experiments", *Proc. 2nd European Conf. Computer Vision.*, *Lecture Notes in Computer Science*, vol. 588, pp. 321-334, 1992.

[2] Marc Pollefeys and Luc Van Gool, "Stratified self-calibration with the modulus constraint", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 707-724, 1999.

[3] Richard I. Hartley, "Estimation of relative camera positions for uncalibrated cameras", *Proc. Second European Conf. Computer Vision*, pp. 579-587, 1992.

[4] T. S. Huang and O. D. Faugeras, "Some properties of the E matrix in two-view motion estimation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1310-1312, 1989.

[5] Paulo R. S. Mendonca and Roberto Cipolla, "A simple technique for self-calibration", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 112-116, 1999.

[6] Gerhard Roth and Anthony Whitehead, "Some improvements on two autocalibration algorithms based on the fundamental matrix", *Proc. of International Conf. on Pattern Recognition*, vol. 2, pp. 312-315, 2002.

[7] Cyril Zeller and Oliver Faugeras, "Camera self-calibration from video sequences: the Kruppa equations revisited", *Research Report 2793*, INRIA, Feb. 1996.

[8] Q.-T. Luong and O.D. Faugeras, "The fundamental matrix: Theory, algorithms and stability analysis", *The International Journal of Computer Vision*, vol. 1, no. 17, pp. 43-76, 1996.

[9] Richard I. Hartley, "In defense of the eight-point algorithm", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, 1997.

[10] Zhengyou Zhang, Q.-T. Luong and O.D. Faugeras, "Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction", *Research Report 2079*, INRIA, June 1994.