Full Camera Calibration from a Single View of Planar Scene

Yisong Chen[1], Horace Ip[2], Zhangjin Huang[1], and Guoping Wang[1]

[1] Key Laboratory of Machine Perception (Ministry of Education), Peking University
[2] Department of Computer Science, City University of Hong Kong

Abstract. We present a novel algorithm that applies conics to realize reliable camera calibration. In particular, we show that a single view of two coplanar circles is sufficiently powerful to give a fully automatic calibration framework that estimates both intrinsic and extrinsic parameters. This method stems from the previous work of conic based calibration and calibration-free scene analysis. It eliminates many *a priori* constraints such as known principal point, restrictive calibration patterns, or multiple views. Calibration is achieved statistically through identifying multiple orthogonal directions and optimizing a probability function by maximum likelihood estimate. Orthogonal vanishing points, which build the basic geometric primitives used in calibration, are identified based on the fact that they represent conjugate directions with respect to an arbitrary circle under perspective transformation. Experimental results from synthetic and real scenes demonstrate the effectiveness, accuracy, and popularity of the approach.

1 Introduction

As an essential step for extracting metric 3D information from 2D images, camera calibration keeps an active research topic in most computer vision applications[9]. Much work has been devoted to camera calibration. They can be classified into two categories: (1D, 2D or 3D) Calibration pattern based algorithms, and multiple view based self-calibration approaches [15,19].

Conics and quadrics are widely accepted as most fundamental patterns in computer vision due to their elegant properties such as simple and compact algebraic expression, invariance under projective transformation, and robustness to image noise. Conics have long been employed to help perform camera calibration and pose estimation [6]. The strategy of using spheres as calibration pattern also draws more and more attention in recent years [1].

Vanishing point and vanishing line also play important roles in a lot of calibration and scene analysis work[12,14]. Under the assumption of zero skew and unit aspect ratio, all intrinsic parameters can be solved from the vanishing points of three mutually orthogonal directions in a single image[3]. Multiple patterns or views can be employed to perform calibration in the cases where not all three vanishing points are available from a single view.

Although recent research has come up with fruitful achievements, most work suffers from the problems of multiple views, restricted patterns or incompleteness of solutions [5,13,18]. Two major obstacles are the mandatory requirements of multiple views and

G. Bebis et al. (Eds.): ISVC 2008, Part I, LNCS 5358, pp. 815–824, 2008.

non-planar scene structures. In this paper, we make an attempt to calibrate the camera from a single image of planar scene. In our approach, coplanar circles are adopted as basic calibration patterns and vanishing points function in a brand new way. Conic based planar rectification[11] and conic based pose estimation[4] are two previous approaches most related to our work. The work of [4] estimates the focal length and the camera pose from two coplanar circles in the image. However, this method implicitly assumes that the principal point is known beforehand and cannot treat non-unit aspect ratio. The work of [11] makes accurate Euclidean measures from coplanar circles in a calibration-free manner. Nevertheless, the analysis is limited on the target plane and cannot be extended to applications in need of camera parameters.

In our work, we propose a full calibration scheme which statistically estimates the focal length, the principal point, the aspect ratio as well as the extrinsic parameters. In particular, we show that circle is a powerful conic in that a single view of two coplanar circles is capable of providing adequate information to do metric calibration. A coarse pipeline of our algorithm is as follows: First, calibration-free planar rectification reported in [11] is performed and extended to recover the vanishing line, the centers of the circles and many orthogonal vanishing point pairs. Second, under different guesses of the principal point the distribution of the focal length are computed from all orthogonal vanishing point pairs. Third, based on the previously computed focal length distribution a statistical optimization routine is designed to estimate the focal length, the principal point and the aspect ratio simultaneously. Fourth, the conic based pose estimation[4,6] are employed to compute the extrinsic parameters. Finally, the calibration result is validated by comparison with the ground truth for synthetic scenes or by augmented reality tests for real scenes.

The major advantage of our work lies in that the algorithm uses only a single image of simple planar scene to achieve a full solution of camera calibration. Therefore, the scene requirement is low in comparison with previous methods. This approach is very practical and works well for many scenes which previous methods fail to treat.

The rest of the paper is organized as follows. Section 2 briefly reviews and extends the previous work of coplanar circles based scene analysis. Section 3 elaborates a feasible scheme which statistically estimates the focal length, the principal point and the aspect ratio simultaneously. Some discussions are also provided in this section. Section 4 presents the experimental results on both synthetic and real scenes. Finally, concluding remarks are given in Section 5.

2 Preliminaries

We will present a calibration algorithm step by step under the practical assumption of zero skew. Our algorithm solves the camera projection matrix $P = K[R|t]$ where K is the zero-skew calibration matrix containing 4 intrinsic parameters as defined in equation (1) and the metric matrix $[R|t]$ fully encodes the 6 extrinsic parameters.

$$K = \begin{pmatrix} \alpha f & 0 & u \\ 0 & f & v \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

We first briefly introduce some related work. Throughout the discussion we adopt the homogeneous presentation which is standard in algebraic projective geometry [17].

In [11] it is suggested that under perspective transformation the images of the two circular points on a plane, $I = (1, i, 0)^T$ and $J = (1, -i, 0)^T$, can be computed by solving the intersection of the images of two coplanar circles, which have the following forms under homogeneous presentation.

$$\begin{aligned} a_1x^2 + b_1xy + c_1y^2 + d_1xw + e_1yw + f_1w^2 = 0 \\ a_2x^2 + b_2xy + c_2y^2 + d_2xw + e_2yw + f_2w^2 = 0 \end{aligned} \tag{2}$$

By solving equation (2) the images of the two circular points, I' and J', can be computed as

$$\begin{aligned} I' &= (x_0, y_0, 1) \\ J' &= (\overline{x_0}, \overline{y_0}, 1) \end{aligned} \tag{3}$$

where (x_0, y_0) are the roots of equation (2) corresponding to the circular points. Afterwards the vanishing line is computed as the cross product of the two circular points:

$$l'_\infty = I' \times J' = (y_0 - \overline{y_0}, \overline{x_0} - x_0, x_0\overline{y_0} - \overline{x_0}y_0) \tag{4}$$

Notice that the vanishing line is a real line although the entries of I' and J' are always complex.

In [4] an algorithm is addressed to estimate the focal length and the camera pose from two coplanar circles. Unfortunately the principal point has to be known a priori and the aspect ratio is fixed to be 1.0 for this algorithm to take effect. The attempt of using this constraint solely to estimate the principal point and the focal length at the same time results in large errors, especially in the presence of non-unit aspect ratio. It is also mentioned in [10] that small changes in the estimated principal point may severely degrade the quality of reconstruction. Therefore, some alternative algorithm is desired to give a reliable estimate of the principal point as well as the aspect ratio.

By integrating and extending the ideas of the above approaches, we propose a calibration scheme which simultaneously estimates the focal length, the principal point, and the aspect ratio. The algorithm is outlined in Section 3.

3 Statistical Camera Calibration

In this section, we present a step-by-step framework that benefits from conjugate direction computation and fully calibrates the camera. We start from Some preparing theories and give an orthogonal direction identification algorithm based on coplanar circles [17].

3.1 Orthogonal Vanishing Point Pairs Identification

To make full use of the geometric cues in the image we turn to the following fact: the line at the infinity, l_∞ , is the polar line of the circle center, o_i, of an arbitrary circle

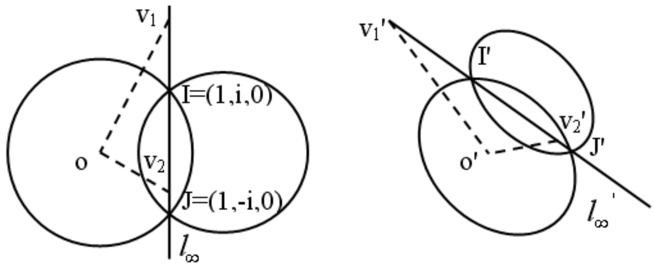

Fig. 1. Vanishing line and orthogonal point pair computation under original and perspective view. v_1 and v_2 are orthogonal vanishing points.

C on the plane. In other words, o_i and l_∞ satisfy the pole-polar relation described in equation (5).

$$l_\infty = (l_1, l_2, l_3)^T = Co_i = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{5}$$

A corollary of the above fact is that two orthogonal directions are conjugate to each other with respect to any circle on the plane. Note that a planar direction can be represented by the corresponding point at infinity. Accordingly, given a circle C and the line at infinity l_∞ on the plane, we can freely choose one point at infinity, v, on l_∞ and determine another point at infinity v' in the orthogonal direction of v using the conjugate property of the two directions. The calculation is formulated with the following equations under homogeneous representation.

$$l = Cv, v' = l \times l_\infty \tag{6}$$

That is, the orthogonal direction of a given point at infinity can be computed by solving the intersection of its polar line, l, and the line at infinity, l_∞. All the above computations are based on the pole-polar relationship, which is invariant under projective transformation. Consequently, the process can be easily transported to determine as many conjugate vanishing point pairs as we want in a perspective view. An illustration of these computations is given in Figure 1. With an image of two coplanar circles, the vanishing line can be computed using equations (2-4). Then many orthogonal directions can be computed using equation (6). This paves the way for our statistical calibration framework, which will be detailed in the next section.

3.2 Statistical Calibration by Maximum Likelihood Estimate

For convenience we first consider the camera model with unit aspect ratio. In the Cartesian image coordinate if the position of the principal point $p(x_0, y_0)$ is given, then for each freely chosen vanishing point $v(x, y)$, a corresponding vanishing point $v'(x', y')$, which represents the orthogonal direction of v, can be identified on the vanishing line using equation (6). Moreover, according to the orthogonal property of v and v', there is a unique focal length corresponding to specified p, v, and v', which can be computed from the following equation [3].

$$f = \sqrt{-(x - x_0)(x' - x_0) - (y - y_0)(y' - y_0)} \quad (7)$$

Different orthogonal vanishing point pairs lead to different estimated focal lengths. Therefore, for each guessed principal point p and a set of orthogonal vanishing points $V = \{\{v_1, v_1'\}, \{v_2, v_2'\}, ...\{v_n, v_n'\}\}$, we can estimate a corresponding set of focal lengths $F = \{f_1, f_2, ...f_n\}$. Our basic idea is to employ the set F containing large amounts of estimated f values to statistically put constraint on the principal point.

We reasonably expect that, if the principal point is correctly estimated, then the f values in the F set surely form a densely distributed cluster. On the contrary, if the guessed principal point is far from the correct position, then the distribution of the focal lengths computed by equation (7) is more likely quite sparse. Therefore, the distribution of the entries of the F set provides a confidence measure of the guess about the principal point (x_0, y_0). Naturally, the variance of the distribution, $D(F)$, is a good candidate to measure such confidence and evaluate the goodness of the guess. In other words, although the probability density function $P(x_0, y_0)$ is hidden from us it can be measured through the observable focal length distribution $D(F)$. Smaller $D(F)$ corresponds to higher confidence of (x_0, y_0). Under this formulation we can use $D(F)$ to characterize the probability density function of the principal point and perform calibration through maximum likelihood estimate. Note that F is determined by (x_0, y_0) and should be more strictly written as $F(x_0, y_0)$. We take $D(F)$ as the cost function and try to solve the following optimization problem:

$$Minimize_{(x_0,y_0)}(D(F(x_0, y_0))) \quad (8)$$

Under this formulation, from every guess about the principal point a confidence value can be estimated and the corresponding focal length can be computed. An optimization routine is required to seek the minimum of equation (8), which corresponds to the maximum likelihood estimate of the intrinsic parameters (x_0, y_0, f). In our study the above statistical function is not easily differentiated analytically. So a derivative-free optimizer is preferred. The downhill simplex method is a good candidate for this type of optimization [16]. In addition, Experiments show that a lot of local minimums exist in the solution space. We solve this problem by employing multiple initial points. Namely, the optimization is repeated several times with multiple randomly chosen starting points and the best result produced is adopted as the final solution. This strategy ensures the reliability and robustness. After the principal point (u, v) and the focal length f are determined, the conic based pose estimate algorithm in [4,8] is employed to calculate the extrinsic parameters. This completes a full single-view based calibration framework.

3.3 Taking Aspect Ratio into Account

Having made the above calibration algorithm work, adding an extra intrinsic parameter, i.e., the scale factor α, becomes straightforward. All we need to do is just introduce α as a fourth unknown variable into the optimization routine. During optimization, for each guessed value of the scale factor, the image is first scaled horizontally to give a corrected 'ideal image'. Afterwards the 'ideal image' is taken as the input of the optimization program. Due to the fact that the 4 intrinsic parameters (u, v, f, α) are somewhat tightly coupled and are not easy to be estimated separately with high precision [2], the risk of instability slightly rises compared with the algorithm considering no scale factor. Nonetheless, the power of the downhill simplex method with multiple initializations effectively prevents such degradation and a good solution can always be reached.

Table 1 briefly describes the major steps of the calibration scheme addressed in this section.

Table 1. The pipeline of our calibration scheme

step	Approach	Techniques	Equations
1	Solve the vanishing line and a group of orthogonal vanishing point pairs	Ellipse detecting and fitting, Algebraic equation solving, pole-polar computation	(2-6)
2	Compute an optimal focal length f for each given guess of (u, v, α)	Vanishing point and vanishing line based intrinsic parameter calibration	(7)
3	Use the distribution of the focal lengths to build an error function based on MLE	Probability density function based maximum-likelihood estimate	(8)
4	Nonlinear optimization by iteratively executing steps 2-4, until convergence	Downhill simplex method with multiple initialization	Ref. [16]
5	Extrinsic parameter calibration from optimal (u, v, α, f)	Conic based camera pose estimation	Ref. [4,8]

3.4 More Discussions

Finally, it is worthwhile to make an insightful comparison between our work and other conic based calibration approaches. Although conics are widely used for calibration purpose, up to date most conic based calibration algorithms are originated from the plane-based calibration framework stated in [18]. In such a framework, the images of the circular points computed from conics of multiple views are used to put constraints on and solve the image of the absolute conic ω(IAC). Then the calibration matrix K is obtained by factorizing it with the well known equation $\omega = K^{-T}K^{-1}$. In contrast, by treating the problem from a different point of view our work does not involve the notion of IAC. The calibration is achieved from statistical information provided by a large set of orthogonal vanishing points in a single view. The advantage of our work is obvious: it does not rely on *a priori* known principal point or aspect ratio as in [4], has no restrictions on circle positions as in [5,13], and above all, utilizes only a single view to achieve reliable calibration. We believe that it is an attractive solution and may represent a promising direction.

Note that in the context of our study, many orthogonal vanishing point pairs $\{v_1, v_1'\}$, $\{v_2, v_2'\}$, ..., $\{v_n, v_n'\}$ can be identified. So one might wonder whether we could use the over-determined constraint $v^T \omega v' = 0$ to solve the image of the absolute conic, ω, and then compute the calibration matrix by employing the methods of [18]. Unfortunately this idea does not work in practice. The reason is that all vanishing points come from a common line, which is a degenerate case for this formulation. The solution obtained

Fig. 2. The original Olympic logo scene and two augmented frames

Table 2. Camera calibration results for the synthetic scene

	Ground truth (u, v, α, f)	*Estimated* (u, v, α, f)
Case1	(-0.005, -0.005, 1.00, 2.56)	*(-0.00983, -0.00278, 1.00068, 2.57648)*
Case2	(0.275, -0.360, 0.89, 2.56)	*(0.26896, -0.38662, 0.89303, 2.55377)*
Case3	(0.500, 0.450, 0.97, 2.56)	*(0.49496, 0.45749, 0.96346, 2.49293)*
Case4	(-0.200, 0.125, 1.05, 2.56)	*(-0.20441, 0.17573, 1.05164, 2.68032)*
Case5	(-0.075, -0.045, 1.16, 2.56)	*(-0.07842, -0.01975, 1.15125, 2.51853)*

in this way bears a high risk of inaccuracy and instability and it often occurs that the ω computed in this manner fails to be decomposed reliably. This problem is avoided in our MLE based optimization scheme.

4 Experimental Results

Both synthetic and real world scenes are tested in our experiments. Ellipse detection and fitting is performed by some robust feature extraction and regression algorithms [7]. One point to stress is that our experiments are done in normalized image coordinates to ensure a steady order of magnitude during the solving process and to achieve a better precision [9].

4.1 Synthetic Scene

The image in our first experiment is a 512*512 synthetic planar scene with an Olympic logo on it. See Figure 2. Different circle combinations can be selected to compute the vanishing line. After the vanishing line is computed by equation (4) the method addressed in section 3 is employed to calibrate the camera and test the performance of the algorithm under different parameter configurations. In the normalized image coordinate the size of one pixel is 0.01. Some results are given in Table 2. The data in the table show that the algorithm adapts stably to a wide range of principal points and aspect ratios. All parameters can be estimated with high accuracy. The maximal deviations of the principal point, the focal length, and the aspect ratio are, respectively, 5

Fig. 3. The original and the augmented plaza scenes

Table 3. Camera calibration results for the real scene

Key circle	u	v	α	f	N
Circle1	0.00757	-0.02641	1.07045	6.86312	(-0.2804,0.7824,0.5561)
Circle2	0.01425	0.03174	1.06088	6.93255	(-0.2846,0.7869,0.5475)
Circle3	0.01577	-0.02191	1.07829	6.92373	(-0.2774,0.7857,0.5529)

pixels, 12 pixels, and 0.009. The plane normal with respect to the camera coordinate, N, can be computed using the method in [8]. In all cases we get consistent results $N \approx (-0.0005, 0.4480, 0.8940)$. This normal vector helps us compute the extrinsic parameters, which define the camera pose.

With all intrinsic and extrinsic parameters figured out, we easily rebuild the transform matrix between the world coordinate and the camera coordinate. To validate the correctness of the result we make an augmented reality experiment by producing a short movie of a 3D soccer model rolling on the plane in the scene. The augmented scene looks quite realistic. Two frames extracted from the movie are illustrated in Figure 2.

4.2 Real Scene

In our second experiment a challenging 640*480 plaza scene is adopted to test the interesting case of concentric circles. See Figure 3. Despite the relatively noisy scene structures in the image, the distinct magenta color of the two rings in the scene allows robust detection of both edges of the inner ring (circle1 and circle2) and the inner edge of the outer ring (circle3). Circle1 and circle2 are too close to each other and represent a typical degenerate case. So this pair is not appropriate and should be discarded for calibration purpose. By contrast, the (circle1, circle3) pair and the (circle2, circle3) pair are sufficiently distant although the two circles are concentric. From either combination the images of the circular points and the circle centers can be successfully estimated. By employing equation (6) we compute an individual orthogonal vanishing point set from each of the 3 circles and try one calibration respectively. The calibration results of the 4 intrinsic parameters and the plane normal N for all 3 cases are given in Table 3. The data in the table show that the results are quite consistent. This strongly justifies the stability of our approach.

Two reference points are selected on the plane to help build a world coordinate system with the ground plane as the x-y plane and its normal as the z direction. Then the transform between the camera coordinate and the world coordinate is established. We additionally validate our algorithm by adding a 3D moving car model into the scene. Metric rectification technique helps tailor the size of the virtual car and fill it seamlessly into the scene. Two frames of the animation are given in Figure 3.

5 Conclusion

We have presented a framework that fully calibrates the camera from only a single perspective view of two coplanar circles. Metric planar rectification and conic based pose estimate are combined in a statistical manner to achieve robust and reliable calibration. The advantage of our work is twofold: First, it is superior to many previous calibration algorithms in that it uses only a single view of planar scene. Second, it is more practical than many calibration-free approaches because it supports 3D augmented reality applications in addition to simple 2D Euclidean measure. Future work includes making more accurate error analysis and treating more complex camera models.

Acknowledgement

We would like to thank all anonymous reviewers for their comments and advices, which greatly help improve the quality of this manuscript.

References

1. Agrawal, M., Davis, L.S.: Camera calibration using spheres: a semi-definite programming approach. In: Proc. of IEEE ICCV, pp. 782–789 (2003)
2. Bougnoux, S.: From projective to Euclidean space under any practical situation, a criticism of selfcalibration. In: Proc. 6th ICCV, pp. 790–796 (January 1998)
3. Caprile, B., Torre, V.: Using vanishing points for camera calibration. International journal of computer vision 4(2), 127–140 (1990)
4. Chen, Q., Hu, H., Wada, T.: Camera calibration with two arbitrary coplanar circles. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 521–532. Springer, Heidelberg (2004)
5. Colombo, C., Comanducci, D., Del Bimbo, A.: Camera calibration with two arbitrary coaxial circles. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 265–276. Springer, Heidelberg (2006)
6. Dhome, M., Lapreste, J., Rives, G., Richetin, M.: Spatial localization of modeled objects of revolution in monocular perspective vision. In: Faugeras, O. (ed.) ECCV 1990. LNCS, vol. 427, pp. 475–485. Springer, Heidelberg (1990)
7. Fitzgibbon, A., Pilu, M., Fisher, R.: Direct least-square fitting of ellipses. IEEE Trans. PAMI (June 1999)
8. Forsyth, D., et al.: Invariant descriptors for 3-D object recognition and pose. IEEE Trans. PAMI 13(10), 250–262 (1991)
9. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University, Cambridge (2003)
10. Hartley, R., Kaucic, R.: Sensitivity of calibration to principal point position. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 433–446. Springer, Heidelberg (2002)

11. Ip, H., Chen, Y.: Planar rectification by solving the intersection of two circles under 2D homography. Pattern recognition 38(7), 1117–1120 (2005)
12. Kanatani, K.: Statistical analysis of focal-length calibration using vanishing points. IEEE Trans. Robotics and Automation 8(6), 767–775 (1992)
13. Kim, J., Gurdjos, P., Kweon, I.: Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. IEEE Trans. PAMI 27(4), 637–642 (2005)
14. Liebowitz, D., Criminisi, A., Zisserman, A.: Creating architectural models from images. In: EuroGraphics1999, vol. 18, pp. 39–50 (1999)
15. Luong, Q.T., Faugeras, O.: Self-calibration of a moving camera from point correspondences and fundamental matrices. IJCV 22(3), 261–289 (1997)
16. Press, W.H., et al.: Numerical recipes in C: the art of scientific computing, 2nd edn. Cambridge University Press, Cambridge (1997)
17. Semple, J.G., Kneebone, G.T.: Algebraic projective Geometry. Clarendon Press, Oxford (1998)
18. Sturm, P.F., Maybank, S.J.: On plane-based camera calibration: a general algorithm, singularities, applications. In: Proc. CVPR, pp. 432–437 (1999)
19. Zhang, Z.: Camera calibration with one-dimensional objects. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 161–174. Springer, Heidelberg (2002)