A Flexible New Technique for Camera Calibration

Zhengyou Zhang, *Senior Member*, *IEEE*

Abstract—We propose a flexible new technique to easily calibrate a camera. It only requires the camera to observe a planar pattern shown at a few (at least two) different orientations. Either the camera or the planar pattern can be freely moved. The motion need not be known. Radial lens distortion is modeled. The proposed procedure consists of a closed-form solution, followed by a nonlinear refinement based on the maximum likelihood criterion. Both computer simulation and real data have been used to test the proposed technique and very good results have been obtained. Compared with classical techniques which use expensive equipment such as two or three orthogonal planes, the proposed technique is easy to use and flexible. It advances 3D computer vision one more step from laboratory environments to real world use. The corresponding software is available from the author's Web page.

Index Terms—Camera calibration, calibration from planes, 2D pattern, flexible plane-based calibration, absolute conic, projective mapping, lens distortion, closed-form solution, maximum likelihood estimation, flexible setup.

1 Motivations

CAMERA calibration is a necessary step in 3D computer vision in order to extract metric information from 2D images. Much work has been done, starting in the photogrammetry community (see [2], [4] to cite a few), and more recently in computer vision ([9], [8], [23], [7], [25], [24], [16], [6] to cite a few). We can classify those techniques roughly into two categories: photogrammetric calibration and self-calibration.

- **Three-dimensional reference object-based calibration.** Camera calibration is performed by observing a calibration object whose geometry in 3D space is known with very good precision. Calibration can be done very efficiently [5]. The calibration object usually consists of two or three planes orthogonal to each other. Sometimes a plane undergoing a precisely known translation is also used [23]. These approaches require an expensive calibration apparatus, and an elaborate setup.
- **Self-calibration.** Techniques in this category do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides in general two constraints [16], [14] on the cameras' internal parameters from one camera displacement by using image information alone. Therefore, if images are taken by the same camera with fixed internal parameters, correspondences between three images are sufficient to recover both the internal and external parameters which allow us to reconstruct 3D structure up to a similarity [15], [12]. While this approach is very flexible, it is not yet mature [1]. Because there are many parameters to estimate, we cannot always obtain reliable results.

Other techniques exist: vanishing points for orthogonal directions [3], [13], and calibration from pure rotation [11], [20].

- *The author is with Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399. E-mail: zhang@microsoft.com.*

Manuscript received 26 Apr. 2000; revised 25 Aug. 2000; accepted 7 Sept. 2000.
Recommended for acceptance by A. Shashua.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111998.

Our current research is focused on a desktop vision system (DVS) since the potential for using DVSs is large. Cameras are becoming inexpensive and ubiquitous. A DVS aims at the general public who are not experts in computer vision. A typical computer user will perform vision tasks only from time to time, so they will not be willing to invest money for expensive equipment. Therefore, flexibility, robustness, and low cost are important. The camera calibration technique described in this paper was developed with these considerations in mind.

The proposed technique only requires the camera to observe a planar pattern shown at a few (at least two) different orientations. The pattern can be printed on a laser printer and attached to a "reasonable" planar surface (e.g., a hard book cover). Either the camera or the planar pattern can be moved by hand. The motion need not be known. The proposed approach, which uses 2D metric information, lies between the photogrammetric calibration, which uses explicit 3D model, and self-calibration, which uses motion rigidity or equivalently implicit 3D information. Both computer simulation and real data have been used to test the proposed technique and very good results have been obtained. Compared with classical techniques, the proposed technique is considerably more flexible: Anyone can make a calibration pattern by him/herself and the setup is very easy. Compared with self-calibration, it gains a considerable degree of robustness. We believe the new technique advances 3D computer vision one step from laboratory environments to the real world.

Note that Triggs [22] recently developed a self-calibration technique from at least five views of a planar scene. His technique is more flexible than ours, but has difficulty to initialize. Liebowitz and Zisserman [13] described a technique of metric rectification for perspective images of planes using metric information, such as a known angle, two equal though unknown angles, and a known length ratio. They also mentioned that calibration of the internal camera parameters is possible provided at least three such rectified planes, although no experimental results were shown.

During the revision of this paper, we notice the publication of an independent but similar work by Sturm and Maybank [21].They use a simplified camera model (image axes are orthogonal to each other) and have studied the degenerate configurations exhaustively for the case of one and two planes, which are very important in practice if only one or two views are used for camera calibration.

The paper is organized as follows: Section 2 describes the basic constraints from observing a single plane. Section 3 describes the calibration procedure. We start with a closed-form solution, followed by nonlinear optimization. Radial lens distortion is also modeled. Section 4 provides the experimental results. Both computer simulation and real data are used to validate the proposed technique. In the Appendix, we provides a number of details, including the techniques for estimating the homography between the model plane and its image.

2 Basic Equations

We examine the constraints on the camera's intrinsic parameters provided by observing a single plane. We start with the notation used in this paper.

2.1 Notation

A 2D point is denoted by $\mathbf{m} = [u, v]^T$. A 3D point is denoted by $\mathbf{M} = [X, Y, Z]^T$. We use $\tilde{\mathbf{x}}$ to denote the augmented vector by adding 1 as the last element: $\tilde{\mathbf{m}} = [u, v, 1]^T$ and $\widetilde{\mathbf{M}} = [X, Y, Z, 1]^T$. A camera is modeled by the usual pinhole: The relationship between a 3D point $\mathbf{M}$ and its image projection $\mathbf{m}$ is given by

$$s\tilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\widetilde{\mathbf{M}}, \qquad \text{with } \mathbf{A} = \begin{bmatrix}\alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1\end{bmatrix}, \tag{1}$$

where s is an arbitrary scale factor, $(\mathbf{R}, \mathbf{t})$, called the extrinsic parameters is the rotation and translation which relates the world coordinate system to the camera coordinate system, and $\mathbf{A}$ is called the camera intrinsic matrix, with (u_0, v_0) the coordinates of the principal point, α and β the scale factors in image u and v axes, and γ the parameter describing the skew of the two image axes.

We use the abbreviation $\mathbf{A}^{-T}$ for $(\mathbf{A}^{-1})^T$ or $(\mathbf{A}^T)^{-1}$.

2.2 Homography between the Model Plane and Its Image

Without loss of generality, we assume the model plane is on $Z = 0$ of the world coordinate system. Let's denote the ith column of the rotation matrix $\mathbf{R}$ by $\mathbf{r}_i$. From (1), we have

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}.$$

By abuse of notation, we still use $\mathbf{M}$ to denote a point on the model plane, but $\mathbf{M} = [X, Y]^T$ since Z is always equal to zero. In turn, $\widetilde{\mathbf{M}} = [X, Y, 1]^T$. Therefore, a model point $\mathbf{M}$ and its image $\mathbf{m}$ is related by a homography $\mathbf{H}$:

$$s\widetilde{\mathbf{m}} = \mathbf{H}\widetilde{\mathbf{M}} \quad \text{with} \quad \mathbf{H} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}. \tag{2}$$

As is clear, the 3×3 matrix $\mathbf{H}$ is defined up to a scale factor.

2.3 Constraints on the Intrinsic Parameters

Given an image of the model plane, an homography can be estimated (see the Appendix). Let's denote it by $\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$. From (2), we have

$$[\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = \lambda\mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}],$$

where λ is an arbitrary scalar. Using the knowledge that $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, we have

$$\mathbf{h}_1^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h}_2 = 0 \tag{3}$$

$$\mathbf{h}_1^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h}_2. \tag{4}$$

These are the two basic constraints on the intrinsic parameters, given one homography. Because a homography has 8 degrees of freedom and there are six extrinsic parameters (three for rotation and three for translation), we can only obtain two constraints on the intrinsic parameters. Note that $\mathbf{A}^{-T}\mathbf{A}^{-1}$ actually describes the image of the absolute conic [15]. In the next section, we will give a geometric interpretation.

2.4 Geometric Interpretation

We are now relating (3) and (4) to the absolute conic [16], [15].

It is not difficult to verify that the model plane, under our convention, is described in the camera coordinate system by the following equation:

$$\begin{bmatrix} \mathbf{r}_3 \\ \mathbf{r}_3^T\mathbf{t} \end{bmatrix}^T\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = 0,$$

where $w = 0$ for points at infinity and $w = 1$ otherwise. This plane intersects the plane at infinity at a line and we can easily see that

$$\begin{bmatrix} \mathbf{r}_1 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbf{r}_2 \\ 0 \end{bmatrix}$$

are two particular points on that line. Any point on it is a linear combination of these two points, i.e.,

$$\mathbf{x}_\infty = a\begin{bmatrix} \mathbf{r}_1 \\ 0 \end{bmatrix} + b\begin{bmatrix} \mathbf{r}_2 \\ 0 \end{bmatrix} = \begin{bmatrix} a\mathbf{r}_1 + b\mathbf{r}_2 \\ 0 \end{bmatrix}.$$

Now, let's compute the intersection of the above line with the absolute conic. By definition, the point $\mathbf{x}_\infty$, known as the *circular point* [18], satisfies: $\mathbf{x}_\infty^T\mathbf{x}_\infty = 0$, i.e., $(a\mathbf{r}_1 + b\mathbf{r}_2)^T(a\mathbf{r}_1 + b\mathbf{r}_2) = 0$, or $a^2 + b^2 = 0$. The solution is $b = \pm ai$, where $i^2 = -1$. That is, the two intersection points are

$$\mathbf{x}_\infty = a\begin{bmatrix} \mathbf{r}_1 \pm i\mathbf{r}_2 \\ 0 \end{bmatrix}.$$

The significance of this pair of complex conjugate points lies in the fact that they are invariant to Euclidean transformations. Their projection in the image plane is given, up to a scale factor, by

$$\widetilde{\mathbf{m}}_\infty = \mathbf{A}(\mathbf{r}_1 \pm i\mathbf{r}_2) = \mathbf{h}_1 \pm i\mathbf{h}_2.$$

Point $\widetilde{\mathbf{m}}_\infty$ is on the image of the absolute conic, described by $\mathbf{A}^{-T}\mathbf{A}^{-1}$ [15]. This gives

$$(\mathbf{h}_1 \pm i\mathbf{h}_2)^T\mathbf{A}^{-T}\mathbf{A}^{-1}(\mathbf{h}_1 \pm i\mathbf{h}_2) = 0.$$

Requiring that both real and imaginary parts be zero yields (3) and (4).

3 Solving Camera Calibration

This section provides the details how to effectively solve the camera calibration problem. We start with an analytical solution, followed by a nonlinear optimization technique based on the maximum-likelihood criterion. Finally, we take into account lens distortion, giving both analytical and nonlinear solutions.

3.1 Closed-Form Solution

Let

$$\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1} \equiv \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2\beta} & \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ -\frac{\gamma}{\alpha^2\beta} & \frac{\gamma^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix}. \tag{5}$$

Note that $\mathbf{B}$ is symmetric, defined by a 6D vector

$$\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T. \tag{6}$$

Let the ith column vector of $\mathbf{H}$ be $\mathbf{h}_i = [h_{i1}, h_{i2}, h_{i3}]^T$. Then, we have

$$\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j = \mathbf{v}_{ij}^T\mathbf{b} \tag{7}$$

with

$$\mathbf{v}_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T.$$

Therefore, the two fundamental constraints (3) and (4), from a given homography, can be rewritten as two homogeneous equations in $\mathbf{b}$:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix}\mathbf{b} = \mathbf{0}. \tag{8}$$

If n images of the model plane are observed, by stacking n such equations as (8), we have

$$\mathbf{V}\mathbf{b} = \mathbf{0}, \tag{9}$$

where $\mathbf{V}$ is a $2n \times 6$ matrix. If $n \geq 3$, we will have in general a unique solution $\mathbf{b}$ defined up to a scale factor. If $n = 2$, we can impose the skewless constraint $\gamma = 0$, i.e., $[0, 1, 0, 0, 0, 0]\mathbf{b} = 0$, which is added as an additional equation to (9). (If $n = 1$, we can only solve two camera intrinsic parameters, e.g., α and β, assuming u_0 and v_0 are known (e.g., at the image center) and $\gamma = 0$, and that is indeed what we did in [19] for head pose determination based on the fact that eyes and mouth are reasonably coplanar. In fact, Tsai [23] already mentions that focal length from one plane is possible, but incorrectly says that aspect ratio is not.) The solution to (9) is well-known as the eigenvector of $\mathbf{V}^T\mathbf{V}$ associated with the smallest eigenvalue (equivalently, the right singular vector of $\mathbf{V}$ associated with the smallest singular value).

Once $\mathbf{b}$ is estimated, we can compute all camera intrinsic parameters as follows. The matrix $\mathbf{B}$, as described in Section 3.1, is estimated up to a scale factor, i.e., $\mathbf{B} = \lambda\mathbf{A}^{-T}\mathbf{A}$ with λ an arbitrary scale. Without difficulty, we can uniquely extract the intrinsic parameters from matrix $\mathbf{B}$.

$$\begin{aligned}
v_0 &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2)\\
\lambda &= B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})]/B_{11}\\
\alpha &= \sqrt{\lambda/B_{11}}\\
\beta &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)}\\
\gamma &= -B_{12}\alpha^2\beta/\lambda\\
u_0 &= \gamma v_0/\alpha - B_{13}\alpha^2/\lambda.
\end{aligned}$$

Once $\mathbf{A}$ is known, the extrinsic parameters for each image is readily computed. From (2), we have

$$\mathbf{r}_1 = \lambda\mathbf{A}^{-1}\mathbf{h}_1, \quad \mathbf{r}_2 = \lambda\mathbf{A}^{-1}\mathbf{h}_2, \quad \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \quad \mathbf{t} = \lambda\mathbf{A}^{-1}\mathbf{h}_3$$

with $\lambda = 1/\|\mathbf{A}^{-1}\mathbf{h}_1\| = 1/\|\mathbf{A}^{-1}\mathbf{h}_2\|$. Of course, because of noise in data, the so-computed matrix $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ does not, in general, satisfy the properties of a rotation matrix. The best rotation matrix can then be obtained through for example singular value decomposition [10], [26].

3.2 Maximum-Likelihood Estimation

The above solution is obtained through minimizing an algebraic distance which is not physically meaningful. We can refine it through maximum-likelihood inference.

We are given n images of a model plane and there are m points on the model plane. Assume that the image points are corrupted by independent and identically distributed noise. The maximum-likelihood estimate can be obtained by minimizing the following functional:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)\|^2, \tag{10}$$

where $\hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)$ is the projection of point $\mathbf{M}_j$ in image i, according to (2). A rotation $\mathbf{R}$ is parameterized by a vector of three parameters, denoted by $\mathbf{r}$, which is parallel to the rotation axis and whose magnitude is equal to the rotation angle. $\mathbf{R}$ and $\mathbf{r}$ are related by the Rodrigues formula [5]. Minimizing (10) is a nonlinear minimization problem, which is solved with the Levenberg-Marquardt Algorithm as implemented in `Minpack` [17]. It requires an initial guess of $\mathbf{A}, \{\mathbf{R}_i, \mathbf{t}_i | i = 1..n\}$ which can be obtained using the technique described in the previous section.

Desktop cameras usually have visible lens distortion, especially the radial components. We have included these while minimizing (10). Refer to the technical report, [26], for more details.

3.3 Summary

The recommended calibration procedure is as follows:

1. Print a pattern and attach it to a planar surface.
2. Take a few images of the model plane under different orientations by moving either the plane or the camera.
3. Detect the feature points in the images.
4. Estimate the five intrinsic parameters and all the extrinsic parameters using the closed-form solution, as described in Section 3.1
5. Refine all parameters, including lens distortion parameters, by minimizing (10).

There is a degenerate configuration in my technique when planes are parallel to each other. Refer to the technical report, [26], for a more detailed description.

4 Experimental Results

The proposed algorithm has been tested on both computer simulated data and real data. The closed-form solution involves finding a singular value decomposition of a small $2n \times 6$ matrix, where n is the number of images. The nonlinear refinement within the Levenberg-Marquardt Algorithm takes 3 to 5 iterations to converge. Due to space limitation, we describe in this section one set of experiments with real data when the calibration pattern is at different distances from the camera. The reader is referred to [26] for more experimental results with both computer simulated and real data, and to the following Web page: http://research.microsoft.com/~zhang/Calib/ for some experimental data and the software.

The example is shown in Fig. 1. The camera to be calibrated is an off-the-shelf PULNiX CCD camera with 6 mm lens. The image resolution is 640×480. As can be seen in Fig. 1, the model plane contains 9×9 squares with nine special dots which are used to identify automatically the correspondence between reference points on the model plane and square corners in images. It was printed on a A4 paper with a 600 DPI laser printer and attached to a cardboard.

In total, 10 images of the plane were taken (six of them are shown in Fig. 1). Five of them (called Set A) were taken at close range, while the other five (called Set B) were taken at a larger distance. We applied our calibration algorithm to Set A, Set B, and also to the whole set (called Set A+B). The results are shown in Table 1. For intuitive understanding, we show the estimated angle between the image axes, ϑ, instead of the skew factor γ. We can see that the angle ϑ is very close to 90°, as expected with almost all modern CCD cameras. The cameras parameters were estimated consistently for all three sets of images, except the distortion parameters with Set B. The reason is that the calibration pattern only occupies the central part of the image in Set B, where lens distortion is not significant and therefore cannot be estimated reliably.

5 Conclusion

In this paper, we have developed a flexible new technique to easily calibrate a camera. The technique only requires the camera to observe a planar pattern from a few different orientations. Although the minimum number of orientations is two if pixels are square, we recommend four or five different orientations for better quality. We can move either the camera or the planar

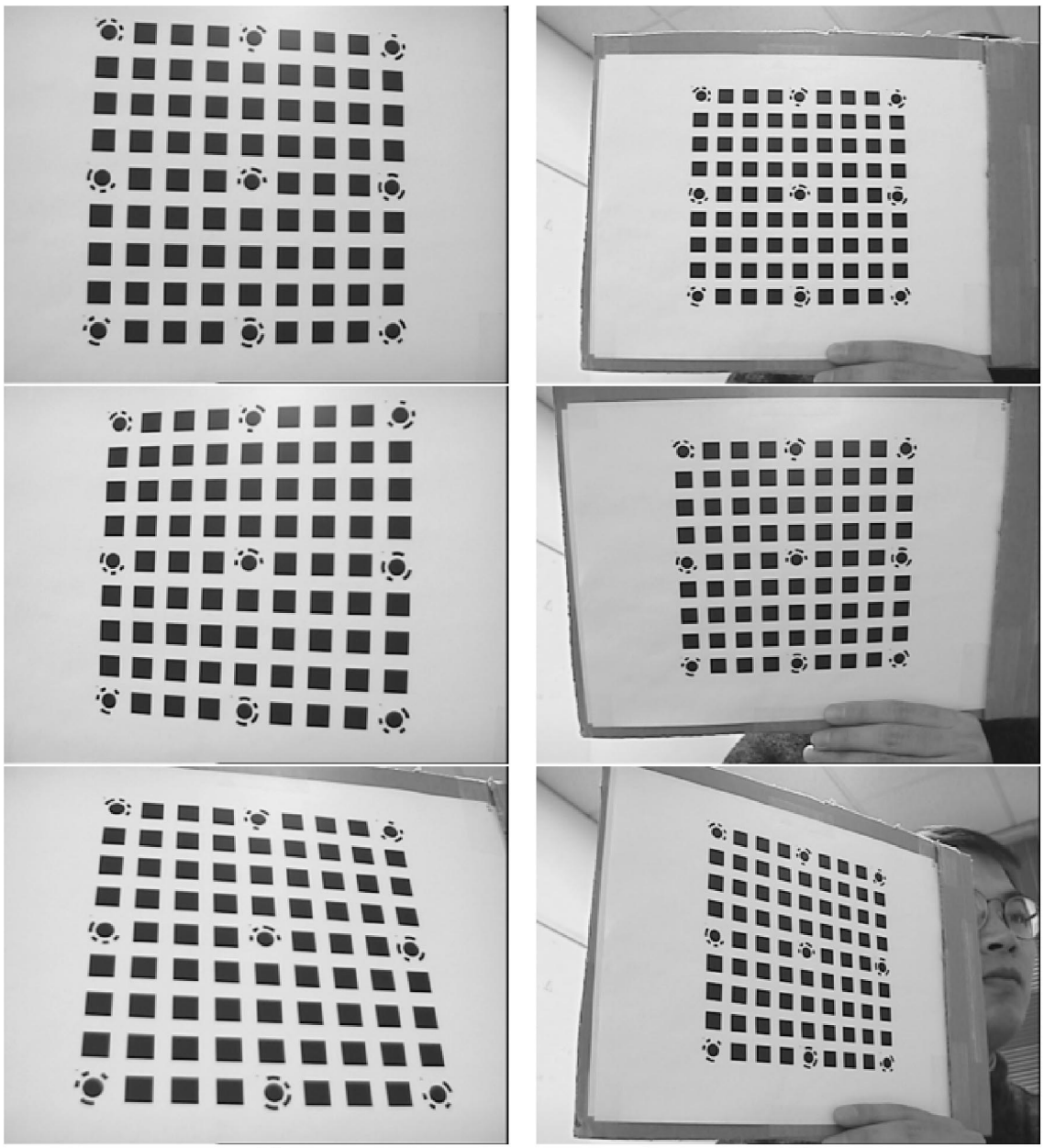

Fig. 1. Two sets of images taken at different distances to the calibration pattern. Each set contains five images. (a) Three images from the set taken at a close distance are shown. (b) Three images from the set taken at a larger distance are shown.

pattern. The motion does not need to be known, but should not be a pure translation. When the number of orientations is only two, one should avoid positioning the planar pattern parallel to the image plane. The pattern could be anything, as long as we know the metric on the plane. For example, we can print a pattern with a laser printer and attach the paper to a reasonable planar surface such as a hard book cover. We can even use a book with known size because the four corners are enough to estimate the plane homographies.

Radial lens distortion is modeled. The proposed procedure consists of a closed-form solution, followed by a nonlinear refinement based on a maximum-likelihood criterion. Both computer simulation and real data have been used to test the proposed technique and very good results have been obtained. Compared with classical techniques which use expensive equipment such as two or three orthogonal planes, the proposed technique gains considerable flexibility.

APPENDIX

ESTIMATING HOMOGRAPHY BETWEEN THE MODEL PLANE AND ITS IMAGE

There are many ways to estimate the homography between the model plane and its image. Here, we present a technique based on a maximum-likelihood criterion. Let $\mathbf{M}_i$ and $\mathbf{m}_i$ be the model and image points, respectively. Ideally, they should satisfy (2). In practice, they don't because of noise in the extracted image points. Let's assume that $\mathbf{m}_i$ is corrupted by Gaussian noise with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Lambda}_{\mathbf{m}_i}$. Then, the maximum-likelihood estimation of $\mathbf{H}$ is obtained by minimizing the following functional

TABLE 1
Calibration Results with the Images Shown in Fig. 1

image set	α	β	ϑ	u_0	v_0	k_1	k_2
A	834.01	839.86	89.95°	305.51	240.09	-0.2235	0.3761
B	836.17	841.08	89.92°	301.76	241.51	-0.2676	1.3121
A+B	834.64	840.32	89.94°	304.77	240.59	-0.2214	0.3643

$$\sum_i (\mathbf{m}_i - \hat{\mathbf{m}}_i)^T \mathbf{\Lambda}_{\mathbf{m}_i}^{-1} (\mathbf{m}_i - \hat{\mathbf{m}}_i),$$

where

$$\hat{\mathbf{m}}_i = \frac{1}{\bar{\mathbf{h}}_3^T \mathbf{M}_i} \begin{bmatrix} \bar{\mathbf{h}}_1^T \mathbf{M}_i \\ \bar{\mathbf{h}}_2^T \mathbf{M}_i \end{bmatrix} \quad \text{with } \bar{\mathbf{h}}_i, \text{the } i\text{th row of } \mathbf{H}.$$

In practice, we simply assume $\mathbf{\Lambda}_{\mathbf{m}_i} = \sigma^2 \mathbf{I}$ for all i. This is reasonable if points are extracted independently with the same procedure. In this case, the above problem becomes a nonlinear least-squares one, i.e., $\min_{\mathbf{H}} \sum_i \|\mathbf{m}_i - \hat{\mathbf{m}}_i\|^2$. The nonlinear minimization is conducted with the Levenberg-Marquardt Algorithm as implemented in `Minpack` [17]. This requires an initial guess, which can be obtained as follows:

Let $\mathbf{x} = [\bar{\mathbf{h}}_1^T, \bar{\mathbf{h}}_2^T, \bar{\mathbf{h}}_3^T]^T$. Then, (2) can be rewritten as

$$\begin{bmatrix} \widetilde{\mathbf{M}}^T & \mathbf{0}^T & -u\widetilde{\mathbf{M}}^T \\ \mathbf{0}^T & \widetilde{\mathbf{M}}^T & -v\widetilde{\mathbf{M}}^T \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

When we are given n points, we have n above equations, which can be written in matrix equation as $\mathbf{L}\mathbf{x} = \mathbf{0}$, where $\mathbf{L}$ is a $2n \times 9$ matrix. As $\mathbf{x}$ is defined up to a scale factor, the solution is well-known to be the right singular vector of $\mathbf{L}$ associated with the smallest singular value (or equivalently, the eigenvector of $\mathbf{L}^T\mathbf{L}$ associated with the smallest eigenvalue). In $\mathbf{L}$, some elements are constant 1, some are in pixels, some are in world coordinates, and some are multiplication of both. This makes $\mathbf{L}$ poorly conditioned numerically. Much better results can be obtained by performing a simple data normalization prior to running the above procedure.

ACKNOWLEDGMENTS

The author would like to thank Brian Guenter for his software on corner extraction and for many discussions and to Bill Triggs for insightful comments. He would also like to thank Andrew Zisserman for bringing his CVPR '98 work [13] to the authors' attention. It uses the same constraint but in different form. He would also like to thank the members of the Vision Group at Miscrosoft Research for encouragement and discussions. Anandan and Charles Loop have checked the English of an early version. The constructive comments from the anonymous reviewers are gratefully acknowledged which have helped the author to improve the paper.

REFERENCES

[1] S. Bougnoux, "From Projective to Euclidean Space under any Practical Situation, a Criticism of Self-Calibration," *Proc. Sixth Int'l Conf. Computer Vision*, pp. 790-796, Jan. 1998.
[2] D.C. Brown, "Close-Range Camera Calibration," *Photogrammetric Eng.*, vol. 37, no. 8, pp. 855-866, 1971.
[3] B. Caprile and V. Torre, "Using Vanishing Points for Camera Calibration," *Int'l J. Computer Vision*, vol. 4, no. 2, pp. 127-140, Mar. 1990.
[4] W. Faig, "Calibration of Close-Range Photogrammetry Systems: Mathematical Formulation," *Photogrammetric Eng. and Remote Sensing*, vol. 41, no. 12, pp. 1,479-1,486, 1975.
[5] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.
[6] O. Faugeras, T. Luong, and S. Maybank, "Camera Self-Calibration: Theory and Experiments," *Proc Second European Conf. Computer Vision*, pp. 321-334, May 1992.
[7] O. Faugeras and G. Toscani, "The Calibration Problem for Stereo," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 15-20, June 1986.
[8] S. Ganapathy, "Decomposition of Transformation Matrices for Robot Vision," *Pattern Recognition Letters*, vol. 2, pp. 401-412, Dec. 1984.
[9] D. Gennery, "Stereo-Camera Calibration," *Proc. 10th Image Understanding Workshop*, pp. 101-108, 1979.
[10] G. Golub and C. van Loan, *Matrix Computations*, Baltimore: The John Hopkins Univ. Press, third ed. 1996.
[11] R. Hartley, "Self-Calibration from Multiple Views with a Rotating Camera," *Proc. Third European Conf. Computer Vision*, pp. 471-478, May 1994.
[12] R.I. Hartley, "An Algorithm for Self-Calibration from Several Views," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 908-912, June 1994.
[13] D. Liebowitz and A. Zisserman, "Metric Rectification for Perspective Images of Planes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 482-488, June 1998.
[14] Q.-T. Luong, "Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus Grande Autonomie des Systèmes Robotiques," PhD thesis, Université de Paris-Sud, Centre d'Orsay, Dec. 1992.
[15] Q.-T. Luong and O. Faugeras, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices," *Int'l J. Computer Vision*, vol. 22, no. 3, pp. 261-289, 1997.
[16] S.J. Maybank and O.D. Faugeras, "A Theory of Self-Calibration of a Moving Camera," *Int'l J. Computer Vision*, vol. 8. no. 2, pp. 123-152, Aug. 1992.
[17] J. More, "The Levenberg-Marquardt Algorithm, Implementation, and Theory," *Numerical Analysis*, G.A. Watson, ed., Springer-Verlag, 1977.
[18] J. Semple and G. Kneebone, *Algebraic Projective Geometry*. Oxford: Clarendon Press, 1952.
[19] I. Shimizu, Z. Zhang, S. Akamatsu, and K. Deguchi, "Head Pose Determination from One Image Using a Generic Model," *Proc. IEEE Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 100-105, Apr. 1998.
[20] G. Stein, "Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Error," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 230-236, June 1995.
[21] P. Sturm and S. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 432-437, June 1999.
[22] B. Triggs, "Autocalibration from Planar Scenes," *Proc. Fifth European Conf. Computer Vision*, pp. 89-105, June 1998.
[23] R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE J. Robotics and Automation*, vol. 3, no. 4, pp. 323-344, Aug. 1987.
[24] G. Wei and S. Ma, "A Complete Two-Plane Camera Calibration Method and Experimental Comparisons," *Proc. Fourth Int'l Conf. Computer Vision*, pp. 439-446, May 1993.
[25] J. Weng, P. Cohen, and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965-980, Oct. 1992.
[26] Z. Zhang, "A Flexible New Technique for Camera Calibration," Technical Report MSR-TR-98-71, Microsoft Research, Dec. 1998. Available together with the software at http://research.microsoft.com/~zhang/Calib/.