Camera Calibration with Distortion Models and Accuracy Evaluation

Juyang Weng, *Member, IEEE*, Paul Cohen, and Marc Herniou

Abstract— **The objective of stereo camera calibration is to estimate the internal and external parameters of each camera. Using these parameters, the 3-D position of a point in the scene, which is identified and matched in two stereo images, can be determined by the method of triangulation. In this paper, we present a camera model that accounts for major sources of camera distortion, namely, radial, decentering, and thin prism distortions. The proposed calibration procedure consists of two steps. In the first step, the calibration parameters are estimated using a closed-form solution based on a distortion-free camera model. In the second step, the parameters estimated in the first step are improved iteratively through a nonlinear optimization, taking into account camera distortions. According to minimum variance estimation, the objective function to be minimized is the mean-square discrepancy between the observed image points and their inferred image projections computed with the estimated calibration parameters. We introduce a type of measure that can be used to directly evaluate the performance of calibration and compare calibrations among different systems. The validity and performance of our calibration procedure are tested with both synthetic data and real images taken by tele- and wide-angle lenses. The results consistently show significant improvements over less complete camera models.**

Index Terms— **Camera calibration, lens distortion, optimization, stereo.**

I. Introduction

CALIBRATION OF cameras is considered as an important issue in computer vision. Accurate calibration of cameras is especially crucial for applications that involve quantitative measurements such as dimensional measurements, depth from stereoscopy, or motion from images.

One aspect of camera calibration is to estimate the internal parameters of the camera. These parameters determine how the image coordinates of a point are derived, given the spatial position of the point with respect to the camera. The estimation of the geometrical relation between the camera and the scene, or between different cameras, is also an important aspect of calibration. The corresponding parameters that characterize such a geometrical relation are called external parameters. It is well known that actual cameras are not perfect and sustain a variety of aberrations. For geometrical measurements, the main concern is camera distortion, which relates to the position of image points in the image plane but not directly to the image quality. For example, the position of a point in a slightly blurred image can still be measured as the center of the blurred point. However, if the image position of a point is not accurate, the results that depend on its image coordinates will be erroneous.

Camera calibration has long been an important issue in the photogrammetry community. With the increasing need for higher accuracy measurement in computer vision, it has also attracted research efforts in the computer vision community. Compared with the high-quality metric cameras used in photogrammetry, the cameras commonly used in computer vision have the following characteristics: a) Image spatial resolution is defined by spatial digitization and is relatively low (e.g., a typical CCD sensing array has about 512×480 pixels); b) lenses used for video cameras are nonmetric off-the-shelf lenses and sustain a substantial amount of distortion; c) camera assembly sustains considerable internal misalignment (e.g., the CCD sensing array may not be orthogonal to the optical axis, and the center of the array may not coincide with the optical principal point, i.e., the intersection of the optical axis and the image plane).

The existing techniques for camera calibration can be classified into the following categories.

1) Direct Nonlinear Minimization: In this category, equations that relate the parameters to be estimated with the 3-D coordinates of control points and their image plane projections are established. The search for the parameters involves using an iterative algorithm with the objective of minimizing residual errors of some equations. Most of the classical calibration techniques in photogrammetry belong to this category (e.g., [3], [1], [15], [4], [11]). One advantage of this type of technique is that the camera model can be very general to cover many types of distortion. Some simple distortion-free models for computer vision applications have also employed this type of technique (e.g., [7], [9]). Another advantage is that the algorithm may achieve high accuracy, provided that the estimation model is good, and correct convergence has been reached. However, since the algorithm is iterative, the procedure may end up with a bad solution unless a good initial guess is available. Furthermore, once distortion parameters are included in the parameter space, the minimization may be unstable if the procedure of iterations is not properly designed. The interaction between the distortion parameters and external parameters can lead to divergence or false solutions.

Manuscript received January 18, 1990; revised February 24, 1992. This research was supported by the Natural Sciences and Engineering Council of Canada under Grant A3890 and grants from Centre de Recherche Informatique de Montreal and Ecole Polytechnique de Montreal. Recommended for acceptance by Associate Editor J. Mundy.

J. Weng is with the Department of Computer Science, Michigan State University, East Lansing, MI 48824–1027.

P. Cohen and M. Herniou are with the Department of Electrical Engineering, Ecole Polytechnique de Montreal, Montreal, Canada H3C 3A7.

IEEE Log Number 9201774.

2) Closed-Form Solution: With this type of scheme, parameter values are computed directly through a noniterative algorithm based on a closed-form solution (e.g., [1], [15], [8], [6]). A set of intermediate parameters is defined in terms of the original parameters. The intermediate parameters can be computed by solving linear equations, and the final parameters are determined from those intermediate parameters. Since no iteration is required, the algorithms are fast. However, such methods have the following disadvantages. First, camera distortion cannot be incorporated, and therefore, distortion effects cannot be corrected. It is worth mentioning that the direct linear transformation (DLT) introduced by Abdel-Aziz and Karara [1] has been extended to incorporate distortion parameters. However, the corresponding formulation is not exact; depth components of control points, in a camera-centered coordinate system, are assumed to be constant. Second, due to the objective to construct a noniterative algorithm, the actual constraints in the intermediate parameters are not considered. Consequently, in the presence of noise, the intermediate solution does not satisfy the constraints, and the accuracy of the final solution is relatively poor.

3) Two-Step Methods: The methods of this type involve a direct solution for most of the calibration parameters and some iterative solution for the other parameters. The existing techniques include those presented by Tsai [13] and Lenz and Tsai [10]. A radial alignment constraint is used to derive a closed-form solution for the external parameters and the effective focal length of the camera. Then, an iterative scheme is used to estimate three parameters: the depth component in the translation vector (external parameter), the effective focal length, and a radial distortion coefficient. In [10], two additional parameters (the image coordinates of the principal point, which were considered to be known in [13]) have been included into the set of iteratively determined parameters. The advantages of their method are as follows: a) A closed-form solution is derived for most of the parameters; b) the closed-form solution is immune to lens radial distortion; c) the number of parameters to be estimated through iterations is relatively small. The disadvantages are as follows: a) Their method can only handle radial distortion and cannot be extended to other types of distortion; b) the solution is not optimal because the information provided by the calibration points has not been fully utilized. The radial component of a point is discarded completely, and only the tangential component is used for the solution. However, although it is generally true that the radial component is less reliable than the tangential component, the tangential component is not error free either. From a statistical point of view, discarding relatively unreliable radial components (which account for half of the observations) results in a less reliable estimator.

In this paper, we address the above problems and introduce our approach to camera calibration.

1) For the calibration of off-the-shelf nonmetric cameras, we derive a model that takes into account major distortions and investigate the amount of penalty caused by less general camera models. This will allow us to address important questions such as the following:

- Is it at all necessary to consider distortion for popular off-the-shelf video cameras with, for instance, a typical 512×512 image resolution?
- How much improvement is brought about by considering one or several types of distortion phenomena in the model?

It was reported in [6] that triangulation with an accuracy of one in 2000 parts had been achieved by calibration based on a distortion-free model. A comparable result was reported in [13] with a simple one-parameter radial distortion model and no consideration of tangential distortion. In Faig's very elaborate model [4], four types of distortion were considered for high-accuracy photogrammetric applications: radial symmetric distortion, decentering distortion, affinity distortion, and distortion caused by nonperpendicularity of axis. We will find some answers to the above questions in our experiments with three different models: the distortion-free model, the radial distortion model, and our complete model that includes both radial and tangential distortions.

2) We adopt a two-step approach to the calibration of our stereo camera system. The first-step consists of using a noniterative algorithm to directly compute a closed-form solution for all the external parameters and some major internal parameters based on a distortion-free camera model. The second step is a nonlinear optimization based on a camera model that incorporates distortion using the solution of the first step as an initial guess. One major difference between our two-step computational approach and those of [13] and [10] is that our second step is an optimization that computes and improves *all* the parameters, whereas in the approaches of [13] and [10], the second step iteratively computes only *a few* parameters that cannot be provided by the first step. Our approach was motivated by the following considerations: a) Since the algorithm that computes a closed-form solution is noniterative, it is fast, and a solution is generally guaranteed. It gives a complete solution to all external and internal parameters of a distortion-free camera model. Since the center of an image has little distortion, only the points near the center of the image (which are called central points) are used for the closed-form solution. Consequently, the closed-form solution is not affected very much by distortion and is good enough to be used as an initial guess for further optimization. b) The nonlinear optimization step can take into account different types of distortion. c) Even if the actual camera is distortion free, nonlinear optimization can still improve the closed-form solution. Indeed, the closed-form solution is usually the solution that minimizes mean-square residuals of some equations. However, because various uncertainties in the different equations have not been properly taken into account, the solution is not optimal. The optimization step will then compute the "best" solution. d) In general, a nonlinear optimization algorithm may converge to a local extremum that is not globally optimal. However, if an approximate solution is given as an initial guess, the number of iterations can be significantly reduced, and the globally optimal solution can be reliably reached. Such a two-step algorithm is much more reliable than a direct iterative algorithm that starts with an arbitrarily chosen initial guess (e.g., "zero" guess). e) Real-time calibration is not necessary in most applications. Usually,

calibration needs to be done only once, and the estimated parameters are used for real-time operation. This is also true for time-varying systems in which the calibrated camera parameters are given as functions of camera control signals. Therefore, the calibration algorithm can use considerable time to obtain highly accurate parameters.

3) We introduce a new measure to assess the accuracy of camera calibration with digital images. The existing measures depend on the setup of the system, especially on the field of view, the baseline length, and the camera-to-object distance. This makes it very difficult to compare techniques of calibrations with different setups. As we know, image resolution inherently limits the accuracy of calibration. The calibrated parameters cannot be exact even if the camera is ideal. However, a good calibration with a lower image resolution may exceed the accuracy of a poor calibration with a higher resolution. It is reasonable to measure the quality of calibration based on image resolution. The measure introduced in this paper is called *normalized stereo camera error*, which is an error measure normalized according to the resolution of the images. Such a normalized measure is important because a) the performance of different calibration approaches can be quantitatively evaluated and compared, and b) this measure indicates whether a particular camera model is elaborate enough under the image resolution. Since the resolution of a digital image is well defined by the number of pixels, this measure is especially convenient for calibrations that are based on digital images.

We present our camera models in the next section. The estimation of the parameters is discussed in Section III. In Section IV, we introduce the normalized stereo camera error. The performance of our method is demonstrated by the results of simulations and experiments with real camera systems in Section V. Finally, Section VI gives concluding remarks.

II. Camera Models

In this section, we consider two types of camera models. The first model is a pinhole camera model that neglects all optical distortion. The second model takes into account several types of distortion. The first model can be viewed as a first approximation of the second one and will actually be considered as such in the calibration procedure.

A. Distortion-Free Camera Model

Let (x, y, z) represent the coordinates of any visible point P in a fixed reference system (world coordinate system), and let (x_c, y_c, z_c) represent the coordinates of the same point in a camera-centered coordinate system. As illustrated in Fig. 1, the origin of the camera-centered coordinate system coincides with the optical center of the camera, and the z_c axis coincides with its optical axis. The image plane, which corresponds to the image sensing array, is assumed to be parallel to the (x_c, y_c) plane and at a distance f to the origin, where f represents the (effective) focal length of the camera. The relationship between the world- and camera-centered coordinate systems

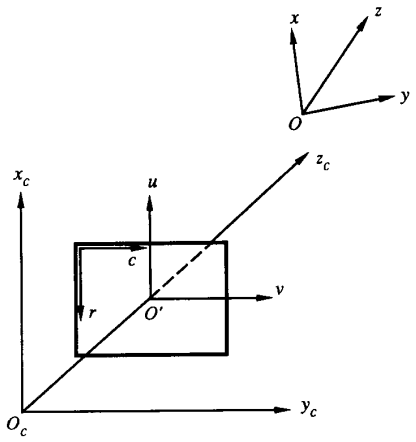


Fig. 1. Coordinate systems.

is given by

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = R \begin{pmatrix} x \\ y \\ z \end{pmatrix} + T \tag{1}$$

where $R = (r_{i,j})$ is a 3×3 rotation matrix defining the camera orientation, and $T = (t_1, t_2, t_3)^\top$ is a translation vector defining the camera position.

We now define, in the image plane, the image coordinate system (O', u, v), where O' represents the principal point of the image plane (i.e., the intersection of the image plane with the optical axis) and where the u and v axes are chosen parallel to the x_c and y_c axes. It should be stressed that owing to possible misalignment of the CCD array, O' does not necessarily coincide with the geometrical center of the image plane. The image plane coordinates of the point P are given by the equations

$$\begin{aligned} u &= f x_c / z_c \\ v &= f y_c / z_c. \end{aligned} \tag{2}$$

Finally, if we denote by (r, c) the position of the corresponding pixel in the digitized image, this position is related to the image-plane coordinates by the expressions

$$\begin{aligned} r - r_0 &= s_u u \\ c - c_0 &= s_v v \end{aligned} \tag{3}$$

where (r_0, c_0) denotes the pixel position of the principal point O'. The coordinates (r, c) can be considered to be the row and column numbers in a CCD array. In other words, the x_c and y_c axes are chosen to be parallel to row and column directions, respectively. As can be noticed in Fig. 1, the adopted conventions impose s_u to be negative and s_v to be positive. Combining (1), (2), and (3) leads to the following expressions that relate pixel position, the world coordinates, and the various parameters to be calibrated

$$\begin{aligned} \frac{u}{f} &= \frac{r - r_0}{f_u} = \frac{r_{1,1}x + r_{1,2}y + r_{1,3}z + t_1}{r_{3,1}x + r_{3,2}y + r_{3,3}z + t_3} \stackrel{\text{def}}{=} \dot{u} \\ \frac{v}{f} &= \frac{c - c_0}{f_v} = \frac{r_{2,1}x + r_{2,2}y + r_{2,3}z + t_2}{r_{3,1}x + r_{3,2}y + r_{3,3}z + t_3} \stackrel{\text{def}}{=} \dot{v} \end{aligned} \tag{4}$$

where $(\dot{u}, \dot{v})$ defines the coordinates in the normalized image plane that is located at $z = 1$, and $f_u = s_u f$ and $f_v = s_v f$ are called row focal length and column focal length, respectively. If we scale the camera, with respect to the focal point, so that the row-to-row (column-to-column) pixel spacing is equal to 1, the corresponding focal length of such a scaled camera is equal to the row (column) focal length. Most conventional cameras produce rectangular images with a ratio of about 4/3 between horizontal and vertical dimensions. If the ratio of row-to-row pixel distances to column-to-column distance is q, the ratio $|f_u/f_v|$ is roughly equal to q^{-1}. However, because of the timing errors, unstability of the scanning electronics, and a possible tilt of the CCD array, the ratio $|f_u/f_v|$ is not exactly q^{-1}, although q can be directly computed from the dimensional specifications of the CCD sensing elements provided by camera manufacturers.

With this camera model, the calibration problem is expressed in the following terms:

Given a sufficient number of visible points whose world coordinates (x_i, y_i, z_i) are known with a high precision, as well as their corresponding observed pixel positions (r_i', c_i'), estimate, in some optimal sense, the value of the internal camera parameters r_0, c_0, f_u, f_v and the external parameters R and T. In general, observed pixel locations (r_i', c_i') are not equal to locations (r_i, c_i) resulting from (4) because of acquisition and spatial digitization noise and point extraction errors.

B. Geometrical Distortion

Geometrical distortion concerns the position of image points in the image plane. As a result of several types of imperfections in the design and assembly of lenses composing the camera optical system, the expressions in (2) do not hold true and must be replaced by expressions that explicitly take into account the positional error thus introduced:

$$\begin{aligned} u' &= u + \delta_u(u, v) \\ v' &= v + \delta_v(u, v) \end{aligned} \tag{5}$$

where u and v are the nonobservable, distortion-free image coordinates, and u' and v' are the corresponding coordinates with distortion. As indicated by (5), the amount of positional error along each coordinate usually depends on the point position. In order to correct distortion, we need to analyze various sources of distortion and model their effects in the image plane.

In this paper, we consider three types of distortion. The first one is caused by imperfect lens shape and manifests itself by radial positional error only, whereas the second and the third types of distortion are generally caused by improper lens and camera assembly and generate both radial and tangential errors in point positions (see Fig. 2).

1) Radial Distortion: Radial distortion causes an inward or outward displacement of a given image point from its ideal location. This type of distortion is mainly caused by flawed radial curvature curve of the lens elements. A negative radial displacement of the image points is referred to as barrel distortion. It causes outer points to crowd increasingly together and the scale to decrease. A positive radial displacement is referred to as pincushion distortion. It causes outer points to spread and the scale to increase. This type of distortion is strictly symmetric about the optical axis. Fig. 3 illustrates the effect of radial distortion.

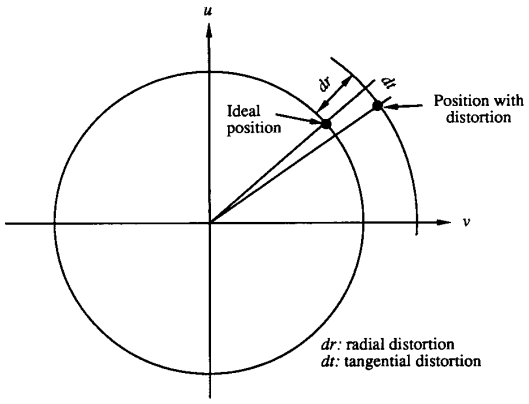


Fig. 2. Radial and tangential distortions.

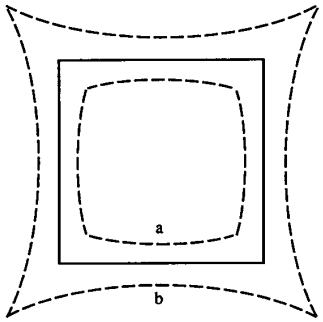


Fig. 3. Effect of radial distortion. Solid lines: no distortion; dashed lines: with radial distortion (a: negative, b: positive).

The radial distortion of a perfectly centered lens is governed by an expression of the following form [11], [4]:

$$\delta_{\rho r} = k_1\rho^3 + k_2\rho^5 + k_3\rho^7 + \cdots \tag{6}$$

where ρ is the radial distance from the principal point of the image plane, and $k_1, k_2, k_3, \ldots$ are the coefficients of radial distortion. At each image point represented by polar coordinates (ρ, φ), radial distortion corresponds to the distortion along radial (ρ) direction. The image point can also be expressed in terms of the Cartesian coordinates (u, v) with

$$\begin{aligned} u &= \rho\cos\varphi \\ v &= \rho\sin\varphi \end{aligned} \tag{7}$$

and the amount of distortion along each of the Cartesian image coordinates can be represented by

$$\begin{aligned} \delta_{ur} &= k_1 u(u^2 + v^2) + O\big[(u, v)^5\big], \\ \delta_{vr} &= k_1 v(u^2 + v^2) + O\big[(u, v)^5\big]. \end{aligned} \tag{8}$$

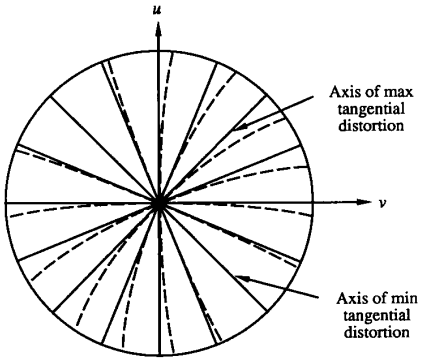


Fig. 4. Effect of tangential distortion. Solid lines: no distortion; dashed lines: with tangential distortion.

2) Decentering Distortion: Actual optical systems are subject to various degrees of decentering, that is, the optical centers of lens elements are not strictly collinear. This defect introduces what is called decentering distortion. This distortion has both radial and tangential components, which can be described analytically by the following expressions [3], [11]:

$$\begin{aligned}\delta_{\rho d} &= 3(j_1\rho^2 + j_2\rho^4 + \cdots)\sin(\varphi - \varphi_0)\\ \delta_{td} &= (j_1\rho^2 + j_2\rho^4 + \cdots)\cos(\varphi - \varphi_0)\end{aligned} \tag{9}$$

where the quantity φ_0 is the angle between the positive u axis and a line of reference known as the axis of maximum tangential distortion as shown in Fig. 4.

The resulting amount of distortion along the u and v axes are given in terms of $\delta_{\rho d}$ and δ_{td} by

$$\begin{pmatrix}\delta_{ud}\\ \delta_{vd}\end{pmatrix} = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}\delta_{\rho d}\\ \delta_{td}\end{pmatrix}. \tag{10}$$

Noting that $\cos\varphi = u/\rho$ and $\sin\varphi = v/\rho$ and letting $p_1 = -j_1\sin\varphi_0$ and $p_2 = j_1\cos\varphi_0$, from (9) and (10), we have

$$\begin{aligned}\delta_{ud} &= p_1(3u^2 + v^2) + 2p_2uv + O\big[(u,v)^4\big],\\ \delta_{vd} &= 2p_1uv + p_2(u^2 + 3v^2) + O\big[(u,v)^4\big].\end{aligned} \tag{11}$$

3) Thin Prism Distortion: Thin prism distortion arises from imperfection in lens design and manufacturing as well as camera assembly (for example, slight tilt of some lens elements or the image sensing array). This type of distortion can be adequately modeled by the adjunction of a thin prism to the optical system, causing additional amounts of radial and tangential distortions [3], [4]. Such distortions can be expressed as

$$\begin{aligned}\delta_{\rho p} &= (i_1\rho^2 + i_2\rho^4 + \cdots)\sin(\varphi - \varphi_1)\\ \delta_{tp} &= (i_1\rho^2 + i_2\rho^4 + \cdots)\cos(\varphi - \varphi_1)\end{aligned} \tag{12}$$

where φ_1 is the angle between the positive u axis and the axis of maximum tangential distortion shown in Fig. 4. Letting $s_1 = -i_1\sin\varphi_1$ and $s_2 = i_1\cos\varphi_1$, a derivation similar to the preceding one leads to the following expressions for the resulting distortions along the u and v axes:

$$\begin{aligned}\delta_{up} &= s_1(u^2 + v^2) + O\big[(u,v)^4\big],\\ \delta_{vp} &= s_2(u^2 + v^2) + O\big[(u,v)^4\big].\end{aligned} \tag{13}$$

4) Total Distortion: We have discussed three types of distortion. Although the decentering distortion and the thin prism distortion lead to similar forms of coefficients in (9) and (12), they model different types of distortion, and in particular, each may have a different axis of maximum tangential distortion. When all the above distortions are present, the effective distortion can be modeled by addition of the corresponding expressions [4], [1]. Combining (8), (11), and (13) gives the total amount of distortion along the u and v axes. Assuming that terms of order higher than 3 are negligible, we obtain

$$\begin{aligned}\delta_u(u,v) &= s_1\big(u^2 + v^2\big) + 3p_1u^2 + p_1v^2 + 2p_2uv\\ &\quad + k_1u\big(u^2 + v^2\big)\\ \delta_v(u,v) &= s_2\big(u^2 + v^2\big) + 2p_1uv + p_2u^2 + 3p_2v^2\\ &\quad + k_1v\big(u^2 + v^2\big).\end{aligned} \tag{14}$$

Letting $g_1 = s_1 + p_1, g_2 = s_2 + p_2, g_3 = 2p_1, g_4 = 2p_2$, the expressions in (14) become

$$\begin{aligned}\delta_u(u,v) &= (g_1 + g_3)u^2 + g_4uv + g_1v^2 + k_1u(u^2 + v^2)\\ \delta_v(u,v) &= g_2u^2 + g_3uv + (g_2 + g_4)v^2 + k_1v(u^2 + v^2).\end{aligned} \tag{15}$$

C. Our Complete Camera Model

Taking into account the distortion along the u and v axes according to (5), the relationship between the distortion-free image point (u, v) and its corresponding pixel location is given by

$$\begin{aligned}u + \delta_u(u,v) &= (r - r_0)/s_u\\ v + \delta_v(u,v) &= (c - c_0)/s_v.\end{aligned} \tag{16}$$

Introducing the new variables

$$\begin{aligned}\hat{u} &= (r - r_0)/f_u\\ \hat{v} &= (c - c_0)/f_v\end{aligned} \tag{17}$$

(16) gives

$$\begin{aligned}\frac{u}{f} &= \hat{u} - \frac{\delta_u(u,v)}{f}\\ \frac{v}{f} &= \hat{v} - \frac{\delta_v(u,v)}{f}.\end{aligned} \tag{18}$$

Because the exact u, v cannot be obtained from actual noise-contaminated observations, the arguments of the modeled distortion are replaced by $\hat{u}, \hat{v}$ [3], [4], [1], [11], which leads to

$$\begin{aligned}\frac{u}{f} &= \hat{u} + \delta'_u(\hat{u},\hat{v})\\ \frac{v}{f} &= \hat{v} + \delta'_v(\hat{u},\hat{v}).\end{aligned} \tag{19}$$

This replacement is reasonable because a) the distortion at the exact image plane projection is approximately equal to that in

the actual projection, and b) the actual distortion coefficients in δ'_u and δ'_v will be estimated based on $\hat{u}$ and $\hat{v}$; therefore, the actual model fitting will be better than what is stated in a).

Redefining the coefficients k_1 and g_1 to g_4, the expressions in (4), (15), and (19) lead to our complete camera model (20), which is shown at the bottom of this page. One can notice that these expressions are linear with respect to the distortion coefficients k_1, g_1, g_2, g_3, g_4. This property will simplify their estimation, as will be explained in Section III.

The calibration problem can now be stated in the following terms:

Given a sufficient number of visible points (x_i, y_i, z_i) and their corresponding pixel locations (r'_i, c'_i), estimate in some optimal sense the set of external and internal nondistortion parameters

$$m = (r_0, c_0, f_u, f_v, T, \alpha, \beta, \gamma)^\top$$

(where α, β, and γ are three independent parameters of rotation matrix R) and the set of distortion parameters

$$d = (k_1, g_1, g_2, g_3, g_4)^\top.$$

After calibration is done for the camera, the estimated calibration parameters m and d can be used to correct distortion and determine the 3-D back-projection line of each sensed point as follows: The measured r and c values of the sensed point give $\hat{u}$ and $\hat{v}$ according to (17). Then, the values of $(\hat{u}, \hat{v})$ are used to evaluate the right-hand sides of the two equations in (20), whose values correspond to the distortion-corrected projection of the point in the normalized image plane $(\dot{u}, \dot{v})$. Finally, the two equations in (20) determine the back-projection line of the sensed point in the world coordinate system.

It should be mentioned that the above camera model has considered only some types of distortions, and it is far from exhausting all possible distortions. In fact, however, it is impossible and unnecessary to consider all types of distortion. Only the major types need to be considered in practice. As can be seen from (20), the three types of distortion lead to two polynomials in $\hat{u}$ and $\hat{v}$, whose coefficients are related. These polynomials can also be interpreted from a different point of view: The calibration problem is to fit polynomial functions in (20) to the measured image data. Such a polynomial fitting can be applied not only to the three types of distortion from which the model is derived but also to other types of distortion. Although this model is most effective to the types of distortion it models, it is less effective to other types. To an extreme, a camera model can be so general that it corresponds to general high-order polynomials without considering all optical characteristics of lens elements and camera assembly at all. However, such an extremely general model would involve excessively many free parameters and would be computationally inefficient and unreliable for usual cameras, whose distortion results from typical optical and mechanical flaws.

III. Resolution Strategies

As established in the previous section, the complete camera model consists of two sets of parameters to be estimated: 1) the vector m, which comprises the external parameters (rotation and translation) and the internal nondistortion parameters and 2) the vector d of internal distortion parameters. The overall estimation problem is nonlinear. We need to investigate the structure of the problem in order to design a stable and efficient estimation procedure.

A. Procedure to Compute the Optimal Solution

Let Ω denote the set of all 3-D control points and ω the set of corresponding image points. Since image-point positions are affected by acquisition and digitization noise, the calibration problem is equivalent to an optimization problem in which the calibration parameters (m^*, d^*) are determined in order to minimize an objective function F (to be defined later):

$$F(\Omega, \omega, m^*, d^*) = \min_{m,d} F(\Omega, \omega, m, d). \tag{21}$$

In order to solve the optimization problem with the nonlinear objective function in (21), we adopt the following procedure:

1. First, let $d = 0$.
2. Compute m, which minimizes $F(\Omega, \omega, m, d)$ with d fixed:
$$\min_m F(\Omega, \omega, m, d). \tag{22}$$
3. Then, with m fixed as current estimate, compute vector d of distortion parameters, which minimizes $F(\Omega, \omega, m, d)$:
$$\min_d F(\Omega, \omega, m, d). \tag{23}$$
4. Go to step 2 unless a certain number of iterations have been performed, and the procedure terminates.

The above procedure was motivated by the following considerations. a) The distortion parameter vector d can be coupled with m to give false minima in $F(\Omega, \omega, m, d)$ because distortion parameters can drastically modify the image position of the points. b) Without a good estimate of the external parameters and the major internal parameters represented by m, the distortion parameters, which are represented by d, cannot be reliably determined. c) Since m corresponds to all parameters in a distortion-free camera model, it cannot be well estimated under significant distortion. Careful measures

$$\frac{r_{1,1}x + r_{1,2}y + r_{1,3}z + t_1}{r_{3,1}x + r_{3,2}y + r_{3,3}z + t_3} = \hat{u} + (g_1 + g_3)\hat{u}^2 + g_4\hat{u}\hat{v} + g_1\hat{v}^2 + k_1\hat{u}(\hat{u}^2 + \hat{v}^2)$$
$$\frac{r_{2,1}x + r_{2,2}y + r_{2,3}z + t_2}{r_{3,1}x + r_{3,2}y + r_{3,3}z + t_3} = \hat{v} + g_2\hat{u}^2 + g_3\hat{u}\hat{v} + (g_2 + g_4)\hat{v}^2 + k_1\hat{v}(\hat{u}^2 + \hat{v}^2). \tag{20}$$

should be taken to deal with the above three problems. The first measure we take is the partition of m and d so that each is computed when the other is fixed. This significantly reduces the harmful interactions between them. The second measure is to compute a good initial m in step 2. If we select points whose projections fall around the center of the image only, we can confidently assume that distortion is small for points in this region and that the assumption $d = 0$ in step 1 will not affect the quality of the estimation of m in step 2 for the first iteration. One should be careful, however, not to impose excessive concentration of the image-points around the image center since such a restriction will affect the accuracy in the estimation of the external parameters. A tradeoff should be adopted between the amount of scattering of the control points and the validity of the assumption $d = 0$. Such a tradeoff depends on the amount of distortion in the lens under calibration, but an optimal tradeoff is not necessary since what we need is just an initial m that is good enough for the start of step 2. (In our implementation, all the image points lying within a radius that equals a quarter of the image side length are considered to be central points.) Once a good estimate of m is obtained, d can be estimated in step 3, using all points in image plane, in order to utilize all available information about distortion. A few repeated iterations lead to improved m and d. Therefore, we select the set Ω of points in such a way that it includes a sufficient subset Ω_1 of central points. Equation (22) of step 2 will be restricted to the subset Ω_1 if step 2 is performed for the first time following step 1.

B. Estimation of m with d Fixed

To give a good preliminary solution to m as a good initial guess for further nonlinear optimization, we first present a closed-form solution to m assuming $d = 0$.

1) Linear Estimation Procedure: Each visible point and its corresponding pixel position give two linear equations derived from (4):

$$\begin{aligned}(r'_i - r_0)x_i r_{3,1} + (r'_i - r_0)y_i r_{3,2} \\ + (r'_i - r_0)z_i r_{3,3} + (r'_i - r_0)t_3 \\ - f_u x_i r_{1,1} - f_u y_i r_{1,2} - f_u z_i r_{1,3} - f_u t_1 = 0 \\ (c'_i - c_0)x_i r_{3,1} + (c'_i - c_0)y_i r_{3,2} \\ + (c'_i - c_0)z_i r_{3,3} + (c'_i - c_0)t_3 \\ - f_v x_i r_{2,1} - f_v y_i r_{2,2} - f_v z_i r_{2,3} - f_v t_2 = 0.\end{aligned} \tag{24}$$

The parameters to be determined in (24) include six external parameters represented by the rotation matrix R (three degrees of freedom) and the translation vector T, as well as four internal parameters r_0, c_0, f_u, and f_v. Each point gives two equations in (24). With these ten unknown parameters, at least five control points are required to give ten nonlinear equations as in (24). However, the solution is not guaranteed even if more points are available because one generally needs to perform an iterative search for solutions, which is not always successful.

In order to derive a closed-form solution to the calibration parameters so that a noniterative algorithm can be used to directly compute the solution, we define the following intermediate parameters [8], [6]:

$$\begin{aligned}W_1 &= f_u R_1 + r_0 R_3 \\ W_2 &= f_v R_2 + c_0 R_3 \\ W_3 &= R_3 \\ w_4 &= f_u t_1 + r_0 t_3 \\ w_5 &= f_v t_2 + c_0 t_3 \\ w_6 &= t_3\end{aligned} \tag{25}$$

where column vectors R_1, R_2 and R_3 correspond to three rows of rotation matrix R. Then, the set of equations in (24) provided by n control points can be expressed in a matrix form

$$AW = 0 \tag{26}$$

where A is a matrix with $2n$ rows and 12 columns, as is shown at the bottom of this page, and W is the vector of all unknown parameters:

$$W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ w_4 \\ w_5 \\ w_6 \end{pmatrix}. \tag{28}$$

Among the multiple solutions to the linear homogeneous equation (26), the one corresponding to the calibration parameters must satisfy two conditions:

- The norm of vector W_3 must be equal to unity since W_3 is equal to the last row of rotation matrix R.
- The sign of w_6 must be compatible with the position of the camera in the world coordinate system. Namely, w_6 must be positive (negative) if the camera is in front of (behind) the (x, y) plane.

$$A = \begin{pmatrix} -x_1 & -y_1 & -z_1 & 0 & 0 & 0 & r'_1 x_1 & r'_1 y_1 & r'_1 z_1 & -1 & 0 & r'_1 \\ 0 & 0 & 0 & -x_1 & -y_1 & -z_1 & c'_1 x_1 & c'_1 y_1 & c'_1 z_1 & 0 & -1 & c'_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -x_n & -y_n & -z_n & 0 & 0 & 0 & r'_n x_n & r'_n y_n & r'_n z_n & -1 & 0 & r'_n \\ 0 & 0 & 0 & -x_n & -y_n & -z_n & c'_n x_n & c'_n y_n & c'_n z_n & 0 & -1 & c'_n \end{pmatrix} \tag{27}$$

Once W is determined (as discussed later), the actual parameters will be given by

$$S = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ s_4 \\ s_5 \\ s_6 \end{pmatrix} = \pm \frac{1}{\|W_3\|} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ w_4 \\ w_5 \\ w_6 \end{pmatrix} \tag{29}$$

where the sign is chosen to satisfy the second condition mentioned above.

Under the assumption of a distortion-free camera (considering central points only), the calibration problem is expressed by the homogeneous system (26). Since the world coordinate system can be defined such that $t_3 \neq 0$, one way to solve for the solution W is to impose the temporary constraint

$$w_6 = t_3 = 1$$

and transform (26) into nonhomogeneous linear system

$$A'W' + B' = 0 \tag{30}$$

where A' denotes the matrix consisting of the first 11 columns of A, B' denotes the last column of A, and W' represents the corresponding reduced unknown vector. In the presence of noise, we can now solve for the vector W' in the linear least-squares system

$$\min_{W'} \|A'W' + B'\|. \tag{31}$$

As indicated previously, the actual solution S must satisfy two constraints and is established according to (29). It then leads to the initial estimate $\bar{m}$ of the external and internal nondistortion parameters

$$\begin{aligned} \bar{r}_0 &= S_1^\top S_3 \\ \bar{c}_0 &= S_2^\top S_3 \\ \bar{f}_u &= -\|S_1 - \bar{r}_0 S_3\| \\ \bar{f}_v &= \|S_2 - \bar{c}_0 S_3\| \\ \bar{t}_1 &= (s_4 - \bar{r}_0 s_6)/\bar{f}_u \\ \bar{t}_2 &= (s_5 - \bar{c}_0 s_6)/\bar{f}_v \\ \bar{t}_3 &= s_6 \\ \bar{R}_1 &= (S_1 - \bar{r}_0 S_3)/\bar{f}_u \\ \bar{R}_2 &= (S_2 - \bar{c}_0 S_3)/\bar{f}_v \\ \bar{R}_3 &= S_3. \end{aligned} \tag{32}$$

These first estimates are established without imposing the constraint in W' (R should be orthonormal). A possible improvement of the solution then consists of solving for the orthonormal matrix $\tilde{R}$, which verifies

$$\|\bar{R} - \tilde{R}\| = \min_R \|\bar{R} - R\| \tag{33}$$

using the closed-form solution presented in the Appendix. This leads to the improved initial estimate $\tilde{m}$ of the other calibration parameters

$$\begin{aligned} \tilde{r}_0 &= S_1^\top \tilde{R}_3 \\ \tilde{c}_0 &= S_2^\top \tilde{R}_3 \\ \tilde{f}_u &= -\|S_1 - \tilde{r}_0 \tilde{R}_3\| \\ \tilde{f}_v &= \|S_2 - \tilde{c}_0 \tilde{R}_3\| \\ \tilde{t}_1 &= (s_4 - \tilde{r}_0 s_6)/\tilde{f}_u \\ \tilde{t}_2 &= (s_5 - \tilde{c}_0 s_6)/\tilde{f}_v \\ \tilde{t}_3 &= \bar{t}_3. \end{aligned} \tag{34}$$

Despite the improvement brought about by the orthonormality constraint, several critiques can be made with respect to overall quality of the estimate $\tilde{m}$. First of all, least-squares estimation is optimal, in the sense of minimum variance, provided that the equation residual consists of uncorrelated zero-mean random noise with equal variance. If the residual components have different variances or if they are correlated, the straightforward least-squares solutions will overtrust the less reliable components and undertrust the more reliable components. Since the elements of A are functions of the pixel coordinates, the above condition for residual distribution is obviously not satisfied. Second, even though the addition of the orthonormality constraint of $\tilde{R}$ brings about an improvement in the accuracy of the estimation procedure, as confirmed by our extensive simulation results, it cannot be proved that the obtained $\tilde{m}$ minimizes (31). Therefore, the solution so obtained has room for improvement. The next section describes an optimal nonlinear procedure for improving m.

2) Nonlinear Optimal Estimation Procedure: Let $U = (u_1, v_1, \ldots, u_n, v_n)^\top$ denote the noise-free image coordinates of n control points, and let $U' = U + N$ denote the corresponding noisy image coordinates. In addition, let $U = f(m, d)$ represent the nonlinear image projection process. We first assume that d is given. Based on the estimate $\tilde{m}$ established through the preceding linear procedure, we can predict the corresponding image projection $\tilde{U}$ of the control points:

$$\tilde{U} = f(\tilde{m}, d). \tag{35}$$

Linearizing function f with respect to m at $\tilde{m}$ leads to

$$U = f(m, d) = f(\tilde{m}, d) + \frac{\partial f(\tilde{m}, d)}{\partial m}(m - \tilde{m}) + h.o.t.. \tag{36}$$

Neglecting higher order terms, (36) can be rewritten as

$$U = \tilde{U} + J(m - \tilde{m}) \tag{37}$$

where J represents the Jacobian matrix of function f in (36).

Expressing U in terms of its noisy version U' results in the linear equation

$$J(m - \tilde{m}) = U' - \tilde{U} - N. \tag{38}$$

The linear minimum variance estimator of m that minimizes $\mathcal{E}\|\tilde{m} - m\|^2$ is the estimator $\tilde{m}$ that minimizes

$$(J(\tilde{m} - m) - U' + \tilde{U})^\top \Gamma^{-1} (J(\tilde{m} - m) - U' + \tilde{U}) \tag{39}$$

or, alternatively, using (37), m minimizes:

$$(U' - f(m, d))^\top \Gamma^{-1} (U' - f(m, d)) \tag{40}$$

where Γ is the covariance matrix of N. Since the image noise N can be reasonably assumed to have a zero mean and be uncorrelated between different components, its covariance matrix Γ is reduced to a diagonal matrix. If we assume that the noise in u and v is proportional to the spacing between consecutive rows and columns, from (16) and (40), the objective function to be minimized is

$$\sum_{i=1}^{n} \left\{ [r_i' - r_i(m,d)]^2 + [c_i' - c_i(m,d)]^2 \right\} \tag{41}$$

where we have added subscript i to denote the variables that correspond to the ith control point, and $r_i(m,d)$ and $c_i(m,d)$ represent the pixel coordinates as functions of the camera parameters (m,d) in accordance with (4). Therefore, the objective function is the sum of the squared discrepancy between the computed and actually observed row and column numbers (which may be real values with a subpixel accuracy). Equation (41) defines the objective function F in Section III-A.

If d is not given, the analogous discussion leads to the same objective function (41) for which both m and d need to be searched for. In Section III-A, we presented a procedure that alternatively searched for m and d, where in each step, either m or d is fixed in order to reduce the interaction between m and d. In fact, for step 3, a closed-form solution can be computed directly as shown in the following section.

C. Estimation of d with m Fixed

Minimizing the objective function in (41) with a given m is a linear least-squares problem. This can be seen directly from the definition of $\hat{u}$ and $\hat{v}$ in (17) and the definition of d in (20). For example

$$\begin{aligned} r(m,d) - r' &= r_0 + f_u \hat{u} - r' \\ &= r_0 + f_u \left(\frac{r_{1,1}x + r_{1,2}y + r_{1,3}z + t_1}{r_{3,1}x + r_{3,2}y + r_{3,3}z + t_3} \right. \\ &\quad \left. - (g_1 + g_3)\hat{u}^2 - g_4 \hat{u}\hat{v} - g_1 \hat{v}^2 - k_1 \hat{u}(\hat{u}^2 + \hat{v}^2) \right) \\ &\quad - r' \end{aligned} \tag{42}$$

which is a linear equation in the components of d. Therefore, with n control points, we can construct a matrix Q and a vector C such that the objective function in (41) becomes

$$\|Qd + C\|^2 \tag{43}$$

where Q is a $2n \times 5$ matrix, C is a $2n$-dimensional vector, and both are constructed from the given m, the 3-D coordinates of the control points, and the measured row and column numbers of the image points. Then, the vector d that minimizes (43) is computed directly without resorting to iterations.

The outline of the estimation procedure has been given in Section III-A. When step 2 is first entered from step 1, a solution to m is computed by a closed-form solution followed by nonlinear optimization for m. When step 2 is entered from step 3, the previous estimate of m is available to be used as an initial estimate for nonlinear optimization for m. A detailed flowchart of the algorithm is presented in Fig. 5.

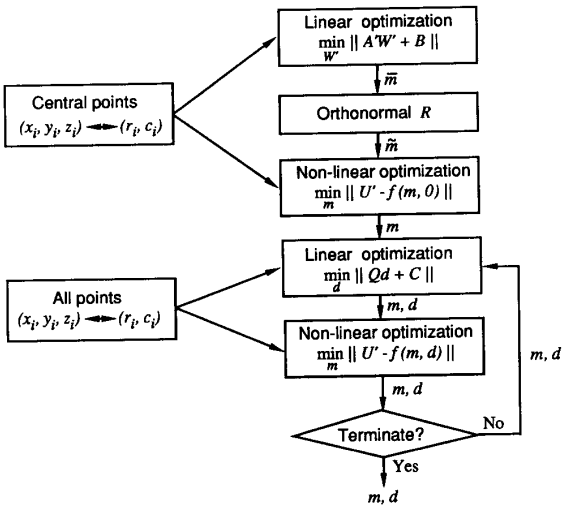


Fig. 5. Flowchart of the estimation procedure. The form of the objective functions for the nonlinear optimizations is used here just for notation simplification. The exact form is given in (41).

IV. Evaluation of Calibration Accuracy

Classical criteria are currently used in computer vision to assess the accuracy of calibration. In [13], three types of measures are mentioned: type I—accuracy of 3-D coordinate measurement obtained through stereo triangulation using the calibrated parameters; type II—radius of ambiguity zone in ray tracing; type III—accuracy of 3-D measurement. All these measurements depend very much on the actual stereo setup, especially on the length of the baseline between the stereo cameras, the field of view of the cameras, and the depth range of the object. For example, with a fixed image resolution, good measures in terms of the above types can be obtained simply by changing the setup in one or more of the following aspects: a) using telelenses so that the array of pixels focus on a smaller area in the scene, b) increasing the baseline between the two cameras in order to reduce the triangulation uncertainty, and c) reducing the distance between the scene and the cameras to take a "close look." Another criterion used in [6] and [13] is the point depth error from triangulation divided by the actual depth. Again, a telelens with a long stereo baseline results in a good measure. However, the improvement resulting from such system modifications is not the merit of calibration. The task of calibration, in the first place, is to obtain the best results based on the system at hand. Only when the calibration cannot meet the system specifications may some modification be made to the system design. Such a system modification might increase the cost or induce difficulties for other modules of the system. For example, a long baseline makes stereo matching more difficult. Often, the system parameters such as working range and field of view are determined by the application and cannot be arbitrarily altered. Therefore, different systems work under different conditions, and it means very little to compare the accuracy of calibration in terms of the above criteria.

Even when the comparison between different calibrations is not of major concern, one still needs a type of measure

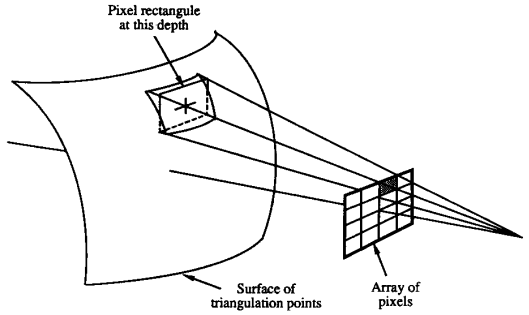


Fig. 6. Backward projection of a pixel to the object surface. The lateral size of the projected pixel at this depth is represented by a pixel rectangle. The lateral error of stereo triangulation is compared with the size of this rectangle to assess the accuracy of calibration.

with which a good calibration has a known range of values. Given a certain value in terms of, e.g., type I error mentioned above, it is not immediately known whether the corresponding calibration is well done. To know it, one needs to investigate how much potential the system (e.g., image resolution) has offered and how much the calibration has obtained. The type of measure to be introduced in the following compares the obtained accuracy with the potential provided by image resolution.

The accuracy of digital image-based camera calibration is mainly limited by the particular resolution of the digital images. There is a major difference between classical film-based camera calibration and digital image-based calibration. While the effective resolution of a film varies with film brand, film grade, film speed, and developing procedure, the resolution of a digital image is determined by the number of pixels. This pixel resolution provides a basis for the evaluation of the calibration accuracy.

Imagine that the array of pixels in an image is projected back to the scene and that each back-projected pixel covers a certain area of the object surface. This area indicates the uncertainty of the basic resolution at this distance (see Fig. 6). Since the orientation of the surface is also related to this area, we consider a plane that is orthogonal to the optical axis and go through the back projection of the pixel center on the object surface. Let the depth of this plane be equal to z, and let the row and column focal lengths be f_u and f_v as defined as in Section II. The back projection of the pixel on this plane is a rectangle (known as a pixel rectangle at this depth in Fig. 6) of size $a \times b$ with a in the u direction and b in the v direction. We have

$$\frac{a}{z} = \frac{1}{|f_u|}, \qquad \frac{b}{z} = \frac{1}{|f_v|}. \tag{44}$$

Uniform digitization noise in a rectangle of $a \times b$ has a variance

$$(a^2 + b^2)/12 = z^2(f_u^{-2} + f_v^{-2})/12. \tag{45}$$

This value characterizes the variance of the lateral error of the back projection of a point at this depth caused by spatial digitization. Let the true coordinates of the ith control point be (x_i, y_i, z_i) represented in the camera-centered coordinate system, and let its coordinates reconstructed through triangulation be $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$. We define *normalized stereo calibration error* (NSCE) to be

$$\text{NSCE} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}{\hat{z}_i^2(f_u^{-2} + f_v^{-2})/12}\right]^{1/2}. \tag{46}$$

This measure represents the mean of the ratio of the lateral triangulation error to the lateral standard deviation of the pixel digitization noise at the depth. Another alternative form of the normalized stereo calibration error uses the root mean squared ratio

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}{\hat{z}_i^2(f_u^{-2} + f_v^{-2})/12}\right]^{1/2}. \tag{47}$$

Since these measures are related only to lateral errors instead of 3-D errors, they are insensitive to baseline length of the stereo setup. These measures are also insensitive to factors such as digital image resolution, field of the view of the cameras, and object-to-camera distance because those factors have similar effects on both the numerators and denominators.

It can be seen from the above definition that the error of each of the two cameras will contribute to the NSCE measure. In other words, the NSCE represents the combined effects of the two calibrations. NSCE < 1 corresponds to a triangulation whose error is lower, on average, than the digitization noise of a pixel at this depth. NSCE $\simeq 1$ indicates a good calibration in which residual distortion is negligible compared with image digitization noise at this depth. NSCE $\gg 1$ reveals a poor calibration.

The NSCE measure is not applicable to a system in which only one camera is involved because the depth of the test points cannot be determined by the single calibrated camera. However, for the purpose of evaluating the accuracy of single camera calibration, the depth of test points can use the ground truth values. This leads to the definition of what we called the *normalized calibration error* (NCE), which is the same as the NSCE except that $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ is evaluated as the intersection of the back-projection line of the sensed test point and the plane $z = z_i$. The NCE measure has advantages similar to those of the NSCE measure and can be used to judge a single camera calibration.

Before we conclude this section, it is worth noting that in our formulation of the optimal solution presented in Section III, we do not take into account the error in the 3-D coordinates of the control points because it should be negligible. In practice, the positions of control points should be determined with a high accuracy (much higher than that represented by the size of the back-projected pixel at the corresponding depth). For example, with a 512 x 512-pixel image taken from an f= 8 mm video camera looking at a scene 1.5 m away, the size of a back-projected pixel is round 2.5mm x 2.5mm. It is not technically difficult to make a calibration pattern with an accuracy that is several orders higher than this. It is not a reasonable practice using a poorly made or poorly positioned calibration pattern in an attempt to correct lens distortion.

V. Experimental Results and Discussions

The camera model and the parameter estimation strategy described in the previous sections have been tested in two different situations. First, the camera model was simulated with specific parameters values, and the estimation strategy was extensively tested on the synthetic data. This experiment allows a thorough investigation of the performance of estimation procedure through a large number of trials, where each has a comparison between the estimated parameters and the ground truth. Second, the overall procedure has been tested with real data provided by off-the-shelf cameras with real stereo setups. This second series of experiments demonstrates the validity of the overall calibration method.

A. Simulations with Synthetic Data

This series of experiments consists first of choosing specific external and internal camera parameters, which are referred to later as ground truth. We specify a rotation by a rotation axis, which is represented by a unit vector (N_x, N_y, N_z) and a rotation angle θ about this axis [2], [14]. Control points are randomly generated and checked for visibility. Since noise is random, the accuracy of the solutions is also random. To show the average performance of our method, the results shown in this section are averaged over 50 random trials, where each has a new set of randomly generated control points. In accordance with previous remarks, the set of control points must contain a sufficient number of points whose projections fall near the center of the image plane. Only these central points will be considered in the linear estimation procedure for the external and nondistortion parameters. Let the simulated camera have a unit focal length $f = 1$ and its image plane be a rectangle with sides $3/4$ unit long in u the direction and 1 unit long in the v direction. This image plane is identical to the normalized image plane defined in (4) since $f = 1$ here. The pixel array has k rows and k columns of pixels ($k = 512$ in the simulations), and therefore, each pixel is a rectangle with side length $a = |f_u^{-1}| = 3/(4k)$ in the u direction and $b = f_v^{-1} = 1/k$ in the v direction. The corresponding noisy image projections of control points are established by adding zero mean uncorrelated Gaussian noise to the exact projections according to (16). The amount of noise added is determined by the corresponding digitization noise. Since the uniform round-off noise with a pixel-to-pixel spacing c has a variance $c^2/12$, the variance of the Gaussian noise to be added is equal to $a^2/12$ in the u component and $b^2/12$ in the v component. Therefore, the standard deviation of the 2-D positional error is

$$\mu_0 = \sqrt{\frac{a^2 + b^2}{12}} = \sqrt{\frac{f_u^{-2} + f_v^{-2}}{12}}. \tag{48}$$

μ_0 represents the amount of uncertainty in the knowledge of an image point location caused by the spatial digitization process involved in the digital image acquisition.

However, as will be explained in the next section, one should design calibration patterns in such a way that control points can be localized in images with a subpixel accuracy. For example, the choice of a calibration pattern containing square shapes allows the definition of square vertices as control points. The image locations of those vertices can be established as the intersections of straight lines interpolated through edge points along the square sides, which results in a subpixel accuracy. Consequently, despite the limit imposed by the digitization process, the projections of control points can be known with a smaller positional error. The simulation here should take into account this subpixel accuracy. Supposing that a reduction by a factor of about 5 in the root-mean-square positional error can be obtained through this interpolation procedure, we simulate this by reducing the aforementioned Gaussian noise by a factor of 5, which leads to the standard deviation of 2-D positional error

$$\mu = \frac{1}{5}\sqrt{\frac{f_u^{-2} + f_v^{-2}}{12}}. \tag{49}$$

Uncertainty parameter μ will be instrumental in the evaluation of the performance of the estimation procedure. Indeed, once estimates m^* and d^* are obtained, the residual image positional error in the normalized image plane (see (4)) will be evaluated as

$$\mu' = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{r_i' - r_i(m^*, d^*)}{f_u}\right)^2 + \left(\frac{c_i' - c_i(m^*, d^*)}{f_u}\right)^2\right]\right\}^{\frac{1}{2}}. \tag{50}$$

Since the standard deviation of the noise added to the normalized image plane is μ, the residual μ' corresponding to the optimal solution should be close to μ.

We first show the results of simulations with distortion-free data. The actual ground truth values and estimation results from the linear and nonlinear procedures are shown in Table I.

TABLE I
Estimation with Synthetic Distortion-Free Data

Ground truth			
Rotation axis	$N_x = 0.2$	$N_y = 1$	$N_z = 5$
Rotation angle	$\theta = 5°$		
Translation	$t_1 = 10$	$t_2 = 6$	$t_3 = 156.5$
Center coordinates	$r_0 = 258$	$c_0 = 254$	
Control points	$n = 64$		

Estimation results		
	Linear	Nonlinear
μ	0.000139	0.000139
μ'	0.000139	0.000137
$\|R - R^*\|/\|R\|$	0.003063	0.003004
$\|T - T^*\|/\|T\|$	0.006793	0.006703
$\|f_u - f_u^*\|/\|f_u\|$	0.007365	0.004643
$\|f_v - f_v^*\|/f_v$	0.004605	0.004595
$\|r_0 - r_0^*\|/r_0$	0.008270	0.008264
$\|c_0 - c_0^*\|/c_0$	0.003415	0.003398

The linear estimation procedure (no distortion assumed) used only those control points that are near the image center, i.e., within a rectangle with a side length $(3/4)/2$ in the u direction and $1/2$ in the v direction and centered at the image plane whose size is $3/4$ by 1. The nonlinear procedure used all the control points and started with the results from the linear estimation procedure as initial guesses. From Table I, it can be seen that the improvement due to the latter procedure is marginal. This phenomenon can be explained by the fact that the linear estimation has already obtained accurate results. This explanation can be confirmed by the closeness of the μ and μ'

TABLE II
ESTIMATION ON SYNTHETIC DATA WITH DISTORTION

Ground truth				
Rotation axis	$N_x = 0.2$	$N_y = 1$	$N_z = 5$	
Rotation angle	$\theta = 5°$			
Translation	$t_1 = 10$	$t_2 = 6$	$t_3 = 156.5$	
Center coordinates	$r_0 = 258$	$c_0 = 254$		
Distortion parameters	$g_1 = 0.02$ $k_1 = 0.01$	$g_2 = -0.009$	$g_3 = -0.02$	$g_4 = 0.009$
Control points	$n = 64$			

Estimation results				
	All points for linear part		Central points for linear part	
	Linear	Nonlinear	Linear	Nonlinear
μ	0.000143	0.000143	0.000139	0.000139
μ'	0.001693	0.000330	0.000174	0.000146
$\|R - R^*\|/\|R\|$	0.010999	0.004015	0.012966	0.012330
$\|T - T^*\|/\|T\|$	0.039840	0.039356	0.017141	0.017163
$\|f_u - f_u^*\|/\|f_u\|$	0.042891	0.025855	0.051549	0.004950
$\|f_v - f_v^*\|/f_v$	0.026249	0.025798	0.004993	0.004943
$\|r_0 - r_0^*\|/r_0$	0.053477	0.049928	0.040025	0.039708
$\|c_0 - c_0^*\|/c_0$	0.019933	0.018734	0.008985	0.008899
$\|k_1 - k_1^*\|/k_1$		0.109362		0.047399
$\|g_1 - g_1^*\|/g_1$		0.033883		0.012728
$\|g_2 - g_2^*\|/g_2$		0.061110		0.020606
$\|g_3 - g_3^*\|/g_3$		0.185069		0.605030
$\|g_4 - g_4^*\|/g_4$		0.239160		0.464835

parameters observable in the linear estimation part of Table I. Since both components of each image point are used in linear estimation, the performance of the linear estimation procedure is already very good with ideal distortion-free cameras.

We now discuss simulations with distortion. Table II gives the ground truth values with nonzero distortion parameters that result in a maximum distortion of about 3 to 4 pixels. To show whether it is important to use central points for linear estimation, we include two different results: that of using all points and that of using central points only. In each case, nonlinear optimization starts with the results from the corresponding linear estimation. Several observations can be made based on the results shown in Table II:

- Concerning the results obtained through the linear estimation procedure, it appears that the selection of only central control points, in the estimation of m, results in a reduction of the residual positional error μ' by a factor of 10. This substantial improvement is confirmed by a lower relative error, in general, for the nondistortion parameters. This improvement has a direct impact on the performance of the subsequent nonlinear part of the estimation procedure, in terms of accuracy as well as convergence time, since the results from the linear procedure were used as the initial estimates in the nonlinear optimization.
- The comparison with the results of the nonlinear optimization shows that a decrease by a factor of 2 in the residual positional error μ' is brought about by the near-center constraint on control points. This improvement is confirmed by a better accuracy in the estimation of most of the nondistortion parameters and in the estimation of most of the distortion parameters. It was also observed, in a substantial number of simulations, that the absence of the near-center constraint caused the nonlinear estimation to diverge because of the insufficient accuracy in the initial estimate $\tilde{m}$.
- Although the positional error is significantly reduced when the near-center constraint is in effect, some of the distortion parameters (e.g., g_3 and g_4) have been less accurately estimated. This phenomenon could be explained by a compensating effect of the polynomial fitting in the presence of noise. Different coefficients of the polynomial can compensate one another somehow to give a similar polynomial shape. On the other hand, since the absolute values of distortion parameters are small, the relative errors do not have to be very small to result in a reasonable distortion correction. There is also a type of compensation between nondistortion parameters m and distortion parameters d in minimizing the objective function shown in (41) and, consequently, μ'.

Fig. 7. Calibration setup.

B. Experiments with Real Images

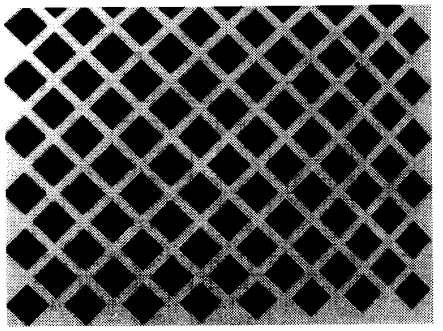
With real images, no *a priori* ground truth is available except for the 3-D coordinates of the control points. The accuracy of the calibration is measured in terms of the accuracy in the reconstructing test points through triangulation. Therefore, calibration needs to be done for each of the two cameras.

1) Description of the Calibration Setup: The setup used in our calibration experiments is illustrated in Fig. 7. It consists of a calibration pattern mounted on a custom-made calibration stand. The images of the pattern were taken by a pair of CCD cameras. The effective part of the CCD sensor array in the cameras has 512 × 480 pixels and a digitizer gives digital images with 8 b/pixel.

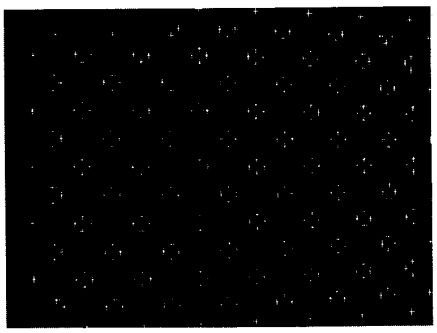
The calibration pattern consists of an array of black diamonds as illustrated in Fig. 8(a).

This pattern was custom made on a 16 × 16-in ultraflat optical glass plate by means of a high-precision photographic process. The positional error of the pattern is lower than 50 μm. The plate is mounted on a custom-made precision stand that can move it vertically with high accuracy by means of a micrometric screw. Each 360° turn of the screw moves the surface 1 mm vertically, and the vertical uncertainty can be controlled within 50 μm. The calibration plate was positioned at several distances from the cameras; therefore, a large number of points could be sensed at various range values.

Two types of camera lens have been tested. The first is a Cosmicar $f = 25$-mm telelens that gives an effective diagonal field of view of about 23°. The second is a Cosmicar $f = 8.5$ wide-angle lens that gives an effective diagonal field of view of approximately 64°.



(a)



(b)

Fig. 8. Calibration pattern and the detected control points: (a) Image of the calibration pattern taken by the first camera with an $f = 8.5$ lens; (b) vertices of complete diamonds are detected with a subpixel accuracy (marked with crosses). Those vertices are used as control points for the calibration.

2) Control and Test Points: The control points are those points used for calibration, and test points are points used for testing the accuracy of calibration. The 3-D control and test points are chosen to be the vertices of completely visible diamonds. Their spatial positions (x_i, y_i, z_i) are reliably established owing to the accuracy of the calibration plate and the precision stand. The corresponding image-point locations are estimated, with a subpixel accuracy, as the intersections of adjacent edge lines, where each is determined by a least-squares fit to edge points along the side of the diamond. The edge-point locations along each side are determined by a third-order polynomial least-squares fit to the intensity profile across the side. Fig. 8(b) displays an example of those detected control points based on the computed data. The correspondence between the 3-D points and their images was established manually.

During the calibration, the external and nondistortion internal parameters are estimated for each of the two stereo cameras. After the calibration is done, the external and internal parameters of each camera are used to determine the back-projection line of each sensed test point in the world coordinate system. The two back-projection lines of the two cameras reconstruct the 3-D position of the test point in the world coordinate system. This is the process of triangulation. Any error in the calibrated parameters will result in error in the reconstructed 3-D position of the test point. Since we know the positions of the test points in the world coordinate system, we can use the difference between the 3-D positions of the reconstructed test points and their true positions to indicate the accuracy of the calibration.

In addition to the NSCE, we will also consider three other conventional measures:

- The first measure M_1 is the average norm of the 3-D positional difference between the reconstructed test points and the true test points.
- The second measure M_2 is the average norm of the x-y plane projection (in the world coordinate system) of the positional error of the test points. Since the two cameras aligned along the y axis, this measure indicates the lateral errors. This measure together with M_1 provide a feeling of the relative distribution of lateral errors and longitudinal (in depth direction) errors.
- The third measure M_3 is the ratio of the average estimated depth (z value) to the average absolute depth error. It provides an assessment of the relative accuracy in the depth measurement of a stereo system, and it can be referred to as "one in M_3 parts." However, as we discussed before, M_3 will be small with both wide-angle lenses and short distances to the test points.

As discussed in Section IV, among the above four types of measures, only the value of the NSCE directly indicates the accuracy of calibration.

Results for the f=25 mm Telelens: Triangulation tests have been performed for the linear estimation step, which corresponds to a distortion-free model, and the nonlinear estimation step, which uses our complete distortion model. In this way, we can see whether the distortion model brings about any improvement for this type of lens. Only central points are used as control points for the linear estimation procedure and its residual measurement. The calibration plate is positioned at four positions: $z = 0$, $z = -50$, $z = -100$, and $z = -200$. The points at $z = -100$ are used as test points, and the remaining points are used as control points for calibration. Table III lists the results obtained. The results of triangulation are measured in the coordinate system centered at the first camera.

On the basis of the M_1, M_2, and M_3 measurements, it can be observed from Table III that the triangulation accuracy has been significantly improved by the estimation of the distortion parameters. In particular, M_3 gives an indication of the increased accuracy in depth estimation. Comparison between the values of the image positional error parameter μ' confirms the improvement of the nonlinear optimization over the linear estimation procedure even though the later is measured only over central points.

The performance of calibration is directly indicated by the NSCE values. As presented in Table III, the measured NSCE value of the linear estimation is equal to 1.7478, which is about 75% larger than 1, which means that the accuracy of this calibration has not reached the potential accuracy allowed by the image resolution. The NSCE value of the nonlinear estimation is, however, just about 6% larger than 1, which indicates a good calibration.

TABLE III
CALIBRATION RESULTS FOR THE $f = 25$ mm LENS AND TESTING

Results of Calibration					
		Linear		Nonlinear	
		First camera	Second camera	First camera	Second camera
Residual	μ'	0.000287	0.000234	0.000253	0.000207
Control points	n	73	77	264	284
Rotation axis	N_x	0.6202	-0.0171	0.6194	-0.0173
	N_y	-0.7801	-0.9834	-0.7807	-0.9831
	N_z	-0.0825	-0.1806	-0.0829	-0.1822
Rotation angle (°)	θ	11.5572	8.8356	11.5618	8.8382
Translation (mm)	t_1	-162.74	-158.43	-162.76	-158.46
	t_2	-190.99	-189.21	-190.98	-189.18
	t_3	866.03	901.59	864.89	901.07
Focal length	f_u	-1923.20	-1914.07	-1921.81	-1913.68
	f_v	1587.44	1580.53	1587.21	1580.38
Center	r_0	257.28	273.15	257.33	273.22
coordinates	c_0	275.13	270.59	275.14	270.63
Distortion	k_1	0.0	0.0	0.2632	0.2325
parameters	g_1	0.0	0.0	-0.000110	-0.002818
	g_2	0.0	0.0	0.003521	0.008344
	g_3	0.0	0.0	0.004880	0.007216
	g_4	0.0	0.0	-0.003583	-0.007339

Results of Triangulation Test			
		Linear	Nonlinear
Average depth	z	822.83mm	821.78mm
Test points	n	128	128
Test	M_1	0.6436mm	0.4365mm
parameters	M_2	0.3391mm	0.2054mm
	M_3	1653.4	2256.9
	NSCE	1.7478	1.0598

TABLE IV
CALIBRATION RESULTS AND TESTING FOR THE $f = 8$ mm LENS WITH A DISTORTION-FREE CAMERA MODEL

Results of Calibration					
		Linear		Nonlinear	
		First camera	Second camera	First camera	Second camera
Residual	μ'	0.000822	0.000772	0.004231	0.003968
Control points	n	113	133	428	424
Rotation axis	N_x	0.9849	-0.9478	0.9919	-0.9606
	N_y	0.1714	0.2583	0.1254	0.2037
	N_z	0.0226	-0.1870	0.0203	-0.1891
Rotation angle (°)	θ	5.8062	6.5499	6.2003	6.5144
Translation (mm)	t_1	-138.71	-136.03	-139.76	-138.23
	t_2	-178.23	-203.85	-179.82	-203.57
	t_3	354.52	400.87	357.02	402.13
Focal length	f_u	-637.56	-637.61	-630.99	-626.37
	f_v	525.61	525.64	519.15	515.39
Center	r_0	241.16	244.67	239.54	241.34
coordinates	c_0	252.08	261.89	254.31	261.30
Distortion	k_1	0.0	0.0	0.0	0.0
parameters	g_1	0.0	0.0	0.0	0.0
	g_2	0.0	0.0	0.0	0.0
	g_3	0.0	0.0	0.0	0.0
	g_4	0.0	0.0	0.0	0.0

Results of Triangulation Test			
		Linear	Nonlinear
Average depth	z	317.67mm	317.53mm
Test points	n	188	188
Test	M_1	3.7574mm	2.1910mm
parameters	M_2	0.6773mm	0.9098mm
	M_3	87.340	170.30
	NSCE	2.9952	3.9793

The absolute values of the radial and tangential parameters do not directly give the relative amount of distortion in the cameras. As can be seen in (20), k_1 is the coefficient of the third power of image coordinates, and the g_i's are the coefficients of the second power. From (4), the maximum absolute value of $\hat{u}$ is around $240/f_u \simeq 0.125$, and that of $\hat{v}$ is $256/f_v \simeq 0.162$. According to the measured distortion parameters, the maximum (symmetrical) radial distortion discussed in Section II-B-1 is about five to seven times larger than that of the combination of decentering and thin prism distortions. The coefficient of radial distortion also indicates that the maximum radial distortion is about 3 to 4 pixels, which occurs at the border of the images.

We have also modified this experiment so that all the control points, and not just central points, are used for the linear estimation. This modification caused the linear estimation to give a slightly worse result compared with that listed in Table III. From this worse result, the following nonlinear estimation procedure reached a triangulation accuracy similar to the corresponding one shown in Table III, but it took 130% more CPU time to converge.

In the above experiments, the plane of test points is located between the planes of control points. To investigate the ability of the calibrated cameras to measure points beyond the range of control points, we let the points at $z = -50$ and $z = -200$ be control points and the points at $z = 0$ be test points. The measured test parameters from the nonlinear estimation are $M_1 = 0.6476$, $M_2 = 0.2335$, $M_3 = 1615.7$ mm, and NSCE $= 1.0735$, which are nearly as good as those listed in Table III. For example, the NSCE value increased by about 1.3%.

4) Results for the f=8 mm Wide-Angle Lens: With a wide-angle lens, the distortion is expected to be more significant than with a telelens. First, we present in Table IV the results with a distortion-free camera model.

Two sets of results are presented. The first set comes from the linear estimation procedure using central control points. The second set comes from the nonlinear estimation procedure assuming $d = 0$ and using the results from linear estimation as initial estimates. The motivation of this arrangement is to show what will happen if a distortion-free camera model is applied to cameras with considerable distortion.

As can be seen from Table IV, the accuracy is very poor for both linear and nonlinear estimation procedures. Since the result from the linear estimation is poor, the nonlinear estimation procedure with $d = 0$ fails to find good estimates and leads to an even worse NSCE value. The residual parameter μ' of the nonlinear estimation is much larger than that of the linear estimation because the former was measured in the entire image plane where boundary regions sustain significant distortion. The poor performance of the nonlinear estimation can be explained as follows. The procedure tries to fit a distortion-free model to the images with significant distortion, but it cannot fit well. To further minimize the objective function in (41), the nondistortion parameter m has to be changed, which results in a worse m. Such a worse m directly leads to a bad NSCE. This is the consequence of neglecting all kinds of distortions.

Table V shows the results from the nonlinear optimization procedure with different completeness of distortion parameters: a) radial distortion only and b) our complete distortion model. The results from the linear estimation procedure shown in Table IV were used as the initial estimates.

Compared with the case of $f = 25$ mm in Table III, the residual values of μ' in Table V are larger mainly because the focal length of the cameras here is much shorter, and consequently, the size of the normalized image plane in the

TABLE V
CALIBRATION RESULTS AND TESTING FOR THE $f = 8$ mm LENS WITH DISTORTION CAMERA MODELS

Results of Calibration					
		Radial only		Radial & tangential	
		First camera	Second camera	First camera	Second camera
Residual	μ'	0.001245	0.001573	0.000992	0.001064
Control points	n	428	424	428	424
Rotation axis	N_x	0.9884	-0.9524	0.9851	-0.9482
	N_y	0.1507	0.2404	0.1701	0.2569
	N_z	0.0201	-0.1874	0.0226	-0.1870
Rotation angle (°)	θ	5.8092	6.4395	5.7869	6.5574
Translation (mm)	t_1	-139.03	-136.33	-138.68	-136.00
	t_2	-178.29	-204.09	-178.22	-203.84
	t_3	353.64	400.54	354.12	400.62
Focal length	f_u	-636.69	-637.48	-636.86	-636.81
	f_v	525.58	526.48	525.62	525.77
Center coordinates	r_0	240.90	244.47	241.15	244.70
	c_0	252.22	261.83	252.19	261.90
Distortion parameters	k_1	0.1742	0.1690	0.1731	0.1720
	g_1	0.0	0.0	0.001640	0.002777
	g_2	0.0	0.0	0.001779	0.000846
	g_3	0.0	0.0	0.014065	0.020896
	g_4	0.0	0.0	0.003981	0.009532

Results of Triangulation Test			
		Radial only	Radial & tangential
Average depth	z	314.03mm	313.84mm
Test points	n	188	188
Test parameters	M_1	0.5811mm	0.4653mm
	M_2	0.3149mm	0.2599mm
	M_3	722.86	897.46
	NSCE	1.4081	1.1627

space $(\dot{u}, \dot{v})$ defined in (4) is much larger. In addition, because of the shorter focal length, the value of M_3 is smaller as has been discussed. Again, the parameter value that directly indicates the accuracy of the calibration is the NSCE. From Table V, we can see that with only radial distortion under consideration, the NSCE value is 41% larger than 1, whereas with our complete distortion model, the NSCE value reduces to 1.1627. With this value of NSCE, we know that most of the distortion has been corrected, compared with the effect of digitization noise.

These results show the importance of accommodating distortion and, in particular, the tangential distortion. Although correcting symmetrical radial distortion leads to significant improvement from the distortion-free model, further correcting remaining distortion makes the NSCE decrease from 1.41 to 1.16. The estimated distortion parameters indicate that the maximum symmetrical radial distortion is about 23 pixels, whereas the maximum of the combination of decentering and thin prism distortions is around 3 pixels. Notice that the estimated value of the radial distortion parameter k_1 is virtually the same from a radial distortion-only model to the complete model, which appears to indicate that the parameters of the decentering and thin prism distortions do little to compensate the radial distortion or vice versa. It is also interesting to note from Table V that for both cameras, the estimated distortion parameter g_3 is significantly larger than other distortion parameters g_is, which implies that the axis of maximum tangential distortion is roughly aligned with the v axis. According to our derivation of the distortion parameters, the parameters j_1, φ_0 in (9) and i_1, φ_1 in (12) can be computed from g_1, g_2, g_3, and g_4. For example, for the first camera, the axis of maximum tangential decentering distortion is computed as $\varphi_0 = -74.20°$ from the u direction.

VI. SUMMARY AND CONCLUSIONS

With a goal of investigating the effects of radial and tangential distortion, the distortion of a camera is modeled by a combination of three typical effects: radial, decentering, and thin prism distortions. These three effects result in five distortion parameters that accommodate up to third-order terms.

In the calibration procedure, nondistortion parameters are first computed using a closed-form solution, and they are then used as initial estimates for further nonlinear optimization. The nondistortion and distortion parameters are decoupled in our procedure so that their harmful interactions are suppressed, which makes the nonlinear optimization more stable.

We have introduced a type of normalized accuracy measure for stereo or single camera calibration, which can be used to evaluate the accuracy of calibration and compare different calibrations without being significantly affected by the differences among systems.

Our experimental results have demonstrated the performance of our approach with both synthetic and real data. Generally, a wide-angle lens causes more distortion than a telelens. Therefore, distortion correction is more important for wide-angle lenses. Symmetrical radial distortion correction results in significant improvement. Correcting tangential distortion also leads to a considerable improvement over just correcting radial distortion.

APPENDIX

The solution of R that satisfies

$$\|RC - D\| = \min \tag{51}$$

where $C = [C_1 C_2 \dots C_n]$ and $D = [D_1 D_2 \dots D_n]$, subject to R (which is a rotation matrix), is as follows.

Define a 4 by 4 matrix B by

$$B = \sum_{i=1}^{n} B_i^\top B_i \tag{52}$$

where

$$B_i = \begin{bmatrix} 0 & (C_i - D_i)^\top \\ D_i - C_i & [D_i + C_i]_\times \end{bmatrix} \tag{53}$$

where $[\bullet]_\times$ is a mapping from a 3-D vector to a 3 by 3 matrix:

$$[(x, y, z)]_\times = \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}. \tag{54}$$

Let $q = (q_0, q_1, q_2, q_3)^\top$ be a unit eigenvector of B associated with the smallest eigenvalue. The solution of rotation matrix R in (51) is shown at the top of the next page. For proofs, see [12], [5], and [14].

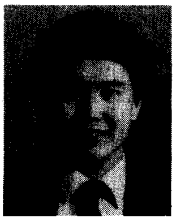
$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_2q_1 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_3q_1 - q_0q_2) & 2(q_3q_2 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}. \quad (55)$$

ACKNOWLEDGMENT

The instrumentation support from H. Nguyen is gratefully acknowledged.

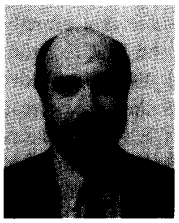
REFERENCES

[1] Y. I. Abdel-Aziz and H. M. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry," in *Proc. Symp. Close-Range Photogrammetry* (Urbana, IL), Jan. 1971, pp. 1–18.
[2] O. Bottema and B. Roth, *Theoretical Kinematics*. Amsterdam: North-Holland, 1979.
[3] D. C. Brown, "Decentering distortion of lenses," *Photogrammetric Eng. Remote Sensing*, May 1966, pp. 444–462.
[4] W. Faig, "Calibration of close-range photogrammetric systems: Mathematical formulation," *Photogrammetric Eng. Remote Sensing*, vol. 41, no. 12, pp. 1479–1486, Dec. 1975.
[5] O. D. Faugeras and M. Hebert, "A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces," in *Proc. 8th Int. Joint Conf. Artificial Intell.* (Karlsruhe, W. Germany), Aug. 1983, pp. 996–1002.
[6] O. D. Faugeras and G. Toscani, "Calibration problem for stereo," in *Proc. Int. Conf. Comput. Vision Patt. Recogn.* (Miami Beach, FL), June 1986, pp. 15–20.
[7] D. B. Gennery, "Stereo-camera calibration," in *Proc. 10th Image Understanding Workshop*, 1979, pp. 101–108.
[8] S. Ganapathy, "Decomposition of transformation matrices for robot vision," in *Proc. IEEE Int. Conf. Robotics Automat.* (Atlanta), Mar. 1984, pp. 130–139.
[9] A. Isaguirre, P. Pu, and J. Summers, "A new development in camera calibration calibrating a pair of mobile cameras," in *Proc. IEEE Int. Conf. Robotics Automat.* (St. Louis), Mar. 1985, pp. 74–79.
[10] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology," in *Proc. IEEE Int. Conf. Robotics Automat.* (Raleigh, NC), Mar. 1987, pp. 68–75.
[11] *Manual of Photogrammetry*. Amer. Soc. Photogrammetry, 1980, 4th ed.
[12] M. D. Shuster, "Approximate algorithms for fast optimal attitude computation," in *Proc. AIAA Guidance Contr. Spec. Conf.* (Palo Alto, CA), Aug. 1978, pp. 88–95.
[13] 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 Automat.*, vol. RA-3, no. 4, pp. 323–344, Aug. 1987.
[14] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: Algorithm, error analysis and error estimation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 5, pp. 451–476, May. 1989.
[15] K. W. Wong, "Mathematical formulation and digital analysis in close-range photogrammetry," *Photogrammetric Eng. Remote Sensing*, vol. 41, no. 11, pp. 1355–1373, Nov. 1975.



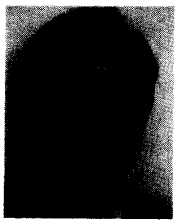
Juyang Weng (S'85–M'88) received the B.S. degree from Fudan University, Shanhai, China, in 1982 and the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, in 1985 and 1988, respectively, all in computer science.

From September 1984 to December 1988, he was a research assistant at the Coordinated Science Laboratory, University of Illinois. In the summer of 1987, he was employed at IBM Los Angeles Scientific Center, Los Angeles, CA. Since January 1989, he has been a researcher at the Centre de Recherche Informatique de Montréal, Montréal, Canada, while he has also been associated with Ecole Polytechnique de Montréal. Since October 1990, he has held a visiting assistant professor position at the University of Illinois at Urbana-Champaign. He is now an assistant professor with the Department of Computer Science, Michigan State University. His current research interests include computer vision, image processing, neural networks, object modeling and representation, parallel architecture for image processing, autonomous navigation, and artificial intelligence.



Paul Cohen received the B.S. degree in electronics from the Insitut National Polytechnique de Toulouse, France, in 1970 and the M.Sc. and Ph.D. degrees in electrical engineering from the Université de Sherbrooke in Quebec in 1972 and 1976, respectively.

Since 1978, he has been with the Department of Electrical and Computer Engineering, Ecole Polytechnique de Montréal, Montréal, Canada, where he is presently a full professor in charge of the Perception and Robotics Laboratory. His research interests are in image processing, computer vision, and pattern recognition.



Marc Herniou received the engineering degree in electronics and computer science from the Université de Paris-Sud-Orsay in 1989.

In 1989, he was at the Ecole Polytechnique de Montréal, Montréal, Canada, where he worked on computer vision. He is currently a Research and Development engineer in Schlumberger, France.