Implicit and Explicit Camera Calibration: Theory and Experiments

Article *in* IEEE Transactions on Pattern Analysis and Machine Intelligence · June 1994

DOI: 10.1109/34.291450 · Source: IEEE Xplore

Implicit and Explicit Camera Calibration: Theory and Experiments

Guo-Qing Wei, *Member, IEEE,* and Song De Ma, *Senior Member, IEEE*

***Abstract*—By implicit camera calibration, we mean the process of calibrating a camera without explicitly computing its physical parameters. Implicit calibration can be used for both three-dimensional (3-D) measurement and generation of image coordinates. In this paper, we present a new implicit model based on the generalized projective mappings between the image plane and two calibration planes. The back-projection and projection processes are modelled separately to ease the computation of distorted image coordinates from known world points. A set of constraints of perspectivity is derived to relate the transformation parameters of the two calibration planes. Under the assumption of the radial distortion model, we present a computationally efficient method for explicitly correcting the distortion of image coordinates in frame buffer without involving the computation of camera position and orientation. By combining with any *linear* calibration techniques, this method makes explicit the camera physical parameters. Extensive experimental comparison of our methods with the classic photogrammetric method and Tsai's method in the aspects of 3-D measurement (both absolute and relative errors), the prediction of image coordinates, and the effect of the number of calibration points, is made using real images from 15 different depth values.**

***Index Terms*—Projective mapping, back-projection, projection, implicit distortion compensation, perspectivity conditions, projective invariants, explicit distortion correction.**

I. Introduction

CAMERA calibration is a preliminary step toward computational vision. Tsai [19] gave a comprehensive survey of the state of the art. Generally, camera calibration means the process of computing a camera's physical parameters, like image center, focal length, position and orientation, etc., [4], [26], [27], [7], [19], [9], [25]. We call this kind of calibration explicit calibration. Explicit calibration is of universal use in all aspects of computer vision. Yet in some specific cases, like stereo vision, the camera physical parameters are not necessarily required. Some intermediate parameters can also be calibrated for either making three-dimensional (3-D) measurement (back-projection) or predicting image coordinates from known world coordinates (projection). We call this kind of calibration implicit calibration. Since we do not attach any physical meanings to the intermediate parameters, the robustness of these parameters is of less concern, as long as they combine together to perform the correct 3-D measurement.

Martins' two-plane method [14] is the first of the latter kind, which was aimed only for 3-D reconstruction. The method considers lens distortions, but the calibrated parameters are generally not globally valid in the whole image plane [22]. The perspective transformation matrix method [5] and the recent work of Mohr and Morin [15] can be used for both 3-D reconstruction and the computation of image coordinates. Yet no lens distortion can be considered.

In this paper, we present an implicit two-plane calibration method, which models both the back-projection and the projection processes globally in the presence of lens distortions. The computation of image coordinates has an analytic form for any types of lens distortions. Besides, the method does not need any *a priori* knowledge about the distortion type. This appears to us to be an important feature. This is because, for explicit camera calibration, Tsai ever noted that any more elaborate distortion model than a radial one not only would not help increase accuracy but also would cause numerical instability [19]. This was confirmed by our explicit calibration experiments. Yet in some other reports (Faugeras and Toscani [6], Weng [25]), especially when wide angle cameras were used, it was found that adding a nonradial distortion component in the distortion model improved accuracy. This means that an explicit calibration method needs a suitable distortion model, which depends on the lenses one uses.

In cases where camera physical parameters are needed, explicit calibration methods have to be used. Classic photogrammetric methods solve for all the camera parameters simultaneously by full-scale nonlinear optimization [4], [25], [26], Under the radial distortion model, Tsai [19] derived a radial alignment constraint (RAC), which is independent of the radial distortion coefficient and some other camera parameters. By assuming a known image center, most of the camera parameters can be solved for *linearly* from the RAC's. Thus the algorithm worked very fast. If the image center is not known in advance, Lenz and Tsai [17] determined its location by a nonlinear approach based on minimizing the RAC residual error. However, we found that the RAC method does not give the global solution in the case of noisy data [23].

In this paper, we present an explicit calibration method, which makes a good compromise between computational cost and accuracy. This is done by using projective invariants for prior radial distortion correction [21]. Invariants have

Manuscript received September 23, 1992. Recommended for acceptance by Associate Editor Dr. R. Bolle.

G.-Q. Wei is with the Institute of Robotics and and System Dynamics, German Aerospace Research Establishment, DLR, 82230 Oberpfaffenhofen, Germany. He is also with the Institue of Automation, Chinese Academy of Sciences, P.R.China.

S. D. Ma is with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R.C.

IEEE Log Number 9215846.

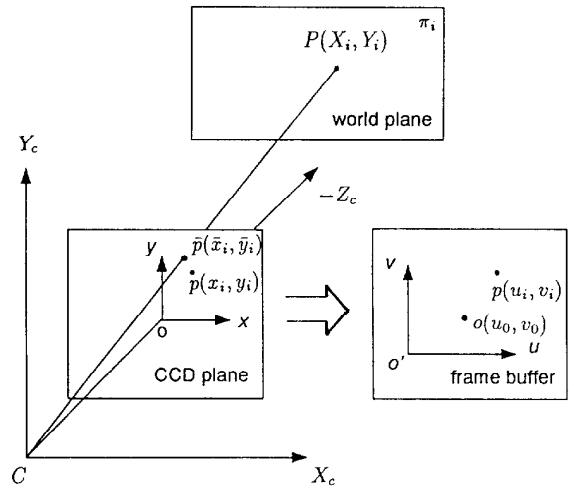

Fig. 1. The imaging geometry.

found some applications in image understanding [12], object recognition [13], and relative positioning [15]. Its use in calibrating a lens' distortion is rare. Our basic idea of using invariants for calibrating a lens' distortion is that distortion makes invariants *vary*, so that we can use the variance to compute the distortion.

The paper is organized as follows. Section II describes the general formulas of our implicit calibration method by using two calibration planes. Section III specializes the method under the pinhole constraint. Section IV presents our explicit calibration method for correcting radial distortion. Section V reports experimental results on both the implicit and the explicit methods. Extensive experimental comparison with several existing calibration methods is made in different aspects. In Section VI, we give some conclusions.

II. Implicit Calibration: The General Case

A. The Back-Projection Problem

Suppose there are two calibration planes π_1 and π_2 in 3-D space, and we define two reference coordinate systems $O_1 : X_1 - Y_1 - Z_1$ and $O_2 : X_2 - Y_2 - Z_2$, with $(X_1, Y_1) \in \pi_1$ and $(X_2, Y_2) \in \pi_2$. Denote the ideal (undistorted) image coordinates of a world point $P : (X_i, Y_i) \in \pi_i$ by $\bar{p} : (\bar{x}_i, \bar{y}_i)$ in the camera's CCD plane, $i = 1, 2$, with the CCD coordinate axes x and y parallel to the camera coordinate axes X_c and Y_c, and the coordinate origin at the image center. The ideal image coordinates, however, are not available. They are distorted by the camera lens to give the distorted CCD coordinates $p : (x_i, y_i)$, which are observed through the image buffer coordinates (u_i, v_i) in pixels, see Fig. 1. Assume the frame buffer coordinate axes are parallel to the CCD coordinate axes. One can thus compute the ideal CCD coordinates $(\bar{x}, \bar{y})$ from the frame buffer coordinates (u, v) as follows [4], [26], [25] (Here we drop the subscript since the computation is independent of the position of the world point.):

$$x = (u - u_0)d_x, \quad y = (v - v_0)d_y, \tag{1}$$

$$\bar{x} = x + \delta^{(x)}(x, y), \ \bar{y} = y + \delta^{(y)}(x, y), \tag{2}$$

$$\delta^{(x)}(x, y) = k_1 x(x^2 + y^2) + (p_1(3x^2 + y^2) + 2p_2 xy) + s_1(x^2 + y^2), \tag{3}$$

$$\delta^{(y)}(x, y) = \underbrace{k_1 y(x^2 + y^2)}_{\text{radial}} + \underbrace{2p_1 xy + p_2(3x^2 + y^2)}_{\text{decentering}} + \underbrace{s_2(x^2 + y^2)}_{\text{thin prism}} \tag{4}$$

where (u_0, v_0) are the coordinates of the image center (the principal point) in the frame buffer; d_x and d_y are the distances between adjacent pixels in the horizontal and vertical directions of the CCD plane, respectively; $(\delta^{(x)}(x, y), \delta^{(y)}(x, y))$ are the distortion corrections to (x, y); k_1, p_1, p_2, s_1, and s_2 are the distortion coefficients.

To formulate our back-projection problem, we have shown [22] that the direct transformation from the image coordinates (u_i, v_i) in the frame buffer to the world coordinates (X_i, Y_i) on the calibration plane while accounting for the above distortions, can be written, for π_1 as

$$X_1 = \frac{\sum_{0 \le i+j \le 3} a_{ij}^{(1)} u_1^i v_1^j}{\sum_{0 \le i+j \le 3} a_{ij}^{(3)} u_1^i v_1^j}, \quad Y_1 = \frac{\sum_{0 \le i+j \le 3} a_{ij}^{(2)} u_1^i v_1^j}{\sum_{0 \le i+j \le 3} a_{ij}^{(3)} u_1^i v_1^j} \tag{5}$$

where $\{a_{ij}^{(k)}\}$ are the transformation coefficients. Equation (5) can be rewritten in the homogeneous form as

$$\rho_1 \mathbf{x}_{w1} = \mathbf{A}\hat{\mathbf{u}}_1 \tag{6}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{33}^{(1)}, & \ldots, & a_{00}^{(1)} \\ a_{33}^{(2)}, & \ldots, & a_{00}^{(2)} \\ a_{33}^{(3)}, & \ldots, & a_{00}^{(3)} \end{pmatrix}, \tag{7}$$

$$\hat{\mathbf{u}}_1 = (u_1^3, u_1^2 v_1, u_1 v_1^2, v_1^3, u_1^2, u_1 v_1, v_1^2, u_1, v_1, 1)^T,$$

and $\rho_1 = \mathbf{a}^{(3)}\hat{\mathbf{u}}_1$, with $\mathbf{a}^{(3)}$ the third row of $\mathbf{A}$; $\mathbf{x}_{w1} = (X_1, Y_1, 1)^T$.

Similarly, the transformation for plane π_2 can be written as

$$\rho_2 \mathbf{x}_{w2} = \mathbf{B}\hat{\mathbf{u}}_2 \tag{8}$$

where the definitions of $\mathbf{x}_{w2}$, $\mathbf{B}$, ρ_2, and $\hat{\mathbf{u}}_2$ are similar to that for π_1.

When given a set of calibration points on each calibration plane and the corresponding image coordinates in the frame buffer, we can then solve (6) and (8) for the coefficients $\mathbf{A}$ and $\mathbf{B}$ without having to identify the camera physical parameters. With the matrices $\mathbf{A}$ and $\mathbf{B}$, we can compute the line of sight of any image point by mapping the image point to the two calibration planes using (6) and (8) and then drawing a line through the obtained points. With the lines of sight from two cameras, 3-D reconstruction can be easily made directly in the world coordinate system. Under this two-plane model, the camera perspective center C can be computed by intersecting two (or more) lines of sight in each camera.

Equations (6) and (8) can be viewed as a generalization of the projective mappings in the distortion-free case through adding higher order power terms. Although the power terms up to the third order were derived, we found that using only up to the second order has already been able to achieve a significant effect of distortion compensation. Thus, according to the order n up to which we use, we will call the general form of (6) and (8) the *nth order projective mapping*, $n = 1$, 2, or 3.

B. The Projection Problem

Equations (6) and (8) combine together to represent the mapping from the image space to the world space. Since they are nonlinear in the image coordinates u and v, it is difficult to inverse them to determine u and v when given a world point (X, Y) on either of the calibration planes. The computation of image coordinates from known world coordinates is posed as the projection problem.

Instead of seeking a way to inverse (6) and (8), we will model the mapping from the world space to the image space in separation from its reverse process. We notice that (3) and (4) are the approximates of their original forms [2], [25]. The exact distortion equations should read as

$$x = \bar{x} - \delta^{(x)}(\bar{x}, \bar{y}), \quad y = \bar{x} - \delta^{(y)}(\bar{x}, \bar{y}) \tag{9}$$

where the meanings of (x, y), $(\bar{x}, \bar{y})$, and the functional forms of $\delta^{(x)}()$ and $\delta^{(y)}()$ are the same as that in (2), (3), and (4), but the coefficient values in $\delta^{(x)}(\bar{x}, \bar{y})$ and $\delta^{(y)}(\bar{x}, \bar{y})$ are different from that in $\delta^{(x)}(x, y)$ and $\delta^{(y)}(x, y)$. Equation (9) models the distortion process in the direction from the ideal coordinates to the distorted coordinates. With this modelling we can establish the projection for π_1 as follows. We know that the transformation from world coordinates to ideal image coordinates can be written, for π_1 as

$$\bar{x}_1 = \frac{\bar{a}_1 X_1 + \bar{b}_1 Y_1 + \bar{c}_1}{\bar{p}_1 X_1 + \bar{q}_1 Y_1 + \bar{r}_1}, \quad \bar{y}_1 = \frac{\bar{d}_1 X_1 + \bar{e}_1 Y_1 + \bar{f}_1}{\bar{p}_1 X_1 + \bar{q}_1 Y_1 + \bar{r}_1} \tag{10}$$

where $(\bar{a}_1, \cdots, \bar{r}_1)$ are the unknown projective transformation coefficients. We substitute (10) into the expansion of (9), and express its left-hand side terms x_1 and y_1 by u_1 and v_1 using (1). Rearranging the terms, we get the transformation from world space to image space for π_1 as:

$$u_1 = \frac{\sum_{0 \le i+j \le 3} \bar{a}_{ij}^{(1)} X_1^i Y_1^j}{\sum_{0 \le i+j \le 3} \bar{a}_{ij}^{(3)} X_1^i Y_1^j}, \quad v_1 = \frac{\sum_{0 \le i+j \le 3} \bar{a}_{ij}^{(2)} X_1^i Y_1^j}{\sum_{0 \le i+j \le 3} \bar{a}_{ij}^{(3)} X_1^i Y_1^j} \tag{11}$$

where $\{\bar{a}_{ij}^{(k)}\}$ are the resulting coefficients in the derivation. Equation (11) can be written in the homogeneous form as

$$\bar{\rho}_1 \mathbf{u}_1 = \bar{\mathbf{A}} \hat{\mathbf{x}}_{w1} \tag{12}$$

where $\bar{\mathbf{A}}$ is the coefficient matrix of $\{\bar{a}_{ij}^{(k)}\}$ defined similarly to (7), $\mathbf{u}_1 = (u_1, v_1, 1)^T$, $\bar{\rho}_1 = \bar{\mathbf{a}}^{(3)} \hat{\mathbf{x}}_{w1}$, and

$$\hat{\mathbf{x}}_{w1} = (X_1^3, X_1^2 Y_1, X_1 Y_1^2, Y_1^3, X_1^2, X_1 Y_1, Y_1^2, X_1, Y_1, 1)^T. \tag{13}$$

Similarly, the transformation for π_2 can be written as

$$\bar{\rho}_2 \mathbf{u}_2 = \bar{\mathbf{B}} \hat{\mathbf{x}}_{w2} \tag{14}$$

with $\bar{\mathbf{B}}$, $\mathbf{u}_2$, $\bar{\rho}_2$, and $\hat{\mathbf{x}}_{w2}$ similarly defined as for π_1.

Using the same calibration data as in the back-projection case, we can then solve (12) and (14) for the elements of $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$. Similar to the back-projection case, it is usually enough to use the power terms up to the second order for the above calibration.

Since we have obtained the perspective center C in the back-projection case, we can now compute the image coordinates of any 3-D world point P in two steps: (a) connect P with the perspective center C, find the intersection of $\overline{CP}$ with either π_1 or π_2; (b) map the intersection point on π_1 or π_2 to the image plane by either (12) or (14). Notice that the computation has modelled the distortion process.

It can be shown [23] that epipolar curves, which are defined as the distortion of epipolar lines, can be computed in closed forms from the above back-projection and projection transformations.

III. Realizing the Pinhole Model

Like the original two-plane model [14] and its extensions [11], [10], our model in the last section does not use the pinhole constraint either. In particular, (6), (8), (12), and (14) do not necessarily represent *perspective* mappings, nor the *same* perspective mapping. Intuitively, since they model the imaging process of the same camera, there should exist some relationship between them. In this section, we shall fully explore this relationship.

A. The Back-Projection Problem

If we use (6) and (8) to map the same image points to the two calibration planes, we get a set of corresponding points. We are now interested in the transformation between the set of corresponding points. Under the pinhole constraint, we will expect the corresponding points to satisfy the same perspective process.

A general projective transformation between the two calibration planes can be described by a matrix $\mathbf{C} = [c_{ij}]_{3\times 3}$ as follows

$$\rho \mathbf{x}_{w2} = \mathbf{C} \mathbf{x}_{w1} \text{ or } \mathbf{x}_{w2} = \frac{\mathbf{C} \mathbf{x}_{w1}}{\mathbf{c}_3 \mathbf{x}_{w1}} \tag{15}$$

where $\mathbf{x}_{w1}$ and $\mathbf{x}_{w2}$ are the correspondence points; $\mathbf{c}_3$ is the third row of $\mathbf{C}$. Next, we will derive conditions for (15) to be a perspective mapping, i.e., a *perspectivity*.

Since the geometrical arrangement of the two calibration planes is known, we can describe it by the rigid transformation between the two coordinate systems O_1 and O_2 as

$$(X_1, Y_1, Z_1)^T = \bar{\mathbf{R}} (X_2, Y_2, Z_2)^T + \bar{\mathbf{t}} \tag{16}$$

where $\bar{\mathbf{R}} = [\bar{r}_{ij}]_{3\times 3}$, $\bar{\mathbf{t}} = (\bar{t}_1, \bar{t}_2, \bar{t}_3)^T$ are the rotation matrix and the translation vector, respectively. Thus, any point P_2 : $(X_2, Y_2) \in \pi_2$ in the O_2 system can be expressed in the O_1

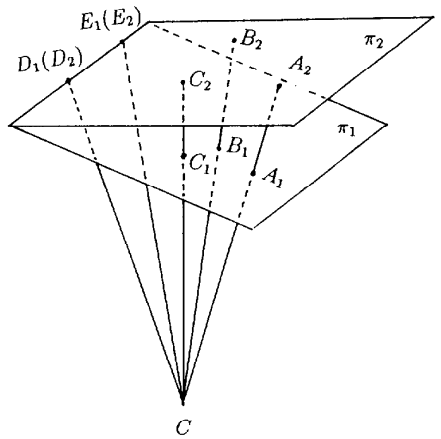

Fig. 2. If A_1, B_1, C_1 on π_1 and A_2, B_2, C_2 on π_2 form a perspective mapping, then the mapping keeps the points D_1, E_1 on the intersection line invariant.

system by

$$\begin{pmatrix} X_2' \\ Y_2' \\ Z_2' \end{pmatrix} - (\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \bar{\mathbf{t}}) \begin{pmatrix} X_2 \\ Y_2 \\ 1 \end{pmatrix} = \hat{\mathbf{R}} \mathbf{x}_{w2} \tag{17}$$

where $\bar{\mathbf{r}}_1$ and $\bar{\mathbf{r}}_2$ are the first and second columns of $\bar{\mathbf{R}}$, and $\hat{\mathbf{R}} = (\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \bar{\mathbf{t}})$.

In terms of (17), the mapping in (15) can be expressed in the same coordinate system O_1 as

$$\begin{pmatrix} X_2' \\ Y_2' \\ Z_2' \end{pmatrix} = \frac{\hat{\mathbf{R}}\,\mathbf{C}\mathbf{x}_{w1}}{\mathbf{c}_3\mathbf{x}_{w1}} = \frac{\mathbf{C}'\mathbf{x}_{w1}}{\mathbf{c}_3\mathbf{x}_{w1}} \tag{18}$$

where $\mathbf{C}' = \hat{\mathbf{R}}\,\mathbf{C} = [c_{ij}']_{3\times 3}$. Equation (18) represents a 3-D projective transformation, i.e., a *projectivity*, expressed in the O_1 system.

Definition [8]: Under a projectivity $\mathcal{P}$ in projective space S, a point $X \in S$ is said to be an invariant point if $\mathcal{P}(X) = X$; a line L is said to be an invariant line if $\mathcal{P}(L) = L$; a line L is said to be pointwise invariant if $\mathcal{P}(L) = L$ and $\forall X \in L : \mathcal{P}(X) = X$.

Lemma 1: A projectivity $\mathcal{P}$ relating two planes in S is a perspectivity if and only if the common line of the two planes is pointwise invariant.

Fig. 2 explains the implication of this lemma, which can be seen as a natural extension of a similar theorem concerning a projectivity relating two lines on a plane [3].

The common line l of the two calibration planes in our case can found from (16) by letting $Z_1 = 0$ and $Z_2 = 0$ as

$$l : k_1 X_1 + k_2 Y_1 + k_3 = 0 \tag{19}$$

where

$$k_1 = \bar{r}_{13},\ k_2 = \bar{r}_{23},\ k_3 = -\bar{r}_{13}\bar{t}_1 - \bar{r}_{23}\bar{t}_2 - \bar{r}_{33}\bar{t}_3.$$

By applying Lemma 1 to (18) and denoting the corresponding mapping by $\mathcal{P}_{12} : \pi_1 \rightarrow \pi_2$, we have the following.

Lemma 2 (Perspectivity Conditions): A set of necessary and sufficient conditions for $\mathcal{P}_{12}$ to be a perspectivity, when the two calibration planes are in ordinary relative positions, is

$$C1 : k_2 c_{31}' - k_1 c_{32}' = 0, \tag{20}$$
$$C2 : k_3 c_{32}' - k_2 c_{33}' = 0, \tag{21}$$
$$C3 : k_2 c_{31} - k_1 c_{32} = 0, \tag{22}$$
$$C4 : k_2(c_{11}' - c_{33}) - k_1 c_{12}' + k_3 c_{32} = 0, \tag{23}$$
$$C5 : k_3 c_{12}' - k_2 c_{13}' = 0. \tag{24}$$

Proof of this lemma can be found in Appendix A.

The constraints above are derived by assuming k_1, k_2 and k_3 all nonzero. This does not lose any generality, since we can always choose the O_1 system such that k_1, k_2 and k_3 are made nonzero. The only exception is when π_1 and π_2 are parallel to each other. In this case, we have

Lemma 2' The perspectivity conditions for parallel calibration planes are

$$C1p : c_{31} = 0, \tag{25}$$
$$C2p : c_{32} = 0, \tag{26}$$
$$C3p : c_{21}' = 0, \tag{27}$$
$$C4p : c_{11}' = c_{22}', \tag{28}$$
$$C5p : c_{12}' = 0. \tag{29}$$

Derivation of these constraints can be found in Appendix B.

In terms of the mapping $\mathbf{C}$, we can now replace the back-projection mapping $\mathbf{B}$ of π_2 by two submappings. One is based on applying $\mathbf{A}$ to map onto π_1 the image coordinates of π_2. The other is based on using $\mathbf{C}$ to map the obtained points on π_1 to π_2. So we have

$$\rho\mathbf{B} = \mathbf{C}\mathbf{A} \tag{30}$$

where ρ is a scale factor. Thus the back-projection calibration satisfying the pinhole model can now be formulated as

$$\rho_1 \mathbf{x}_{w1} = \mathbf{A}\hat{\mathbf{u}}_1, \tag{31}$$
$$\rho_2 \mathbf{x}_{w2} = \mathbf{C}\mathbf{A}\hat{\mathbf{u}}_2, \tag{32}$$
$$s.t.\ C1 \sim C5,$$

which can be posed as a constrained optimization problem by minimizing the residual error of (31) and (32) with respect to $\mathbf{A}$ and $\mathbf{C}$, subject to the perspectivity constraints $C1 \sim C5$. Notice that instead of computing $\mathbf{A}$ and $\mathbf{B}$, we now turn to the computation of $\mathbf{A}$ and $\mathbf{C}$. This computation can be reduced to a unconstrained optimization problem [23], due to the linearity of the constraints. After this computation, $\mathbf{B}$ can be obtained from (30). We shall later refer to the projective mapping with the perspectivity constraints as the perspective mapping.

B. Obtaining Camera Physical Parameters

Defining the O_1 coordinate system as the world coordinate system, we can express the direction of the line of sight of any point $(X_1, Y_1) \in \pi_1$ in the world system, according to (18), as

$$\mathbf{v} = \begin{pmatrix} \frac{\mathbf{c}_1'\mathbf{x}_{w1}}{\mathbf{c}_3\mathbf{x}_{w1}} - X_1 \\ \frac{\mathbf{c}_2'\mathbf{x}_{w1}}{\mathbf{c}_3\mathbf{x}_{w1}} - Y_1 \\ \frac{\mathbf{c}_3'\mathbf{x}_{w1}}{\mathbf{c}_3\mathbf{x}_{w1}} \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{c}_1'\mathbf{x}_{w1} - X_1\mathbf{c}_3\mathbf{x}_{w1} \\ \mathbf{c}_2'\mathbf{x}_{w1} - Y_1\mathbf{c}_3\mathbf{x}_{w1} \\ \mathbf{c}_3'\mathbf{x}_{w1} \end{pmatrix} \tag{33}$$

where $\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{c}'_3$ are the row vectors of $\mathbf{C}'$; the symbol $\Rightarrow$ means taking out a common scale factor, which will not affect the direction of the vector. The vector $\mathbf{v}$ is called the *viewing vector* of (X_1, Y_1). From (33), it can be seen that the viewing vector is a quadratic function of X_1 and Y_1.

Lemma 3: Under the perspectivity conditions, we have the following.

a) The viewing vector can be expressed as a linear function of X_1 and Y_1 in terms of the viewing matrix $\mathbf{V}$ by

$$\mathbf{v} = \mathbf{V}\begin{pmatrix} X_1 \\ Y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} v_{11} & 0 & v_{13} \\ 0 & v_{11} & v_{23} \\ 0 & 0 & v_{33} \end{pmatrix}\begin{pmatrix} X_1 \\ Y_1 \\ 1 \end{pmatrix}. \quad (34)$$

When the two calibration planes are in ordinary relative positions, the elements of $\mathbf{V}$ are

$$v_{11} = -c_{32}/k_2,\ v_{13} = c'_{12}/k_2,$$
$$v_{23} = c'_{21}/k_1,\quad v_{33} = c'_{32}/k_2. \quad (35)$$

When the calibration planes are parallel to each other, the elements of $\mathbf{V}$ are

$$v_{11} = c'_{11} - c_{33}, v_{13} = c'_{13},$$
$$v_{23} = c'_{23}, v_{33} = \bar{t}_3 c_{33}. \quad (36)$$

b) The camera perspective center $(t'_1, t'_2, t'_3)^T$ can be computed as

$$\begin{pmatrix} t'_1 \\ t'_2 \\ t'_3 \end{pmatrix} = \begin{pmatrix} -\frac{v_{13}}{v_{11}} \\ -\frac{v_{23}}{v_{11}} \\ -\frac{v_{33}}{v_{11}} \end{pmatrix}. \quad (37)$$

The proof of this lemma is given in [24].

Except for the perspective center, the remaining physical camera parameters can not be computed in closed forms. Nevertheless, we still have their approximates through the following lemma.

Lemma 4: If lens distortion is not considered and the transformation parameters are computed by the 1st order perspective mapping, all the physical camera parameters can be computed directly from the transformation coefficients in closed forms.

Proof: See [24].

This lemma provides an alternative method to compute the camera physical parameters in the distortion-free case. One can also use the perspective transformation matrix to do the same job [7], [5].

C. The Projection Problem

The constraints of perspectivity can also be applied to the projection problem. Since we have obtained the mapping $\mathbf{C}$, we can use it to map the calibration points $\mathbf{x}_{w1}$ on plane π_1 to plane π_2 by

$$\rho' \mathbf{x}'_{w1} = \mathbf{C}\mathbf{x}_{w1} \quad (38)$$

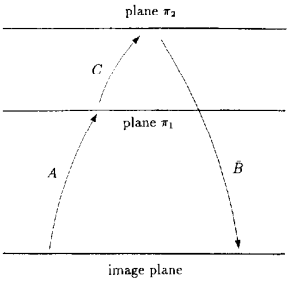

Fig. 3. The mapping procedure for both the back-projection and the projection processes.

where $\mathbf{x}'_{w1} = (X'_1, Y'_1, 1)^T$ is the obtained point, $\rho' = \mathbf{c}_3 \mathbf{x}_{w1}$. In this way, the projection of π_1 can be performed in terms of that of π_2 by

$$\bar{\rho}\mathbf{u}_1 = \bar{\mathbf{B}}\hat{\mathbf{x}}'_{w1} \quad (39)$$

where $\hat{\mathbf{x}}'_{w1}$ is computed from $\mathbf{x}'_{w1}$ in a similar way to (13). Thus, we can combine (39) and (14) to give a least square solution of $\bar{\mathbf{B}}$ using the set of calibration points. In terms of $\bar{\mathbf{B}}$, the prediction of image coordinates can be done in a similar way to that in Section II-B.

It can now be seen that by imposing the perspectivity constraints, we have replaced the mappings in both the back-projection and the projection processes by three mappings: $\mathbf{A}$, $\mathbf{C}$, and $\bar{\mathbf{B}}$. The whole mapping procedure is illustrated in Fig. 3.

In solving the set of equations (6), (8), (12), (14), or (31), (32), and (39), we used the Newton–Raphson method with descent rms errors [24]; with the initial guesses obtained from the linearly solved first order mappings. The details of the solution stages can be found in [24] and [23].

IV. Explicit Correction of Radial Lens Distortion

Under the radial distortion model, we can explicitly compute a camera's physical parameters by first correcting the distortion in image plane and then applying linear calibration methods.

A. Cross Ratio for Distortion Center

Assume there are four collinear points D_1, D_2, D_3, D_4 with known distances between each other on a line L in 3-D space. These points should be ideally projected to $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$, $\bar{p}_4$ in the camera CCD plane. Due to the radial lens distortion, the ideal image points are distorted along radial directions to form the distorted image points p_1, p_2, p_3, p_4. Suppose o is the distortion center. (Here we distinguish the distortion center from the image center due to a reason stated later.) Then, op_i is in the radial direction, and $\bar{p}_i$ should be on $op_i, i = 1 \sim 4$, see Fig. 4. Denote op_i by l_i, $i = 1, 2, 3, 4$. It can be easily seen that D_1, D_2, D_3, D_4 on L and l_1, l_2, l_3, l_4 through o

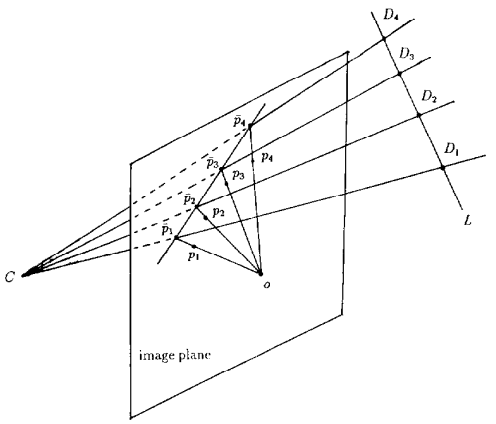

Fig. 4. Projective invariants for radial distortion correction.

form a projective mapping. That is,

$$L(D_1, D_2, D_3, D_4) \bar{\wedge} o(l_1, l_2, l_3, l_4)$$

where $\bar{\wedge}$ represents the projective mapping. We know that under projective mappings the cross ratio is invariant. Thus we have

$$(D_1, D_4; D_3, D_2) = (l_1, l_4; l_3, l_2). \tag{40}$$

Let the cross ratio of the world points be

$$(D_1, D_4; D_3, D_2) = \frac{D_1D_3 \cdot D_2D_4}{D_3D_4 \cdot D_1D_2} = c.$$

Equation (40) can thus be written as

$$c = \frac{\sin\theta_{13} \sin\theta_{24}}{\sin\theta_{34} \sin\theta_{12}} \tag{41}$$

where θ_{ij} is the angle between op_i and op_j. Suppose p_i has the CCD coordinates (x_i, y_i) defined with respect to the *distortion center*, and the frame buffer coordinates of p_i are (u_i, v_i), $i = 1, 2, 3, 4$. Suppose $(\tilde{u}_0, \tilde{v}_0)$ are the coordinates of the distortion center in the frame buffer. Then we have

$$\sin\theta_{ij} = \frac{|\mathbf{op}_i \times \mathbf{op}_j|}{||\mathbf{op}_i||\ ||\mathbf{op}_j||} \tag{42}$$

where

$$\mathbf{op}_i = (x_i, y_i) = ((u_i - \tilde{u}_0)d_x, (v_i - \tilde{v}_0)d_y),$$
$$|\mathbf{op}_i \times \mathbf{op}_j| = x_jy_i - x_iy_j.$$

Substituting (42) into (41), we obtain

$$c = \frac{f_{13}f_{24}}{f_{34}f_{12}} \tag{43}$$

where

$$f_{ij} = (u_j - \tilde{u}_0)(v_i - \tilde{v}_0) - (u_i - \tilde{u}_0)(v_j - \tilde{v}_0).$$

It can be easily seen that (43) is independent of the radial distortion coefficient. This property is similar to that of RAC [19]. Besides, (43) is also independent of the camera position and orientation, the pixel-to-pixel distances d_x, d_y, and the focal length. It is a nonlinear equation only on two variables, namely, the distortion center $(\tilde{u}_0, \tilde{v}_0)$. Theoretically, when given more than one collinear set of calibration points, we can solve for $(\tilde{u}_0, \tilde{v}_0)$. Define

$$F = c - \frac{f_{13}f_{24}}{f_{34}f_{12}}.$$

Using the local linearization about an initial guess and proceeding with iterative search, we can get the solution for $(\tilde{u}_0, \tilde{v}_0)$. Our experience indicates that taking the initial guess of $(\tilde{u}_0, \tilde{v}_0)$ to be the apparent center of the frame buffer will result in the correct solution.

B. Computing the Scale Factor and Distortion Coefficient by Collinearity

Since perspective projection preserves collinearity of collinear points, image points produced by collinear world points should be collinear, if the image distortion has been corrected.

When the distortion is radial, the image coordinates (u, v) in the frame buffer can be corrected to form the ideal coordinates in the CCD plane by simplifying (3) and (4) and then expressing them by u and v according to (1) as

$$\bar{x} = d_x((u-\tilde{u}_0)+(u-\tilde{u}_0)^3k_1'k^2+(u-\tilde{u}_0)(v-\tilde{v}_0)^2k_1'), \tag{44}$$

$$\bar{y} = d_y((v-\tilde{v}_0)+(u-\tilde{u}_0)^2(v-\tilde{v}_0)k_1'k^2+(v-\tilde{v}_0)^3k_1') \tag{45}$$

where $(\bar{x}, \bar{y})$ are defined with respect to the distortion center; $k_1' = k_1d_y^2$; $k = d_x/d_y$ is called the image plane scale factor.

Take the three points p_1, p_2, p_3 for illustration. The collinearity condition for the corrected points $\bar{p}_i = (\bar{x}_i, \bar{y}_i)$ of p_i, $i = 1, 2, 3$, can be written as

$$F = (\bar{x}_3 - \bar{x}_1)(\bar{y}_2 - \bar{y}_1) - (\bar{x}_2 - \bar{x}_1)(\bar{y}_3 - \bar{y}_1) = 0. \tag{46}$$

Inserting (44) and (45) into the above equation, we get a nonlinear equation on the equivalent distortion coefficient k_1' and the scale factor k. Note that the distortion coefficient k_1 can only be determined up to the scale factor d_y^2, but even when d_y is unknown, we can still perform the distortion correction in frame buffer, as will be seen later. The two variables k_1' and k can be solved for in a similar way as for $(\tilde{u}_0, \tilde{v}_0)$. The initial guess for k_1' can be taken to be 0 and that for k to be 1. A thing to note is that, when $k_1' = 0$, we get $\partial F/\partial k \equiv 0$, so that the linearization equation takes the form of

$$\left(\frac{\partial F}{\partial k_1'}, 0\right)|_{M_0}\begin{pmatrix}\Delta k_1' \\ \Delta k\end{pmatrix} = -F|_{M_0} \tag{47}$$

where $M_0 = (0, 1)$ is the initial guess for (k_1', k). Since (47) will cause the coefficient matrix in the normal equations singular, therefore in the first iteration step, we have to make a special treatment by keeping k unchanged and adjusting only k_1' through (47).

After having calibrated the distortion center, the distortion coefficient and the image scale factor, we can now compensate for the radial distortion in the frame buffer. From (44) and (45), we can write the correction equation as

$$\bar{u} = u + (u - \tilde{u}_0)^3k_1'k^2 + (u - \tilde{u}_0)(v - \tilde{v}_0)^2k_1', \tag{48}$$
$$\bar{v} = v + (u - \tilde{u}_0)^2(v - \tilde{v}_0)k_1'k^2 + (v - \tilde{v}_0)^3k_1' \tag{49}$$

where $(\bar{u}, \bar{v})$ are the ideal coordinates of (u, v) in the frame buffer. Later, we shall call the distortion center, the distortion coefficient and the image scale factor as a whole the *correction parameters*. After the above correction, we can use any conventional *linear* camera calibration techniques [1], [5], [9], [21], [27], (including the method in Lemma 4) to obtain the camera physical parameters without considering further lens distortions.

Since the calibration of the correction parameters above relies only on the distortion effect, for lenses with little distortion, the computed distortion center will have a large uncertainty. As an extreme example, when a lens does not have any distortion at all, the distortion center can be anywhere in the image plane without violating the cross ratio invariance. It is also intuitively true that, if there is no distortion, there will be no center of distortion. That is why we do not view the distortion center as the image center. But since the computed distortion coefficient in the no distortion case would be zero, the correction made by (48) and (49) will not change any image coordinates for *any* distortion center. In view of this, we can see that, if a lens has little distortion, the correction process will be insensitive to the uncertainties in the distortion center. On the other hand, if a lens has a large distortion, the uncertainties in the computed distortion center will be small, so that the correction is more reliable. As a conclusion, we can say that the correction process is as a whole reliable.

V. Experimental Results

A. The Set-Up

Two AQUA-TV CCD cameras with normal lenses were mounted on an ordinary z-raiser. The base line and vergence angle of the two cameras are about 200 mm and 20°, respectively. The focal length and the diagonal visual angle of the two lenses are 16 mm and about 30°, respectively. The z-raiser can be manually moved vertically by a screwing gear. The amount of movement can be read within 0.1 mm. A calibration pattern produced by a laser printer contains many black circles of diameter 5 mm, with adjacent circles spaced 20 mm apart in two orthogonal directions on a white background. This calibration pattern was fixed beneath the cameras and was made large enough such that the two cameras in different positions can always "see" some calibration points near the image border, where larger amounts of distortion occur. Due to this arrangement, the number of calibration points acquired in different camera positions is different. The z-raiser were moved to 15 different positions to simulate the existence of 15 world planes. Since the direction of the movement is perpendicular to the calibration pattern, the 15 world planes can be regarded as being spaced in the world Z-direction by the amounts of movement. The origin of the world coordinate system is chosen such that the Z-coordinates of the world planes are approximately their distances to the cameras. The Z-coordinates of the world planes thus chosen are 36 cm, 40 cm, 44 cm, 46 cm, 48 cm, 50 cm, 52 cm, 54 cm, 56 cm, 58 cm, 60 cm, 62 cm, 65 cm, 68 cm, and 71 cm, respectively. Each of the world planes is named according to its Z-coordinate. For example, the plane with a Z-coordinate of 36 cm is named Dis36. The 30 images taken by the two cameras in the 15 positions were processed first by the DRF edge detector [18] (with gradient thresholds interactively selected by displaying the gradient histogram), and then by edge thinning and tracking to produce the boundary of each circle. A contour-based method [20] was used to get the center of gravity of each circle to serve as the calibration points.

B. Implicit Methods

We tried various combinations of two world planes from the 15 planes to serve as the calibration planes. For each calibration, we made 3-D reconstruction for all the image points of the 15 planes. The distance between the reconstructed position and the true 3-D position was calculated as an assessment of the quality of the calibration. To reflect the statistical behavior of the reconstruction process for each world plane, the average error, standard deviation, and maximum error were calculated for each world plane. These measurement errors were plotted as a function of the plane's Z-coordinate. The results have been reported in [23]. The following are the major conclusions: a) The second- order mappings outperform the first- order mappings. This means that the lens distortion can be well accounted for by the second-order mappings; b) The adding of the perspectivity constraints improves the reconstruction accuracy, especially for the planes that are distant from the calibration planes.

C. Explicit Methods

In the calibration of the distortion parameters by our explicit method, the same calibration data was used. Because of the dependence of our method on the distortion effect, we chose only those world points whose image points are near the image border as the calibration points. The world planes near the cameras were not used either, since they contain fewer world points. In forming the four-tuple points from each row and column of the collinear calibration points, the neighboring points within each four-tuple should be selected as distant as possible. This makes our algorithm less sensitive to noise in image data. For each of the ten planes from Dis50 to Dis71, the distortion center, the distortion coefficient and the scale factor were computed. Table I lists the statistics of the ten results (the third and fourth rows). The image centers computed by Tsai's 2-D RAC's (the RAC's for coplanar points) are also listed in the table (the sixth and seventh rows). Due to the weak distortion effect of our lenses (about eight pixels of distortion), the distortion center is not stable, as has been pointed out, whereas the distortion coefficient and the scale factor are relatively more stable. When using each of the ten sets of correction parameters to correct the frame buffer, the largest difference is 2.0 pixels. Despite this variance, the accuracy of 3-D reconstruction using different correction parameters followed by a linear calibration method is almost the same. To illustrate this, we purposely select the set of correction parameters having the largest biases (Dis58) from their averages to correct all the 15 images, and then use two of the corrected images to calibrate the cameras by the linear

TABLE I
THE STATISTICS OF THE TEN SETS OF CORRECTION PARAMETERS AND TSAI'S IMAGE CENTERS

Methods		left camera				right camera			
		$\hat{u}_0$	$\hat{v}_0$	k_1'	k	$\hat{u}_0$	$\hat{v}_0$	k_1'	k
Explicit method	average	272.5	236.1	1.330e-7	1.525	239.6	230.8	1.348e-7	1.505
	deviation	16.2	4.3	7.199e-9	0.052	11.7	3.3	5.876e-9	0.046
	10 planes	260.7	233.7	1.291e-7	1.546	243.6	229.2	1.348e-7	1.506
Tsai's 2D-RAC	average	271.2	235.7	–	–	235.2	231.2	–	–
	deviation	7.3	2.8	–	–	4.62	2.43	–	–

TABLE II
THE CAMERA PHYSICAL PARAMETERS USING Dis62 AND Dis71 AS CALIBRATION PLANES

Methods		Rotation (°)	Translation (mm)	Intrinsic parameters			
		roll-pitch-yaw	t_1, t_2, t_3	u_0, v_0	f_x, f_y	k_1'	k
left camera	M-matrix	-1.01, 10.22, 0.38	-277.4, -101.7, 74.5	255.6, 255.1	922.4, 1420.5	–	1.540
	C'M-matrix	-1.14, 8.79, -0.39	-279.0, -100.9, 65.7	278.4, 236.2	929.7, 1428.1	–	1.536
	Classic	-1.13, 9.79, -0.28	-277.2, -100.8, 70.6	261.3, 238.8	929.7, 1428.3	1.307e-7	1.536
	10C'M-matrix	-1.16, 10.14, -0.48	-277.4, -100.8, 72.4	256.5, 234.7	929.5, 1427.8	–	1.536
	Tsai	-1.18, 8.88, -0.85	-278.9, -100.6, 64.4	276.8, 229.6	926.6, 1423.0	1.394e-7	1.536
right camera	M-matrix	-1.13, -7.37, 0.76	-97.0, -100.6, 7.9	247.8, 240.4	921.7, 1420.1	–	1.541
	C'M-matrix	-1.07, -7.49, 0.23	-96.9, -100.7, 9.3	249.4, 227.8	934.7, 1436.8	–	1.537
	Classic	-1.06, -7.12, 0.14	-95.8, -100.5, 9.6	241.7, 225.3	934.7, 1436.4	1.332e-7	1.537
	10C'M-matrix	-1.07, -7.32, 0.22	-96.9, -100.7, 9.7	246.5, 227.6	935.2, 1437.2	–	1.537
	Tsai	-1.13, -6.50, 0.34	-96.8, -100.8, 9.5	233.5, 231.0	933.4, 1434.5	1.364e-7	1.537

perspective transformation matrix method [5]. The calibration planes chosen were Dis62 and Dis71.

Fig. 5 shows the reconstruction errors for the following three methods:

M-matrix: the linear method based on the perspective transformation matrix, in which no lens distortion can be considered;

C'M-matrix: the method of M-matrix with prior radial distortion correction using our explicit method;

Classic: the classic photogrammetric method [25] using full-scale nonlinear search, considering only radial distortion; (Considering other distortions makes the iteration not converge, see [24])

From Fig. 5, we can see that using the same linear calibration method, the accuracy improvement obtained through our explicit distortion correction is significant. The maximum reduction in the maximum error is 3.29 mm. The achieved accuracy is comparable to that of the full-scale nonlinear search method.

Table II lists the camera physical parameters obtained by the methods of M-matrix, C'M-matrix, Classic, 10C'M-matrix (using the ten planes together for computing the correction parameters), and Tsai's method. The correction parameters obtained by using all the ten planes are listed in Table I. Notice that using the ten planes together is not the same as using multiple planes for complete camera calibration, since the relative positions and orientations between the 10 planes (and the calibration lines) are assumed *unknown*. It can be seen from Table II that the camera parameters produced by our method are very close to the classic method.

Since the measurement accuracy of our explicit method is similar to that of the classic method, in the following we will concentrate on the comparison of three methods: our implicit method, the classic method, and Tsai's method.

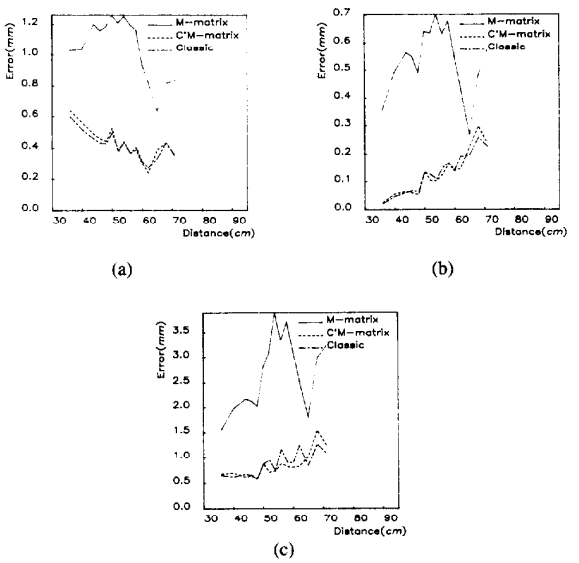

Fig. 5. Absolute reconstruction errors for the explicit methods using Dis62 and Dis71 as calibration planes. (a) average error; (b) standard deviation; (c) maximum error.

D. Some Cross Comparison

First, a short description about the implementation of Tsai's method is in need. We know from Table I that the method of 2-D RAC does not give the same image center for different calibration planes. Thus, when using two planes for camera calibration, we think that the method of 3-D RAC (the RAC for non-coplanar points) gives a better least square solution than just taking the average of the two 2-D RAC's. For this reason, we chose the method based on the 3-D RAC in the comparison. In practice, we found that the image

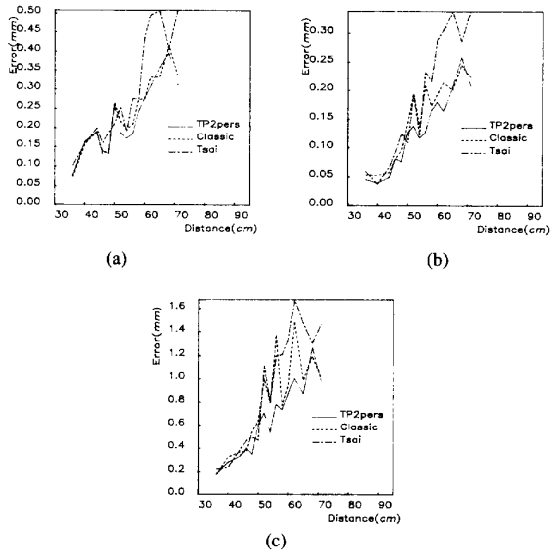

Fig. 6. Absolute reconstruction errors for both the implicit and the explicit methods using Dis36 and Dis71 as calibration planes. (a) average error; (b) standard deviation; (c) maximum error.

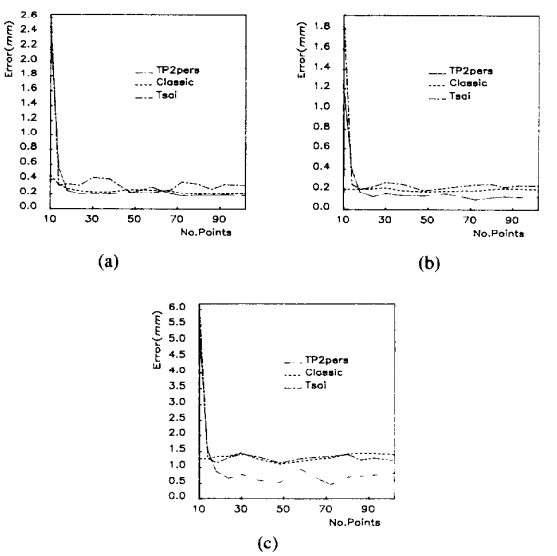

Fig. 7. Absolute reconstruction errors of the test plane Dis56 as a function of the number of calibration points when using Dis36 and Dis71 as calibration planes. (a) average error; (b) standard deviation; (c) maximum error.

center obtained by averaging that of two 2-D RAC's does not differ from that of the 3-D RAC by more than six pixels for our lenses. Therefore, the 2-D RAC is still appropriate for determining the image center. However, the use of 2-D RAC's to determine other camera parameters would need the knowledge about the pixel-to-pixel distances in the CCD plane [19]. Besides, a uncertainty scale factor has to be calibrated separately [17], [16].

Absolute 3-D Error: Fig. 6 shows the absolute 3-D reconstruction errors of our implicit method, the classic method and Tsai's method when using Dis36 and Dis71 as the calibration planes, where TP2pers means the two plane method of second order mapping with perspectivity constraints. The conclusions reached are summarized below.

1) Our implicit method outperforms the explicit methods, especially in the aspects of standard deviation and maximum error. The maximum reductions in the maximum error achieved by our method with respect to the classic and Tsai's method are 0.59 mm and 0.67 mm for this choice of the calibration planes.
2) The classic method outperforms Tsai's method. A detailed analysis of the reason can be found in [23].
3) The measurement accuracy of the three methods near the cameras is almost the same, especially when the chosen calibration planes are also near the cameras. That's because 3-D reconstruction near the cameras is less sensitive to the noise in both the image data and the calibrated parameters. In Tsai's experiment [19], the depth value was 4 inches (about 10 cm), and high accuracy was reported.

It was also found that the major difference of the three methods lies in the depth errors, and for all the three methods the depth errors are larger than the lateral errors.

Among the different choices of the world planes as the calibration planes, we found that using Dis36 and Dis71 as the calibration planes achieved the best accuracy for all the three methods. Thus, in what follows, our comparison will be based on using these two calibration planes. But the conclusions achieved also apply to other cases of calibration planes.

The Influence of the Number of Calibration Points: In the calibration above, all the points on the calibration planes were used. The use of more than the minimum number of calibration points can smooth noise. By a simple counting of unknowns, we know that the three methods of TP2pers, Classic and Tsai need at least 10, 6, and 9 calibration points, respectively. By evenly re-sampling the calibration points, we can study the effect of the number of calibration points on the 3-D measurement. Fig. 7 shows the reconstruction error of a middle plane Dis56 as a function of the number of calibration points. From the figure, we can see that the classic method is less affected by fewer numbers of calibration points. After the number of calibration points has reached 18, all the three methods do not give any further accuracy improvement. This suggests a significant computational saving by using fewer calibration points regularly distributed on the image plane.

Relative 3-D Error: Relative 3-D error means the error in the dimensional measurement like the distance between reconstructed points. By measuring the distance between adjacent points on each world plane, we get a relative error distribution in the lateral direction. By measuring the distance between neighboring points on adjacent planes (the points with the same X, Y coordinates but different Z coordinates), we get the distribution of relative errors in the depth direction. It was found that the relative errors for the three methods are almost the same. This is because the absolute errors for the *neighboring* points are quite correlated. Due to this correlation,

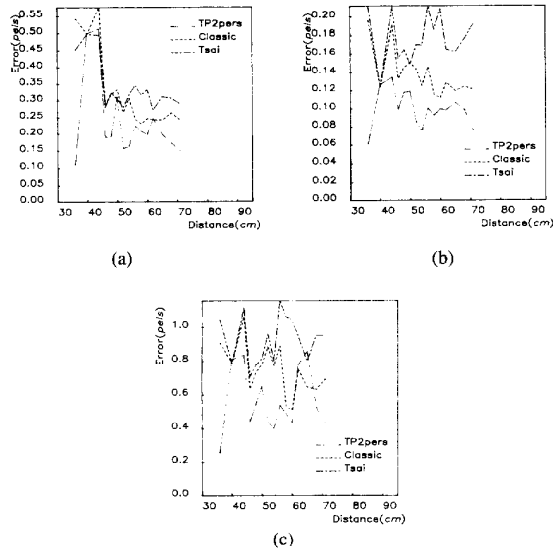

Fig. 8. Errors in generating image coordinates using Dis36 and Dis71 as calibration planes.(a) average error; (b) standard deviation; (c) maximum error.

the relative errors are also smaller than the absolute errors. This fact was ever noticed by Tsai [19]. Again, the relative error in the depth direction is larger than that in the lateral direction.

Errors in Generating Image Coordinates: The calibrated parameters using the explicit methods can also be used to generate ideal image coordinates from known world coordinates, but the computation of *distorted* image coordinates from ideal image coordinates does not have a closed form (see (2)–(4)), except for the radial distortion case [16]. As was already mentioned, we consider only radial distortion in the Classic method, and in Tsai's method, only radial distortion can be considered. For our implicit method, the projection should be specially calibrated. After the calibration, all the world points on the 15 world planes were projected onto the image plane. The generated image coordinates were compared with the actual image coordinates. The errors (in pixels) for our method, the classic method, and Tsai's method are depicted in Fig. 8, from which the accuracy improvement achieved by our method can be clearly seen. The maximum amount of error reductions in the maximum error of our method with respect to the classic and Tsai's methods are, 0.76 pixels and 1.16 pixels, respectively.

A word about the use of the perspectivity constraints in the projection calibration of our implicit method is that the addition of the constraints reduces the maximum error by 0.33 pixels, as compared with that without the constraints.

VI. Conclusion

Experiments have shown that our implicit method gives better accuracy in both the 3-D measurement and the generation of image coordinates. Adding the perspectivity constraints further improves the accuracy. The perspectivity conditions also reduces the minimum number of calibration points required. Computationally, our method is more intensive than the explicit methods due to the more unknown variables involved. The investigation of the effect of the number of calibration points illustrates that much computation can be saved by using fewer calibration points.

Our explicit distortion correction method uses only one-dimensional calibration points, and involves only two sequential two-dimensional search processes. No computation of the camera position and orientation is needed. By combining with other linear calibration methods, the method achieves a reconstruction accuracy that is comparable to that obtained through full-scale nonlinear search methods. A future direction of research is to investigate the possibility of explicitly correcting other types of lens distortions.

Appendix A
Proof of Lemma 2 for the Perspectivity Conditions of Nonparallel Planes

Lemma 2: A set of necessary and sufficient conditions for $\mathcal{P}_{12}$ to be a perspectivity is $C1 \sim C5$ in (20)$\sim$(24).

Proof: The necessity part: By Lemma 1 and (18), we require that $\forall (X_1, Y_1) \in l$:

$$Z_2' = \frac{c_{31}'X_1 + c_{32}'Y_1 + c_{33}'}{c_{31}X_1 + c_{32}Y_1 + c_{33}} \equiv 0, \tag{50}$$

$$X_2' = \frac{c_{11}'X_1 + c_{12}'Y_1 + c_{13}'}{c_{31}X_1 + c_{32}Y_1 + c_{33}} \equiv X_1, \tag{51}$$

$$Y_2' = \frac{c_{21}'X_1 + c_{22}'Y_1 + c_{23}'}{c_{31}X_1 + c_{32}Y_1 + c_{33}} \equiv Y_1. \tag{52}$$

Equation (50) ensures the intersection line of the two calibration planes line-invariant. Equation (51) or (52) ensures point-wise invariant. Since there is only one degree of freedom on the intersection line, (51) and (52) are equivalent.

By (50) and (19), we get $C1, C2$, and

$$C6 : k_1c_{33}' - k_3c_{31}' = 0.$$

The constraint $C6$ is clearly dependent on $C1$ and $C2$.

Next, we solve for Y_1 from (51), and X_1 from (52) as

$$Y_1 = \frac{-c_{31}X_1^2 + (c_{11}' - c_{33})X_1 + c_{13}'}{c_{32}X_1 - c_{12}'},$$

$$X_1 = \frac{-c_{32}Y_1^2 + (c_{22}' - c_{33})Y_1 + c_{23}'}{c_{31}Y_1 - c_{21}'}.$$

Substituting each of them into (19) respectively, we obtain

$$\forall X_1 : (k_1c_{32} - k_2c_{31})X_1^2 + (k_2(c_{11}' - c_{33}) - k_1c_{12}' + k_3c_{32})X_1 + k_2c_{13}' - k_3c_{12}' \equiv 0, \tag{53}$$

$$\forall Y_1 : (k_2c_{31} - k_1c_{32})Y_1^2. + (k_1(c'_{22} - c_{33}) + k_3c_{31} - k_2c'_{21})Y_1 + k_1c'_{23} - k_3c'_{21} \equiv 0. \quad (54)$$

Let the coefficients of each polynomial equal zero. We get $C3, C4, C5$ and

$$C7 : k_1(c'_{22} - c_{33}) + k_3c_{31} - k_2c'_{21} = 0, \quad (55)$$
$$C8 : k_1c'_{23} - k_3c'_{21} = 0. \quad (56)$$

The constraints $C6 \sim C8$ are redundant. The number of independent constraints is 5.

The sufficiency part: In the following derivation, the characters standing above the equation symbol "=" indicate the underlying conditions which lead to the subsequent quantities.

For each $(X_1, Y_1) \in l$, we have from (18)

$$\begin{aligned} X'_2 &= \frac{k_2(c'_{11}X_1 + c'_{12}Y_1 + c'_{13})}{k_2(c_{31}X_1 + c_{32}Y_1 + c_{33})} \\ &\overset{C5,C3}{=} \frac{k_2c'_{11}X_1 + k_2c'_{12}Y_1 + k_3c'_{12}}{k_1c_{32}X_1 + k_2c_{32}Y_1 + k_2c_{33}} \\ &= \frac{k_2c'_{11}X_1 + c'_{12}(k_1X_1 + k_2Y_1 + k_3) - c'_{12}k_1X_1}{c_{32}(k_1X_1 + k_2Y_1 + k_3) - c_{32}k_3 + k_2c_{33}} \\ &\overset{l}{=} \frac{(k_2c'_{11} - k_1c'_{12})X_1}{k_2c_{33} - k_3c_{32}} \overset{C4}{=} X_1, \end{aligned} \quad (57)$$

$$\begin{aligned} Y'_2 &= \frac{k_1(c'_{21}X_1 + c'_{22}Y_1 + c'_{23})}{k_1(c_{31}X_1 + c_{32}Y_1 + c_{33})} \\ &\overset{C8,C3}{=} \frac{k_1c'_{21}X_1 + k_1c'_{22}Y_1 + k_3c'_{21}}{k_1c_{31}X_1 + k_2c_{31}Y_1 + k_1c_{33}} \\ &= \frac{(k_1c'_{22} - k_2c'_{21})Y_1 + c'_{21}(k_1X_1 + k_2Y_1 + k_3)}{c_{31}(k_1X_1 + k_2Y_1 + k_3) - c_{31}k_3 + k_1c_{33}} \\ &\overset{l}{=} \frac{(k_1c'_{22} - k_2c'_{21})Y_1}{k_1c_{33} - k_3c_{31}} \overset{C7}{=} Y_1, \end{aligned} \quad (58)$$

$$\begin{aligned} Z'_2 &= (k_3c'_{31}X_1 + k_3c'_{32}Y_1 + k_3c'_{33})/k_3 \\ &\overset{C6,C2}{=} (k_1c'_{33}X_1 + k_2c'_{33}Y_1 + k_3c'_{33})/k_3 \overset{l}{=} 0. \end{aligned} \quad (59)$$

This means that the common line l of π_1 and π_2 is point-wise invariant. By Lemma 1, we know that $\mathcal{P}_{12}$ is a perspectivity.

Appendix B
Proof of Lemma 2′ for the Perspectivity Conditions of Parallel Planes

When the two calibration planes are parallel to each other, their common line becomes a degenerate line, i.e., a line at infinity. Even in this case, Lemma 1 still holds. Thus we require in (18): .

$$\forall (X_1, Y_1) \to (\infty, \infty) : (X'_2, Y'_2) \to (\infty, \infty), \quad (60)$$

$$\forall k : \lim_{(X_1,Y_1)\to(\infty,\infty)} X'_2/Y'_2 \equiv \lim_{(X_1,Y_1)\to(\infty,\infty)} X_1/Y_1 = k. \quad (61)$$

Equation (60) ensures line-invariant, while (61) ensures point-wise invariant. From (60) and (18), we get $C1p$ and $C2p$.

Using (61), we get

$$\begin{aligned} &\lim_{(X_1,Y_1)\to(\infty,\infty)} X'_2/Y'_2 \\ &= \lim_{(X_1,Y_1)\to(\infty,\infty)} \frac{c'_{11}X_1 + c'_{12}Y_1 + c'_{13}}{c'_{21}X_1 + c'_{22}Y_1 + c_{23'}} \\ &= \frac{c'_{11}k + c'_{12}}{c'_{21}k + c'_{22}} \equiv k, \quad \forall k, \end{aligned} \quad (62)$$

which leads to $C3p$, $C4p$ and $C5p$.

Acknowledgment

The authors are very grateful to Prof. G. Hirzinger, Dr. K. Arbter at the German Aerospace Research Establishment, Oberpfaffenhofen, for providing convenient experimental facilities and having some valuable discussions. The authors are also very thankful to the anonymous reviewers for some critical and helpful comments.

References

[1] Y. I. Abdel-Aziz and H. M. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry," *Symp.Close-Range Photogrammetry*, Univ. of Illinois at Urbana-Champaign, 1971, pp. 1–18.

[2] K. B. Atkinson Ed., *Developments in Close Range Photogrammetry*, London, 1980.

[3] H. S. M Coxeter, *Projective Geometry*, second ed. New York: Applied Science Publ., 1987, p. 35.

[4] I. W. Faig, "Calibration of close-range photogrammetric systems: Mathematical formulation," *Photogrammetric Eng. Remote Sensing*, vol. 41, pp. 1479–1486, 1975.

[5] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," *in Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*,1986, pp. 15–20.

[6] ——,"Camera calibration for 3-D computer vision," *in Pro. Int. Workshop Industrial Applicat. of Machine Vision and Machine Intell.*, Japan, 1987, pp. 240–247.

[7] S. Ganapathy, "Decomposition of transformation matrices for robot vision," *Proc. Int. Conf. Robotics*, 1984, pp. 130–139.

[8] L. E. Garner, *An outline of Projective Geometry*. Amsterdam: Elsevier North Holland, 1981, p. 51.

[9] W. I. Grosky and L. A. Tamburino, "A Unified approach to the linear camera calibration problem," *Proc. Int. Conf. Comput. Vision*, 1987, pp. 511–515.

[10] K. D. Gremban, C. E. Thorpe, and T. Kanade, "Geometric camera calibration using systems of linear equations," *in Proc. IEEE Int. Conf. on Robotics Automation* 1988, pp. 562–567.

[11] A. Izaquire, P. Pu, and J. Summers, "A new development in camera calibration calibrating a pair of mobile cameras," *in Proc. IEEE Int. Conf. on Robotics and Automat.*, 1985, pp. 74–79.

[12] K. Kanatani, *Group-Theoretical Methods in Image Understanding.*, New York: Springer-Verlag, 1990.

[13] Y. Lamdan, J. Schwartz, and H. Wolfson, "On recognition of 3-D objects from 2-D images," *in Proc. IEEE Int. Conf. Robotics and Automat.*, 1988, pp. 1407–1413.

[14] H. A. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration plane," *Comput. Graphics and Image Processing*, vol. 17, pp. 173–180, 1981.

[15] R. Mohr and L. Morin, "Relative positioning from geometric invariants," in *Proc. IEEE Conf. on Comput. Vision Pattern Recognit.*, Hawaii, June 1991, pp. 139–144.

[16] M. A. Penna, "Camera calibration: a quick and easy way to determine the scale factor," *IEEE Trans Pattern Anal. Machine Intell.*, vol. 12, no. 12, pp. 1240–1245, Dec. 1991.
[17] R. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 5, pp. 713–720, May 1988.
[18] J. Shen and S. Castan, "An optimal linear filter in edge detection," *in Proc. of the IEEE Conf. Comput. Vision Pattern Recognit.*, Miami, FL, June 22–26, 1986, pp. 104–109.
[19] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3-D Machine vision," *in Proc. IEEE Conf. Comput. Vision Pattern Recognit.*,1986, pp. 364–374.
[20] G.-Q. Wei, Z. He, and S. D. Ma, "Fast moment calculation: Discrete form, analytic form and parallel implementation," *Proc. Int. Conf. Comput. Aided Technol.* Hong Kong, 1988. pp. 331–340.
[21] ——,"Camera calibration by vanishing point and cross ratio," *in Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, pp. 1630–1633, May 1989, pp. 331–340.
[22] G. Q. Wei and S. D. Ma, "Two plane camera calibration: A unified model," *in Proc. IEEE Conf. on Comput. Vision and Pattern Recognit.*, Hawaii, June 1991, pp. 133–138.
[23] ——, "A complete two-plane camera calibration method and experimental comparisons," *Proc. 4th Intern. Conf. Comput. Vision*, Berlin, May 11–14, 1993, pp. 439–446.
[24] ——, "Implicit and explicit camera calibration: theory and experiments," *Tech. Rep.*, National Lab of Pattern Recognit., Chinese Academy of Sciences, P.R.C., 1992.
[25] J. Weng, P. Cohen, and M. Herniou, "Calibration of stereo cameras using a non-linear distortion model," *Proc. Int. Conf. Pattern Recognit.*, 1990, pp. 246–253.
[26] K. W. Wong, "Mathematical formulation and digital analysis in close-range photogrammetry," *Photogrammetric Eng. Remote Sensing*, vol. 41, pp. 1355–1373, 1975.
[27] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Comput. Graphics and Image Processing*, vol. 7, pp. 195–201, 1978.

Guo-Qing Wei (S'88–S'89–M'89–M'90) was born in October 1962 in Jiang-su Province, P.R.C. He received the B.S., M.S., and Ph.D degrees in electrical engineering in 1983, 1986, and 1989 respectively, all from Southeast University (formally Nanjing Institute of Technology), P.R.C.

From 1989 to 1991, he was an Assistant Professor at the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Since 1991, he has been a guest scientist (postdoctor) at the Robot System Group, Institute of Robotics and System Dynamics, German Aerospace Research Establishment, Oberphaffenhofen. His areas of interest include computer vision and neural networks.

Dr. Wei is a member of IEEE Computer Society and a member of the New York Academy of Sciences..

Song De Ma (SM'91) was born in Shanghai, P.R.C. on July 20, 1946. He received the B.S. degree in Automation from Hsinghua University in 1968, and the degrees of Docteur de 3^{ieme} cycle and of Docteur es Sciences d'Etat from University of Paris 6 in 1983 and 1986, respectively.

From 1983 to 1984, he was an invited researcher in the Computer Vision Laboratory, University of Maryland. From 1984 to 1986, he was an invited researcher in the Robotic Vision Group of INRIA in France. Since 1986, he has been a professor at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His current research interests include 3-D computer vision, neural computing, realistic image synthesis, and sensor based robot control.