



PERGAMON

Pattern Recognition 35 (2002) 1617–1635

PATTERN  
RECOGNITION

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

# A comparative review of camera calibrating methods with accuracy evaluation<sup>☆</sup>

Joaquim Salvi<sup>\*</sup>, Xavier Armangué, Joan Batlle

*Computer Vision and Robotics Group, Institute of Informatics and Applications, University of Girona,  
Avda Lluís Santaló, s/n, E-17071, Girona, Spain*

Received 16 June 2000; accepted 14 June 2001

## Abstract

Camera calibrating is a crucial problem for further metric scene measurement. Many techniques and some studies concerning calibration have been presented in the last few years. However, it is still difficult to go into details of a determined calibrating technique and compare its accuracy with respect to other methods. Principally, this problem emerges from the lack of a standardized notation and the existence of various methods of accuracy evaluation to choose from. This article presents a detailed review of some of the most used calibrating techniques in which the principal idea has been to present them all with the same notation. Furthermore, the techniques surveyed have been tested and their accuracy evaluated. Comparative results are shown and discussed in the article. Moreover, code and results are available in internet. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

*Keywords:* Camera calibration; Lens distortion; Parameter estimation; Optimization; Camera modelling; Accuracy evaluation; 3D reconstruction; Computer vision

## 1. Introduction

Camera calibration is the first step towards computational computer vision. Although some information concerning the measuring of scenes can be obtained by using uncalibrated cameras [1], calibration is essential when metric information is required. The use of precisely calibrated cameras makes the measurement of distances in a real world from their projections on the image plane possible [2,3]. Some applications of this capability include:

1. *Dense reconstruction:* Each image point determines an optical ray passing through the focal point of the camera towards the scene. The use of more than a single view of a motionless scene (taken from a stereo-

scopic system, a single moving camera, or even a structured light emitter) permits crossing both optical rays to get the metric position of the 3D points [4–6]. Obviously, the correspondence problem has to be previously solved [7].

2. *Visual inspection:* Once a dense reconstruction of a measuring object is obtained, the reconstructed object can be compared with a stored model in order to detect any manufacturing imperfections such as bumps, dents or cracks. One potential application is visual inspection for quality control. Computerized visual inspection allows automatic and exhaustive examination of products, as opposed to the slow human inspection which usually implies a statistical approach [8].
3. *Object localization:* When considering various image points from different objects, the relative position among these objects can be easily determined. This has many possible applications such as in industrial part assembly [9] and obstacle avoidance in robot navigation [10,11], among others.

<sup>☆</sup> This work has been supported by Spanish project CICYT TAP99-0443-CO5-01.

<sup>\*</sup> Corresponding author. Tel.: +34-72-41-8474; +34-72-41-8098.

*E-mail address:* qsalvi@eia.udg.es (J. Salvi).

4. *Camera localization*: When a camera is placed in the hand of a robot arm or on a mobile robot, the position and orientation of the camera can be computed by locating some known landmarks in the scene. If these measurements are stored, a temporal analysis allows the handler to determine the trajectory of the robot. This information can be used in robot control and path planning [12–14].

Camera calibration is divided into two phases. First, camera modelling deals with the mathematical approximation of the physical and optical behavior of the sensor by using a set of parameters. The second phase of camera calibration deals with the use of direct or iterative methods to estimate the values of these parameters. There are two kinds of parameters in the model which have to be considered. On the one hand, the intrinsic parameter set, which models the internal geometry and optical characteristics of the image sensor. Basically, intrinsic parameters determine how light is projected through the lens onto the image plane of the sensor. The other set of parameters are the extrinsic ones. The extrinsic parameters measure the position and orientation of the camera with respect to a world coordinate system, which, in turn, provides metric information with respect to a user-fixed coordinate system instead of the camera coordinate system.

Camera calibration can be classified according to several different criteria. For instance, (1) Linear versus non-linear camera calibration (usually differentiated depending on the modelling of lens distortion) [15]. (2) Intrinsic versus extrinsic camera calibration. Intrinsic calibration is concerned only with obtaining the physical and optical parameters of the camera [16,17]. Besides, extrinsic calibration concerns the measurement of the position and orientation of the camera in the scene [18,19]. (3) Implicit [20] versus explicit [21] calibration. Implicit calibration is the process of calibrating a camera without explicitly computing its physical parameters. Although, the results can be used for 3D measurement and the generation of image coordinates, they are useless for camera modelling as the obtained parameters do not correspond to the physical ones [22]. Finally, (4) the methods which use known 3D points as a calibrating pattern [23,24] or even a reduced set of 3D points [25,26], with respect to others which use geometrical properties in the scene such as vanishing lines [27] or other line features [28,29].

These different approaches can also be classified regarding the calibration method used to estimate the parameters of the camera model:

1. *Non-linear optimization techniques*. A calibrating technique becomes non-linear when any kind of lens imperfection is included in the camera model. In that case, the camera parameters are usually obtained through iteration with the constraint of minimizing a determined function. The minimizing function is

usually the distance between the imaged points and the modelled projections obtained by iterating. The advantage of these iterating techniques is that almost any model can be calibrated and accuracy usually increases by increasing the number of iterations up to convergence. However, these techniques require a good initial guess in order to guarantee convergence. Some examples are described in classic photogrammetry [30] and Salvi [31].

2. *Linear techniques which compute the transformation matrix*. These techniques use the least squares method to obtain a transformation matrix which relates 3D points with their 2D projections. The advantage here is the simplicity of the model which consists in a simple and rapid calibration. One drawback is that linear techniques are useless for lens distortion modelling, entailing a rough accuracy of the system. Moreover, it is sometimes difficult to extract the parameters from the matrix due to the implicit calibration used. Some references related to linear calibration can be found in Hall [20], Toscani-Faugeras [23,32] and Ito [15].
3. *Two-step techniques*. These techniques use a linear optimization to compute some of the parameters and, as a second step, the rest of the parameters are computed iteratively. These techniques permit a rapid calibration considerably reducing the number of iterations. Moreover, the convergence is nearly guaranteed due to the linear guess obtained in the first step. *Two-step techniques* make use of the advantages of the previously described methods. Some references are Tsai [24], Weng [33] and Wei [22].

This article is a detailed survey of some of the most frequently used calibrating techniques. The first technique was proposed by Hall in 1982 and is based on an implicit linear camera calibration by computing the  $3 \times 4$  transformation matrix which relates 3D object points with their 2D image projections [20]. The latter work of Faugeras, proposed in 1986, was based on extracting the physical parameters of the camera from such a transformation technique, thus it is explained as the second technique [23,32]. The following methods are based on non-linear explicit camera calibration, including the modelling of lens distortion. Hence, the first one is a simple adaptation of the Faugeras linear method with the aim of including radial lens distortion [31,34]. The widely used method proposed by Tsai, which is based on a *two-step technique* modelling only radial lens distortion, is also detailed [24]. Finally, the complete model of Weng, which was proposed in 1992, including three different types of lens distortion, is explained as the last technique [33]. Note that one of the principal problems to understand a calibrating technique in detail is the lack of notation standardization in mathematical equations and the use of different sets of coordinate systems. Both limitations complicate the comparing of techniques, thus a great deal

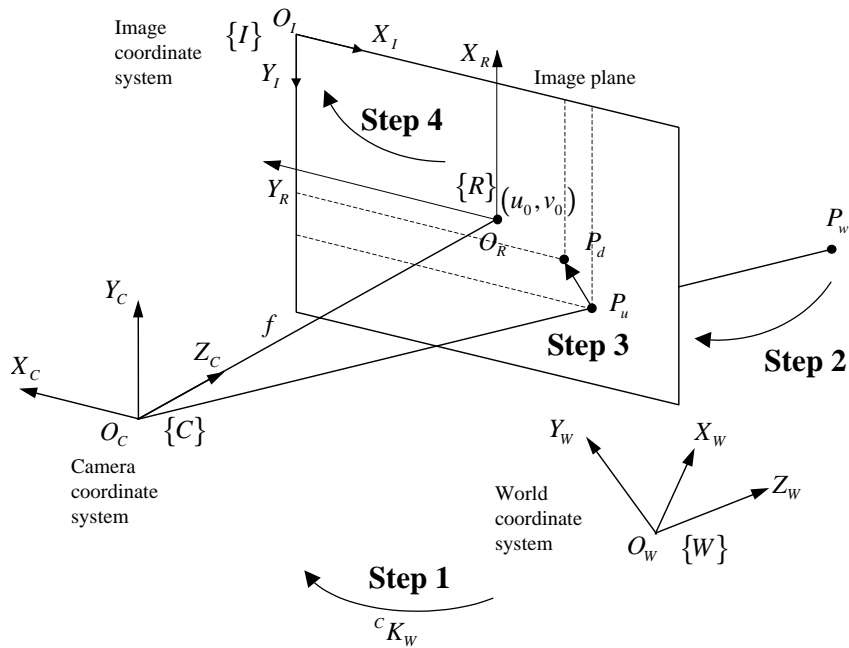


Fig. 1. The geometric relation between a 3D object point and its 2D image projection.

of effort has been made to present the survey using the same notation. All five techniques are explained herein and their 2D and 3D accuracy shown and discussed. A brief overview of camera accuracy evaluation [35] is included with the aim of using the same tools to compare different calibrating techniques implemented.

This article is structured as follows. Section 2 deals with camera modelling and how the camera model is gradually obtained by a sequence of geometrical transformations is explained. Section 3 describes the five different techniques of camera calibration, which estimate the parameters of the camera model. Then, a few methods for accuracy evaluation of camera calibrating techniques are explained in Section 4. Finally, both 2D and 3D accuracy of each calibration technique have been measured and their results are shown and compared. The paper ends with conclusions.

## 2. Camera model

A model is a mathematical formulation which approximates the behavior of any physical device by using a set of mathematical equations. Camera modelling is based on approximating the internal geometry along with the position and orientation of the camera in the scene. There are several camera models to choose from depending on the desired accuracy [15]. The simplest are based on linear transformations without modelling the lens distortion.

However, there are also some non-linear models which accurately model the lens. These are useful for some applications where greater precision is required.

The simplest model is the one proposed by Hall [20]. The goal is to find a linear relationship among the 3D points of the scene with their 2D projecting points on the image plane. This relationship is approximated by means of a transformation matrix,<sup>1</sup> as shown in the equation.

$$\begin{pmatrix} s^I X_d \\ s^I Y_d \\ s \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{pmatrix} \begin{pmatrix} {}^w X_w \\ {}^w Y_w \\ {}^w Z_w \\ 1 \end{pmatrix}. \quad (1)$$

Then, given a 3D point  $P_w$ , expressed with respect to the metric world coordinate system (i.e.  ${}^w P_w$ ), and applying the transformation matrix proposed by Hall, the 2D point  $P_d$  in pixels with respect to the image coordinate system is obtained, i.e.  ${}^I P_d = ({}^I X_d, {}^I Y_d)$ .

However, camera modelling is usually broken down into 4 steps, as is detailed in the following list (see also Fig. 1).

1. The first step consists of relating point  ${}^w P_w$  from the world coordinate system to the camera coordinate system, obtaining  ${}^C P_w$ . This transformation is performed by using a rotation matrix and a translation vector.

<sup>1</sup> The appendix at the end of the paper details the used nomenclature.

2. Next, it is necessary to carry out the projection of point  ${}^C P_w$  on the image plane obtaining point  ${}^C P_u$ , by using a projective transformation.
3. The third step models the lens distortion, based on a disparity with the real projection. Then, point  ${}^C P_u$  is transformed to the real projection of  ${}^C P_d$  (which should coincide with the points captured by the camera).
4. Finally, the last step consists of carrying out another coordinate system transformation in order to change from the metric coordinate system of the camera to the image coordinate system of the computer in pixels, obtaining  ${}^I P_d$ .

In the following, the different camera models of Faugeras–Toscani [32], Faugeras–Toscani with distortion [34], Tsai [24] and Weng [33] are explained in detail with attention on how they carry out the above four steps.

2.1. Changing from the world to the camera coordinate system

Changing the world coordinate system to the camera coordinate system is carried out in the same way in all the surveyed models. This transformation is modelled using a translation vector and a rotation matrix, as shown in the equation.

$$\begin{pmatrix} {}^C X_w \\ {}^C Y_w \\ {}^C Z_w \end{pmatrix} = {}^C R_W \begin{pmatrix} {}^W X_w \\ {}^W Y_w \\ {}^W Z_w \end{pmatrix} + {}^C T_W. \tag{2}$$

Then, given a point  ${}^W P_w$  related to the world coordinate system, and applying Eq. (2), the point  ${}^C P_w$  in relation to the camera coordinate system is obtained. Note that  ${}^C R_W$  expresses the orientation of the world coordinate system  $\{W\}$  with respect to the axis of the camera coordinate system  $\{C\}$ , and that  ${}^C T_W$  expresses the position of the origin of the world coordinate system measured with respect to  $\{C\}$ .

2.2. Projection of the 3D point on the image plane

Consider that any optical sensor can be modelled as a pinhole camera [2]. That is, the image plane is located at a distance  $f$  from the optical center  $O_C$ , and is parallel to the plane defined by the coordinate axis  $X_C$  and  $Y_C$ . Moreover, given an object point ( ${}^C P_w$ ) related to the camera coordinate system, if it is projected through the focal point ( $O_C$ ), the optical ray intercepts the image plane at the 2D image point ( ${}^C P_u$ ). This relation is shown in the equation.

$${}^C X_u = f \frac{{}^C X_w}{{}^C Z_w}, \quad {}^C Y_u = f \frac{{}^C Y_w}{{}^C Z_w}. \tag{3}$$

All the various models reviewed solved the projective transformation by using the same Eq. (3).

2.3. Lens distortion

The third step is based on modelling the distortion of the lenses. However, each model surveyed required a different approach. Eqs. (4) transform the undistorted point  ${}^C P_u$  to the distorted point  ${}^C P_d$ , where  $\delta_x$  and  $\delta_y$  represent the distortion involved.

$${}^C X_u = {}^C X_d + \delta_x, \quad {}^C Y_u = {}^C Y_d + \delta_y. \tag{4}$$

The camera model proposed by Faugeras and Toscani [32] does not model the lens distortion, therefore,  ${}^C P_u$  and  ${}^C P_d$  are the same point. In this case  $\delta_x$  and  $\delta_y$  are zero, as shown in the equations.

$$\delta_x = 0, \quad \delta_y = 0. \tag{5}$$

The Faugeras–Toscani model however can be improved by modelling the radial lens distortion [34]. Tsai [24] has modelled distortion in the same way. As shown in Eqs. (6),  $\delta_x$  and  $\delta_y$  represent the radial distortion [30]. This type of distortion is mainly caused by flawed radial curvature of the lens. See also [33].

$$\delta_x = \delta_{xr}, \quad \delta_y = \delta_{yr}. \tag{6}$$

The displacement given by the radial distortion  $d_r$  can be modelled by Eqs. (7), which consider only  $k_1$  the first term of the radial distortion series. It has been proved that the first term of this series is sufficient to model the radial distortion in most of the applications [24].

$$\delta_{xr} = k_1 {}^C X_d ({}^C X_d^2 + {}^C Y_d^2), \quad \delta_{yr} = k_1 {}^C Y_d ({}^C X_d^2 + {}^C Y_d^2). \tag{7}$$

The model of Weng [33] considers three types of distortion: radial distortion, decentering distortion and thin prism distortion. The total distortion will be the sum of these three distortions.

$$\delta_x = \delta_{xr} + \delta_{xd} + \delta_{xp}, \quad \delta_y = \delta_{yr} + \delta_{yd} + \delta_{yp}. \tag{8}$$

However, Weng proposed to model the lens distortion from the undistorted image point ( ${}^C X_u, {}^C Y_u$ ) instead of the distorted one ( ${}^C X_d, {}^C Y_d$ ). Although both approaches can be considered, it also has to be taken into account that the calibrating parameters will be different. Hence, Eqs. (4) have to be substituted by the equations.

$${}^C X_d = {}^C X_u + \delta_x, \quad {}^C Y_d = {}^C Y_u + \delta_y. \tag{9}$$

The radial distortion is modelled in the same manner Tsai proposed, except that Weng used the undistorted points.

$$\delta_{xr} = k_1 {}^C X_u ({}^C X_u^2 + {}^C Y_u^2), \quad \delta_{yr} = k_1 {}^C Y_u ({}^C X_u^2 + {}^C Y_u^2). \tag{10}$$

The decentering distortion is due to the fact that the optical center of the lens is not correctly aligned with

the center of the camera [33]. This type of distortion introduces a radial and tangential distortion [30], which can be described by the following equations:

$$\begin{aligned} \delta_{xd} &= p_1(3 {}^C X_u^2 + {}^C Y_u^2) + 2p_2 {}^C X_u {}^C Y_u, \\ \delta_{yd} &= 2p_1 {}^C X_u {}^C Y_u + p_2({}^C X_u^2 + 3 {}^C Y_u^2). \end{aligned} \quad (11)$$

The *thin prism distortion* arises from imperfection in lens design and manufacturing as well as camera assembly. This type of distortion can be modelled by adding a thin prism to the optic system, causing radial and tangential distortions [33]. This distortion is modelled by

$$\delta_{xp} = s_1({}^C X_u^2 + {}^C Y_u^2), \quad \delta_{yp} = s_2({}^C X_u^2 + {}^C Y_u^2). \quad (12)$$

By adding the three Eqs. (7), (11) and (12), and carrying out the following variable replacement:  $g_1 = s_1 + p_1$ ,  $g_2 = s_2 + p_2$ ,  $g_3 = 2p_1$  and  $g_4 = 2p_2$ , the following equations are obtained:

$$\begin{aligned} \delta_x &= (g_1 + g_3) {}^C X_u^2 + g_4 {}^C X_u {}^C Y_u + g_1 {}^C Y_u^2 \\ &\quad + k_1 {}^C X_u({}^C X_u^2 + {}^C Y_u^2), \\ \delta_y &= g_2 {}^C X_u^2 + g_3 {}^C X_u {}^C Y_u + (g_2 + g_4) {}^C Y_u^2 \\ &\quad + k_1 {}^C Y_u({}^C X_u^2 + {}^C Y_u^2). \end{aligned} \quad (13)$$

#### 2.4. Changing from the camera image to the computer image coordinate system

This final step deals with expressing the  ${}^C P_d$  point with respect to the computer image plane in pixels  $\{I\}$ . This change of coordinates can be made in two different ways according to the camera models surveyed.

The camera model proposed by Faugeras–Toscani, Faugeras–Toscani with distortion and by Weng use the following equations to carry out such a transformation:

$${}^I X_d = -k_u {}^C X_d + u_0, \quad {}^I Y_d = -k_v {}^C Y_d + v_0, \quad (14)$$

where  $(k_u, k_v)$  are the parameters that transform from metric measures with respect to the camera coordinate system to pixels with respect to the computer image coordinate system, and  $(u_0, v_0)$  are the components that define the projection of the focal point in the plane image in pixels, i.e. the principal point. They are used to determine the translation between both coordinate systems.

The camera model of Tsai proposed other equations to carry out the same transformation. These equations are the following:

$${}^I X_d = -s_x d'_x {}^{-1C} X_d + u_0, \quad {}^I Y_d = -d_y {}^{-1C} Y_d + v_0, \quad (15)$$

where  $(u_0, v_0)$  are the components of the principal point in pixels,  $s_x$  is the image scale factor,  $d'_x = d_x N_{cx} / N_{fx}$ ,  $d_x$  is the center to center distance between adjacent sensor elements in the  $X$  direction,  $d_y$  is the center to center distance between adjacent sensor elements in the  $Y$

direction,  $N_{cx}$  is the number of sensor elements in the  $X$  direction, and  $N_{fx}$  is the number of pixels in an image row as sampled by the computer.

### 3. Calibrating methods

The calibrating method depends on the model used to approximate the behavior of the camera. The linear models, i.e. Hall and Faugeras–Toscani, use a least-squares technique to obtain the parameters of the model. However, non-linear calibrating methods, as with Faugeras–Toscani with distortion, Tsai and Weng, use a two-stage technique. As a first stage, they carry out a linear approximation with the aim of obtaining an initial guess and then a further iterative algorithm is used to optimize the parameters. In this section, each calibrating method is explained detailing the equations and the algorithm used to calibrate the camera parameters.

#### 3.1. The method of Hall

The method used to calibrate the model of Hall is based on expressing Eq. (1) in the following form:

$$\begin{aligned} {}^I X_u &= \frac{A_{11} {}^W X_w + A_{12} {}^W Y_w + A_{13} {}^W Z_w + A_{14}}{A_{31} {}^W X_w + A_{32} {}^W Y_w + A_{33} {}^W Z_w + A_{34}}, \\ {}^I Y_u &= \frac{A_{21} {}^W X_w + A_{22} {}^W Y_w + A_{23} {}^W Z_w + A_{24}}{A_{31} {}^W X_w + A_{32} {}^W Y_w + A_{33} {}^W Z_w + A_{34}}. \end{aligned} \quad (16)$$

By arranging the variables, the following expressions are obtained:

$$\begin{aligned} 0 &= A_{11} {}^W X_w - A_{31} {}^I X_u {}^W X_w + A_{12} {}^W Y_w \\ &\quad - A_{32} {}^I X_u {}^W Y_w + A_{13} {}^W Z_w - A_{33} {}^I X_u {}^W Z_w \\ &\quad + A_{14} - A_{34} {}^I X_u, \\ 0 &= A_{21} {}^W X_w - A_{31} {}^I Y_u {}^W X_w + A_{22} {}^W Y_w \\ &\quad - A_{32} {}^I Y_u {}^W Y_w + A_{23} {}^W Z_w - A_{33} {}^I Y_u {}^W Z_w \\ &\quad + A_{24} - A_{34} {}^I Y_u. \end{aligned} \quad (17)$$

Finally, the unknowns  $A_{ij}$  are arranged in a 12-parameter vector  $(A)$ , obtaining the following equation:

$$QA = 0, \quad (18)$$

where  $A$  is the vector of 12 unknowns of Eq. (19).  $Q$  is a matrix of  $2n \times 12$  where  $n$  is the number of pair points used to calibrate the camera. A pair of points is formed by a 3D point expressed with respect to the world coordinate system  $\{W\}$  and its 2D projection expressed in coordinates from the image plane  $\{I\}$ .

$$A = (A_{11} \ A_{12} \ A_{13} \ A_{14} \ A_{21} \ A_{22} \ A_{23} \ A_{24} \ A_{31} \ A_{32} \ A_{33} \ A_{34})^T. \quad (19)$$

Each pair of points adds to the  $Q$  matrix the two following rows:

$$Q_{2i-1} = \begin{pmatrix} {}^wX_{u_i} \\ {}^wY_{u_i} \\ {}^wZ_{u_i} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -{}^I X_{u_i} {}^wX_{w_i} \\ -{}^I X_{u_i} {}^wY_{w_i} \\ -{}^I X_{u_i} {}^wZ_{w_i} \\ -{}^I X_{u_i} \end{pmatrix}^T, \quad Q_{2i} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ {}^wX_{u_i} \\ {}^wY_{u_i} \\ {}^wZ_{u_i} \\ 1 \\ -{}^I Y_{u_i} {}^wX_{w_i} \\ -{}^I Y_{u_i} {}^wY_{w_i} \\ -{}^I Y_{u_i} {}^wZ_{w_i} \\ -{}^I Y_{u_i} \end{pmatrix}^T \quad (20)$$

Consider then that the 3D position of a set of  $n$  calibrating points and their corresponding 2D projection in the image are known ( $n$  should be bigger or equal to 6). Moreover, consider without loss of generality that  $A_{34} = 1$ . This approximation can be assumed since the transformation matrix is defined up to a scale factor [2]. Then, all the elements of the  $A$  vector can be obtained by using a linear least-squares technique as the pseudo-inverse [20]. With the aim of applying the pseudo-inverse, it becomes necessary to modify Eq. (18) considering that  $A_{34} = 1$ , obtaining:

$$Q' A' = B', \quad (21)$$

where

$$A' = (A_{11} \ A_{12} \ A_{13} \ A_{14} \ A_{21} \ A_{22} \ A_{23} \ A_{24} \ A_{31} \ A_{32} \ A_{33})^T \quad (22)$$

and

$$Q'_{2i-1} = \begin{pmatrix} {}^wX_{u_i} \\ {}^wY_{u_i} \\ {}^wZ_{u_i} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -{}^I X_{u_i} {}^wX_{w_i} \\ -{}^I X_{u_i} {}^wY_{w_i} \\ -{}^I X_{u_i} {}^wZ_{w_i} \\ -{}^I X_{u_i} \end{pmatrix}^T, \quad Q'_{2i} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ {}^wX_{u_i} \\ {}^wY_{u_i} \\ {}^wZ_{u_i} \\ 1 \\ -{}^I Y_{u_i} {}^wX_{w_i} \\ -{}^I Y_{u_i} {}^wY_{w_i} \\ -{}^I Y_{u_i} {}^wZ_{w_i} \\ -{}^I Y_{u_i} \end{pmatrix}^T \quad (23)$$

$$B'_{2i-1} = ({}^I X_{u_i}), \quad B'_{2i} = ({}^I Y_{u_i}). \quad (24)$$

Finally, the vector of unknowns ( $A$ ) is computed by applying the pseudo-inverse shown in the following equation:

$$A' = (Q'^T Q')^{-1} Q'^T B'. \quad (25)$$

### 3.2. The method of Faugeras

In order to obtain the complete model of the camera proposed by Faugeras and Toscani, it is necessary to combine Eqs. (2)–(5) and (14), obtaining the equations.

$${}^I X_u = -k_u f \frac{r_{11} {}^wX_w + r_{12} {}^wY_w + r_{13} {}^wZ_w + t_x}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} + u_0,$$

$${}^I Y_u = -k_v f \frac{r_{21} {}^wX_w + r_{22} {}^wY_w + r_{23} {}^wZ_w + t_y}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} + v_0. \quad (26)$$

Note that Eqs. (26) can be expressed in a matricial form in the following manner:

$$\begin{pmatrix} s {}^I X_d \\ s {}^I Y_d \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} {}^wX_w \\ {}^wY_w \\ {}^wZ_w \\ 1 \end{pmatrix}, \quad (27)$$

where  $\alpha_u = -fk_u$  and  $\alpha_v = -fk_v$ . Then, by computing the product of both matrices, the transformation matrix  $A$  is obtained.

$$\begin{pmatrix} s {}^I X_d \\ s {}^I Y_d \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u r_1 + u_0 r_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v r_2 + v_0 r_3 & \alpha_v t_y + v_0 t_z \\ r_3 & t_z \end{pmatrix} \begin{pmatrix} {}^wX_w \\ {}^wY_w \\ {}^wZ_w \\ 1 \end{pmatrix}. \quad (28)$$

$$A = \begin{pmatrix} \alpha_u r_1 + u_0 r_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v r_2 + v_0 r_3 & \alpha_v t_y + v_0 t_z \\ r_3 & t_z \end{pmatrix}. \quad (29)$$

The camera parameters can be extracted from the symbolic matrix ( $A$ ) by equalling it to the numeric matrix obtained by calibrating the camera with the technique of Hall. Note that the orientation of the vectors  $r_i$  must be orthogonal and that it is also known that the dot product between two vectors is equal to the multiplication of their norms multiplied by the cosine of the angle they form. Using these relationships, the four intrinsic parameters ( $\alpha_u$ ,  $\alpha_v$ ,  $u_0$ ,  $v_0$ ) and the six extrinsic ones ( $r_1, r_2, r_3, t_x, t_y, t_z$ ) can be extracted from Eq. (29) in the following manner:

$$u_0 = A_1 A_3^T, \quad v_0 = A_2 A_3^T,$$

$$\alpha_u = -(A_1 A_1^T - u_0^2)^{1/2}, \quad \alpha_v = -(A_2 A_2^T - v_0^2)^{1/2},$$

$$\begin{aligned}
 r_1 &= \frac{1}{\alpha_u}(A_1 - u_0 A_3), & t_x &= \frac{1}{\alpha_u}(A_{14} - u_0 A_{34}), \\
 r_2 &= \frac{1}{\alpha_v}(A_2 - v_0 A_3), & t_y &= \frac{1}{\alpha_v}(A_{24} - v_0 A_{34}), \\
 r_3 &= A_3, & t_z &= A_{34},
 \end{aligned} \tag{30}$$

where the numerical matrix  $A$  is

$$A = \begin{pmatrix} A_1 & A_{14} \\ A_2 & A_{24} \\ A_3 & A_{34} \end{pmatrix}. \tag{31}$$

However, before estimating the camera parameters, the  $A$  matrix has to be calculated. Faugeras proposed a slightly different method of estimating  $A$  from the one proposed by Hall. Hence, the terms of Eq. (1) have been rearranged in the following way:

$$\begin{aligned}
 A_1 {}^w P_w + A_{14} - {}^I X_u (A_3 {}^w P_w + A_{34}) &= 0, \\
 A_2 {}^w P_w + A_{24} - {}^I Y_u (A_3 {}^w P_w + A_{34}) &= 0.
 \end{aligned} \tag{32}$$

Both equations are then factorized with respect to the unknowns, obtaining,

$$\begin{aligned}
 {}^I X_u &= \frac{A_1}{A_{34}} {}^w P_w + \frac{A_{14}}{A_{34}} - \frac{A_3}{A_{34}} {}^w P_w {}^I X_u, \\
 {}^I Y_u &= \frac{A_2}{A_{34}} {}^w P_w + \frac{A_{24}}{A_{34}} - \frac{A_3}{A_{34}} {}^w P_w {}^I Y_u.
 \end{aligned} \tag{33}$$

At this point, a set of 5 parameters is considered  $X = (T_1, T_2, T_3, C_1, C_2)^T$ , which are  $T_1 = A_1/A_{34}$ ,  $T_2 = A_3/A_{34}$ ,  $T_3 = A_2/A_{34}$ ,  $C_1 = A_{14}/A_{34}$  and  $C_2 = A_{24}/A_{34}$ .

$$\begin{aligned}
 {}^I X_u &= T_1 {}^w P_w + C_1 - T_2 {}^w P_w {}^I X_u, \\
 {}^I Y_u &= T_3 {}^w P_w + C_2 - T_2 {}^w P_w {}^I Y_u.
 \end{aligned} \tag{34}$$

Then, the value of the vector  $X$  is obtained by using a linear least-squares technique.

$$B = QX, \tag{35}$$

where

$$\begin{aligned}
 Q &= \begin{pmatrix} {}^w P_{w_i}^T & -{}^I X_{u_i} {}^w P_{w_i}^T & \dots & 0_{1 \times 3} & 1 & 0 \\ 0_{1 \times 3} & -{}^I Y_{u_i} {}^w P_{w_i}^T & {}^w P_{w_i}^T & {}^w P_{w_i}^T & 0 & 1 \\ & & \dots & & & \end{pmatrix}, \\
 B &= \begin{pmatrix} \dots \\ {}^I X_{u_i} \\ {}^I Y_{u_i} \\ \dots \end{pmatrix}.
 \end{aligned} \tag{36}$$

Hence, vector  $X$  is computed using Eq. (35).

$$X = (Q^T Q)^{-1} Q^T B. \tag{37}$$

Finally, the camera parameters are extracted from  $X$  by using Eq. (28).

$$\begin{aligned}
 T_1 &= \frac{r_3}{t_z} u_0 + \frac{r_1}{t_z} \alpha_u, & C_1 &= u_0 + \frac{t_x}{t_z} \alpha_u, \\
 T_2 &= \frac{r_3}{t_z}, \\
 T_3 &= \frac{r_3}{t_z} v_0 + \frac{r_2}{t_z} \alpha_v, & C_2 &= v_0 + \frac{t_y}{t_z} \alpha_v.
 \end{aligned} \tag{38}$$

At this point, it has to be considered that the norm of the three orientation vectors  $r_i$  is equal to unity by definition. By using Eqs. (38), the parameter  $t_z$  can then be computed. Hence, considering  $r_3 = 1$ ,

$$t_z = \frac{1}{\|T_2\|}. \tag{39}$$

The rest of the parameters can be obtained using the properties of the dot product and the cross product between vectors, which are

$$v_1 v_2 = \|v_1\| \|v_2\| \cos \alpha, \quad v_1 \wedge v_2 = \|v_1\| \|v_2\| \sin \alpha \tag{40}$$

so that,

$$\begin{aligned}
 r_i r_j^T &= 0, \quad i \neq j, & r_i \wedge r_j &= 1, \quad i \neq j, \\
 r_i r_j^T &= 1, \quad i = j, & r_i \wedge r_j &= 0, \quad i = j.
 \end{aligned} \tag{41}$$

The intrinsic parameters can then be obtained in the following way:

$$\begin{aligned}
 u_0 &= \frac{T_1 T_2^T}{\|T_2\|^2}, & v_0 &= \frac{T_1 T_3^T}{\|T_2\|^2}, \\
 \alpha_u &= -\frac{\|T_1^T \wedge T_2^T\|}{\|T_2\|^2}, & \alpha_v &= -\frac{\|T_2^T \wedge T_3^T\|}{\|T_2\|^2}.
 \end{aligned} \tag{42}$$

Moreover, the extrinsic parameters which model the orientation are the following:

$$\begin{aligned}
 r_1 &= -\frac{\|T_2\|}{\|T_1^T \wedge T_2^T\|} \left( T_1 - \frac{T_1 T_2^T}{\|T_2\|^2} T_2 \right), \\
 r_2 &= -\frac{\|T_2\|}{\|T_2^T \wedge T_3^T\|} \left( T_3 - \frac{T_2 T_3^T}{\|T_2\|^2} T_2 \right), \\
 r_3 &= \frac{T_2}{\|T_2\|}.
 \end{aligned} \tag{43}$$

Finally, the extrinsic parameters that model the translation are also obtained from (38).

$$\begin{aligned}
 t_x &= -\frac{\|T_2\|}{\|T_1^T \wedge T_2^T\|} \left( C_1 - \frac{T_1 T_2^T}{\|T_2\|^2} \right), \\
 t_y &= -\frac{\|T_2\|}{\|T_2^T \wedge T_3^T\|} \left( C_2 - \frac{T_2 T_3^T}{\|T_2\|^2} \right), \\
 t_z &= \frac{1}{\|T_2\|}.
 \end{aligned} \tag{44}$$

By using the  $r_i$  vectors in Eqs. (43), the rotation matrix  ${}^C R_W$  is directly obtained. The three angles  $\alpha$ ,  $\beta$  and  $\gamma$  can then be computed by equalling the symbolic rotation matrix to the numeric matrix obtained by calibration. At this point, all the parameters of the linear model of Faugeras are obtained. These parameters determine the relationship between the 3D object points with their projections, as shown in Eq. (28). However, the model of Faugeras can be more accurate if radial lens distortion is included.

### 3.3. The method of Faugeras with radial distortion

When a bright accuracy is necessary, the linear method of Faugeras becomes useless. However, it can be easily modified by including the radial lens distortion as it has been shown in Section 2.3. However, the equations become non-linear, and the linear least-squares technique has to be replaced by an iterative algorithm.

Note that by combining Eqs. (2)–(4), (6) and (7), the following equations are obtained:

$$\begin{aligned}
 {}^C X_d + {}^C X_d k_1 r^2 &= f \frac{r_{11} {}^W X_w + r_{12} {}^W Y_w + r_{13} {}^W Z_w + t_x}{r_{31} {}^W X_w + r_{32} {}^W Y_w + r_{33} {}^W Z_w + t_z}, \\
 {}^C Y_d + {}^C Y_d k_1 r^2 &= f \frac{r_{21} {}^W X_w + r_{22} {}^W Y_w + r_{23} {}^W Z_w + t_y}{r_{31} {}^W X_w + r_{32} {}^W Y_w + r_{33} {}^W Z_w + t_z}, \\
 r &= \sqrt{{}^C X_d^2 + {}^C Y_d^2}. \tag{45}
 \end{aligned}$$

Moreover, Eqs. (14) have to be used to transform from metric coordinates to pixels, Then, Eq. (46) defines the vector of unknowns which can be computed by using an iterative method as, for instance, the method of Newton-Raphson or Levenberg-Marquardt, among others [36].

$$x = (\alpha, \beta, \gamma, t_x, t_y, t_z, k_u, k_v, u_0, v_0, k_1)^T. \tag{46}$$

For example, the general method of Newton-Raphson minimizes the following equation:

$$G(x_k) \approx G(x_{k-1}) + J \Delta x_k, \tag{47}$$

where  $x$  is the unknown vector,  $G(x)$  is the minimization function,  $G(x_k)$  is a value close to the solution, and  $J$  represents the jacobian matrix of the function  $G(x)$ . With the aim of finding a solution of  $\Delta x_k$ , it is necessary to equal  $G(x_k)$  to zero.

$$G(x_k) = 0. \tag{48}$$

Note that one of the problems of convergence in iterative algorithms is the initial guess. However, an initial guess can be obtained by calibrating the linear method of Faugeras–Toscani without including lens distortion, and assuming  $k_1 = 0$ . Moreover, the difference between the initial value and the estimated parameters will be the

error of the function. For each iteration it is necessary to calculate the value of  $\Delta x_k$  to know the new value of  $x$ .

$$J \Delta x_k = -G(x_{k-1}). \tag{49}$$

By, applying Eqs. (45) and (14), and passing all the terms from the equality to the same side, functions  $U(x)$  and  $V(x)$  are defined.

$$\begin{aligned}
 U(x) &= f \frac{r_{11} {}^W X_w + r_{12} {}^W Y_w + r_{13} {}^W Z_w + t_x}{r_{31} {}^W X_w + r_{32} {}^W Y_w + r_{33} {}^W Z_w + t_z} - \frac{({}^I X_d - u_0)}{-k_u} \\
 &\quad - k_1 \left( \left( \frac{({}^I X_d - u_0)}{-k_u} \right)^2 + \left( \frac{({}^I Y_d - v_0)}{-k_v} \right)^2 \right) \\
 &\quad \cdot \frac{({}^I X_d - u_0)}{-k_u}, \\
 V(x) &= f \frac{r_{21} {}^W X_w + r_{22} {}^W Y_w + r_{23} {}^W Z_w + t_y}{r_{31} {}^W X_w + r_{32} {}^W Y_w + r_{33} {}^W Z_w + t_z} - \frac{({}^I Y_d - v_0)}{-k_v} \\
 &\quad - k_1 \left( \left( \frac{({}^I X_d - u_0)}{-k_u} \right)^2 + \left( \frac{({}^I Y_d - v_0)}{-k_v} \right)^2 \right) \\
 &\quad \cdot \frac{({}^I Y_d - v_0)}{-k_v}. \tag{50}
 \end{aligned}$$

In continuation, with the aim of solving the system, it is necessary to apply Eqs. (50) to the  $n$  calibrating points. However, in order to apply Eq. (49), it is necessary to get the symbolic function  $G(x)$  and its partial derivative matrix  $J$ , as it is shown in the following equations:

$$G(x_{k-1}) = \begin{pmatrix} U_1(x_{k-1}) \\ V_1(x_{k-1}) \\ \vdots \\ V_n(x_{k-1}) \end{pmatrix}. \tag{51}$$

$$J = \begin{pmatrix} \frac{\partial U_1(x_{k-1})}{\partial \alpha} & \frac{\partial U_1(x_{k-1})}{\partial \beta} & \dots & \frac{\partial U_1(x_{k-1})}{\partial k_1} \\ \frac{\partial V_1(x_{k-1})}{\partial \alpha} & \frac{\partial V_1(x_{k-1})}{\partial \beta} & \dots & \frac{\partial V_1(x_{k-1})}{\partial k_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial V_n(x_{k-1})}{\partial \alpha} & \frac{\partial V_n(x_{k-1})}{\partial \beta} & \dots & \frac{\partial V_n(x_{k-1})}{\partial k_1} \end{pmatrix}. \tag{52}$$

Finally, the parameters of the model are obtained by applying the pseudo-inverse of Eqs. (53) in each iteration. The more iterations done, the higher the accuracy obtained until convergence is achieved.

$$\begin{aligned}
 \Delta x_k &= -(J^T J)^{-1} J^T G(x_{k-1}), \\
 x_k &= x_{k-1} + \Delta x_k. \tag{53}
 \end{aligned}$$



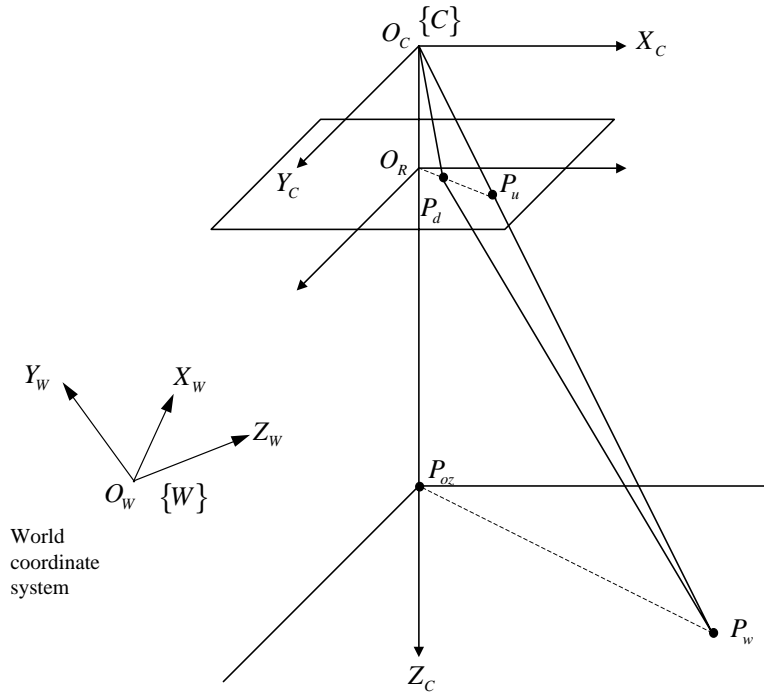


Fig. 2. Illustration of the radial alignment constraint [24].

### 3.4. The method of Tsai

The non-linear method of Faugeras was based on fixing the initial guess without considering lens distortion. Moreover, a large number of iterations are usually necessary to obtain an accurate value of the camera parameters. The method of Tsai [24] also models the radial lens distortion but assumes that there are some parameters of the camera which are provided by manufacturers. This fact reduces the number of calibrating parameters in the first step where an initial guess is estimated. Moreover, although all the parameters are iteratively optimized in the last step, the number of iterations is considerably reduced by using the calibrating algorithm proposed by Tsai.

Firstly, by combining Eqs. (2)–(4), (6) and (7), Eqs. (45) are obtained. Note that at this point model of Tsai is equivalent to the previous model of Faugeras with distortion (45).

Once  ${}^cX'_d$  and  ${}^cY'_d$  are obtained in metric coordinates by using Eq. (15), they can be expressed in pixels  ${}^lX_d$  and  ${}^lY_d$  and the following equations are obtained:

$${}^cX'_{di} = -({}^lX_{di} - u_0)d'_x, \quad {}^cY'_{di} = -({}^lY_{di} - v_0)d_y, \quad (54)$$

where

$${}^cX'_{di} = {}^cX_{di}s_x, \quad {}^cY'_{di} = {}^cY_{di}. \quad (55)$$

It is necessary therefore to find a relationship between the image point  $P_d$  (in metric coordinates) with respect to the object point  $P_w$ . Fig. 2 shows how the radial distortion affects the camera model. It can be observed that the segment  $\overline{O_R P_d}$  is parallel to the segment  $\overline{P_{oz} P_w}$ . Considering this constraint, the following relationship is established:

$$\overline{O_R P_d} // \overline{P_{oz} P_w} \Rightarrow \overline{O_R P_d} \times \overline{P_{oz} P_w} = 0. \quad (56)$$

By using Eq. (56), the following equations are obtained:

$$\overline{O_R P_d} \times \overline{P_{oz} P_w} = 0, \quad (57)$$

$$({}^cX_d, {}^cY_d) \times ({}^cX_w, {}^cY_w) = 0, \quad (58)$$

$${}^cX_d {}^cY_w - {}^cY_d {}^cX_w = 0. \quad (59)$$

Eq. (59) can be arranged expressing the object point  $P_w$  with respect to the world coordinate system, instead of expressing it with respect to the camera coordinate system.

$$\begin{aligned} & {}^cX_d(r_{21} {}^wX_w + r_{22} {}^wY_w + r_{23} {}^wZ_w + t_y) \\ & = {}^cY_d(r_{11} {}^wX_w + r_{12} {}^wY_w + r_{13} {}^wZ_w + t_x). \end{aligned} \quad (60)$$

Operating Eq. (60) and arranging the terms,

$$\begin{aligned}
 {}^C X_d &= {}^C Y_d \frac{{}^W X_w r_{11}}{t_y} + {}^C Y_d \frac{{}^W Y_w r_{12}}{t_y} + {}^C Y_d \frac{{}^W Z_w r_{13}}{t_y} \\
 &+ {}^C Y_d \frac{t_x}{t_y} - {}^C X_d \frac{{}^W X_w r_{21}}{t_y} - {}^C X_d \frac{{}^W Y_w r_{22}}{t_y} \\
 &+ {}^C X_d \frac{{}^W Z_w r_{23}}{t_y}.
 \end{aligned} \tag{61}$$

In order to compute Eq. (61) for the  $n$  points obtained from Eqs. (54), it is necessary to combine Eq. (61) with the Eqs. (55), obtaining,

$$\begin{aligned}
 {}^C X'_{di} &= {}^C Y'_{di} \frac{{}^W X_{wi} s_x r_{11}}{t_y} + {}^C Y'_{di} \frac{{}^W Y_{wi} s_x r_{12}}{t_y} + {}^C Y'_{di} \frac{{}^W Z_{wi} s_x r_{13}}{t_y} \\
 &+ {}^C Y'_{di} \frac{s_x t_x}{t_y} - {}^C X'_{di} \frac{{}^W X_{wi} r_{21}}{t_y} - {}^C X'_{di} \frac{{}^W Y_{wi} r_{22}}{t_y} \\
 &+ {}^C X'_{di} \frac{{}^W Z_{wi} r_{23}}{t_y}.
 \end{aligned} \tag{62}$$

At this point, a system with  $n$  equations and 7 unknowns is obtained, which can be expressed in the following form:

$$\begin{pmatrix} {}^C Y'_{di} \frac{{}^W X_{wi}}{t_y} \\ {}^C Y'_{di} \frac{{}^W Y_{wi}}{t_y} \\ {}^C Y'_{di} \frac{{}^W Z_{wi}}{t_y} \\ {}^C Y'_{di} \\ - {}^C X'_{di} \frac{{}^W X_{wi}}{t_y} \\ - {}^C X'_{di} \frac{{}^W Y_{wi}}{t_y} \\ - {}^C X'_{di} \frac{{}^W Z_{wi}}{t_y} \end{pmatrix}^T \begin{pmatrix} t_y^{-1} s_x r_{11} \\ t_y^{-1} s_x r_{12} \\ t_y^{-1} s_x r_{13} \\ t_y^{-1} s_x t_x \\ t_y^{-1} s_x r_{21} \\ t_y^{-1} s_x r_{22} \\ t_y^{-1} s_x r_{23} \end{pmatrix} = {}^C X'_{di}. \tag{63}$$

In order to simplify the notation, the 7 unknown components of the vector can be renamed  $a_i$ .

$$\begin{aligned}
 a_1 &= t_y^{-1} s_x r_{11}, & a_5 &= t_y^{-1} r_{21}, \\
 a_2 &= t_y^{-1} s_x r_{12}, & a_6 &= t_y^{-1} r_{22}, \\
 a_3 &= t_y^{-1} s_x r_{13}, & a_7 &= t_y^{-1} r_{23}, \\
 a_4 &= t_y^{-1} s_x t_x.
 \end{aligned} \tag{64}$$

Note that the  $a_i$  components can be easily computed by using a least-squares technique. Therefore, the point of interest is to extract the calibrating parameters of the camera from these  $a_i$  components. First  $t_y$  can be obtained by using Eqs. (64) in the following manner:

$$t_y = \frac{\|r_2\|}{\|a_{5,6,7}\|} \tag{65}$$

and Eq. (65) is simplified because the norm of the vector  $r_2$  is equal to the unity, obtaining the

parameter  $t_y$ .

$$|t_y| = \frac{1}{\sqrt{a_5^2 + a_6^2 + a_7^2}}. \tag{66}$$

However, Eq. (66) is insufficient since it does not provide the sign of the  $t_y$  component. In order to determine this sign, a point ( ${}^I X_d, {}^I Y_d$ ) located at the periphery of the image, far from the center, is taken from the set of test points (its corresponding 3D point is also kept). It is then supposed that the  $t_y$  sign is positive, and the following equations are computed:

$$\begin{aligned}
 r_{11} &= a_1 t_y / s_x, & r_{21} &= a_5 t_y, \\
 r_{12} &= a_2 t_y / s_x, & r_{22} &= a_6 t_y, \\
 r_{13} &= a_3 t_y / s_x, & r_{23} &= a_7 t_y, \\
 t_x &= a_4 t_y.
 \end{aligned} \tag{67}$$

By using the corresponding 3D point ( ${}^W X_w, {}^W Y_w, {}^W Z_w$ ), the linear projection of this 3D point on the image plane (sans lens distortion) can be computed by using the equations

$$\begin{aligned}
 {}^C X_u &= r_{11} {}^W X_w + r_{12} {}^W Y_w + r_{13} {}^W Z_w + t_x, \\
 {}^C Y_u &= r_{21} {}^W X_w + r_{22} {}^W Y_w + r_{23} {}^W Z_w + t_y.
 \end{aligned} \tag{68}$$

At this point the  $t_y$  sign can be verified. If both components of the point ( ${}^C X_u, {}^C Y_u$ ) have the same sign as the point ( ${}^I X_d, {}^I Y_d$ ), it means that the  $t_y$  sign was correctly chosen as positive. Otherwise, it has to be considered negative.

The second parameter to be extracted is the scale factor ( $s_x$ ). Note that by arranging Eqs. (64), the following equation is obtained:

$$s_x = \frac{\|a_{1,2,3}\| t_y}{\|r_1\|}, \tag{69}$$

where it is known that the norm of  $r_1$  is the unity and the scale factor is always positive. Then,  $s_x$  is obtained by using the equation.

$$s_x = \sqrt{a_1^2 + a_2^2 + a_3^2} |t_y|. \tag{70}$$

Furthermore, the 2D points, with respect to the camera coordinate system ( ${}^C X_d, {}^C Y_d$ ), can be computed from the same point with respect to the image coordinate system, that is ( ${}^I X_d, {}^I Y_d$ ), by using Eqs. (55). Moreover, by using Eqs. (67) the  $r_1$  and  $r_2$  vectors of the rotation matrix  ${}^C R_W$ , and the first element of the translation vector  ${}^C Y_B$ , i.e.  $t_x$ , can be calculated. Nevertheless, the third orientation vector ( $r_3$ ) can be computed by a cross product between  $r_1$  and  $r_2$  because of the property of orthogonality, (note also that the determinant of any rotation matrix is the unity, i.e.  $|\det R_{ij}| = 1$ ). At this point, the first three steps of the method of Tsai are completed, see Fig. 3.

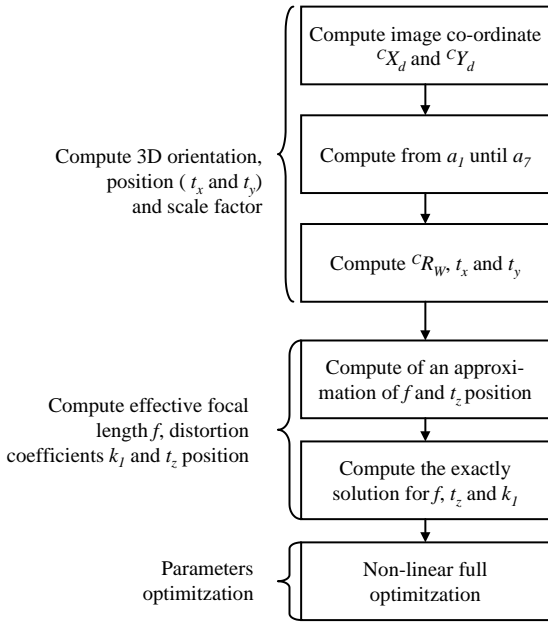


Fig. 3. Flowchart of the method of Tsai.

However, the following parameters are still unknown: the focal distance ( $f$ ), the radial lens distortion coefficient ( $k_1$ ), and the translation of the camera with respect to the  $Z$  axis ( $t_z$ ). In order to compute these last three parameters, a linear approximation is first used without considering the  $k_1$  parameter. The linear approximation is shown in Eq. (71), which was obtained from Eqs. (45).

$$\begin{aligned} & (r_{21} {}^W X_{wi} + r_{22} {}^W Y_{wi} + r_{23} {}^W Z_{wi} + t_y - cY_d) \begin{pmatrix} f \\ t_z \end{pmatrix} \\ &= (r_{31} {}^W X_{wi} + r_{32} {}^W Y_{wi} + r_{33} {}^W Z_{wi}) cY_d. \end{aligned} \quad (71)$$

Eq. (71) has now been applied to the whole set of test points, obtaining a system of  $n$  equations and two unknowns. The linear approximation of both unknowns,  $f$  and  $t_z$ , is obtained by using a pseudo-inverse. However, in order to calculate a better approximation including the  $k_1$  parameter, it is necessary to iterate Eqs. (45) by using an optimization method considering the linear method with  $k_1 = 0$  as an initial solution.

Finally, all the parameters are optimized iteratively with the aim of obtaining an accurate solution. The entire process is explained in Fig. 3.

### 3.5. The method of Weng

The method of Tsai is based on modelling radial lens distortion. The accuracy obtained by Tsai is sufficient for most applications. However, in some cases where the camera lens needs to be accurately modelled, a simple radial approximation is not sufficient. In such situations,

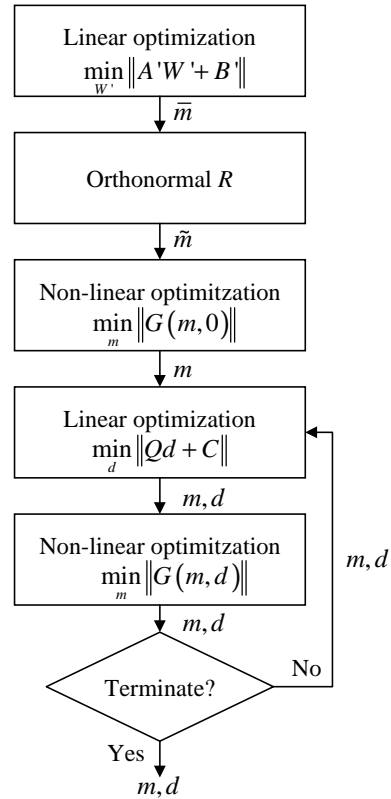


Fig. 4. Flowchart of the method of Weng [33].

Weng [33] modifies the model proposed by Faugeras–Toscani [32] including up to three types of lens distortion, as has been explained in Section 2.3. This fact increases the number of steps needed to calibrate the camera. A flowchart of the entire process is detailed in Fig. 4.

The first step is to obtain the complete model of Weng. However, Weng proposes to simplify the equations by introducing a variable substitution. Hence, equalling Eqs. (9) and (14), the following equations are obtained:

$$\begin{aligned} cX_u + \delta_x(cX_u, cY_u) &= ({}^l X_d - u_0) / -k_u, \\ cY_u + \delta_y(cX_u, cY_u) &= ({}^l Y_d - v_0) / -k_v. \end{aligned} \quad (72)$$

At this point, two new unknowns are introduced, in the following manner:

$$c\hat{X}_d = ({}^l X_d - u_0) / \alpha_u, \quad c\hat{Y}_d = ({}^l Y_d - v_0) / \alpha_v. \quad (73)$$

A substitution is then applied to simplify Eqs. (72), obtaining the equations

$$\begin{aligned} \frac{cX_u}{f} &= c\hat{X}_d - \frac{\delta_x(cX_u, cY_u)}{f}, \\ \frac{cY_u}{f} &= c\hat{Y}_d - \frac{\delta_y(cX_u, cY_u)}{f}. \end{aligned} \quad (74)$$

This replacement of unknowns is necessary because the value of  $({}^cX_u, {}^cY_u)$  cannot be obtained by observation. This fact makes it necessary to compute the distortion from the observed points after representing them with respect to the camera coordinate system, that is from  $({}^c\hat{X}_d, {}^c\hat{Y}_d)$  [30,33]. This replacement is reasonable because the distortion on the image plane suffered by the point  $({}^cX_u, {}^cY_u)$  is approximately equal to the distortion suffered by the point  $({}^c\hat{X}_d, {}^c\hat{Y}_d)$ . Therefore, the distortion coefficients in  $\delta'_x$  and  $\delta'_y$  will be estimated from  $({}^c\hat{X}_d, {}^c\hat{Y}_d)$ , instead of  $\delta_x$  and  $\delta_y$ , which was estimated from  $({}^cX_u, {}^cY_u)$ . As a result, the equations which relate distorted to undistorted points are the following:

$$\begin{aligned}\frac{{}^cX_u}{f} &= {}^c\hat{X}_d + \delta'_x({}^c\hat{X}_d, {}^c\hat{Y}_d), \\ \frac{{}^cY_u}{f} &= {}^c\hat{Y}_d + \delta'_y({}^c\hat{X}_d, {}^c\hat{Y}_d).\end{aligned}\quad (75)$$

Finally, redefining the coefficients  $k_1$  and  $g_1$  up to  $g_4$ , and combining Eqs. (2), (3) and (75) the complete camera model is obtained,

$$\begin{aligned}\frac{r_{11} {}^wX_w + r_{12} {}^wY_w + r_{13} {}^wZ_w + t_x}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} &= {}^c\hat{X}_d + (g_1 + g_3) {}^c\hat{X}_d^2 \\ &+ g_4 {}^c\hat{X}_d {}^c\hat{Y}_d + g_1 {}^c\hat{Y}_d^2 + k_1 {}^c\hat{X}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2), \\ \frac{r_{21} {}^wX_w + r_{22} {}^wY_w + r_{23} {}^wZ_w + t_y}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} &= {}^c\hat{Y}_d + g_2 {}^c\hat{X}_d \\ &+ g_3 {}^c\hat{X}_d {}^c\hat{Y}_d + (g_2 + g_4) {}^c\hat{Y}_d^2 + k_1 {}^c\hat{Y}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2).\end{aligned}\quad (76)$$

In order to be able to calibrate all the parameters of the model, Weng proposes to obtain a first approximation of the linear parameters, i.e. the extrinsic and intrinsic parameters without distortion. The  $m$  vector is now defined containing these linear parameters.

$$m = (u_0, v_0, \alpha_u, \alpha_v, t_x, t_y, t_z, \alpha, \beta, \gamma)^T. \quad (77)$$

Furthermore, the non-linear parameters which model the lens define a new vector  $d$ .

$$d = (k_1, g_1, g_2, g_3, g_4)^T. \quad (78)$$

Moreover, the calibration is based on the 3D test points and their projections. Let us call  $F$  the camera model,  $\Omega$  the set of 3D points, and  $\omega$  the set of their projections. Then, the calibration problem is the same as optimizing the parameters  $(m^*, d^*)$  which minimize the equation  $F$  by using both sets of test points.

$$F(\Omega, \omega, m^*, d^*) = \min_{m,d} F(\Omega, \omega, m, d). \quad (79)$$

This problem of optimization can be solved by using a non-linear method, in the following manner:

1. Fix  $d = 0$ .
2. Calculate  $m$ , which minimizes  $F$  by fixing  $d$ , that is:  $\min_m F(\Omega, \omega, m, d)$
3. Calculate  $d$ , which minimizes  $F$  by fixing  $m$ , that is:  $\min_d F(\Omega, \omega, m, d)$
4. Return to step 2 until the minimization error is sufficiently tolerable.

This method of optimization is used to solve diverse problems. First, the vector  $d$  can be coupled with  $m$  making the minimization of  $F$  false. Second, the intrinsic parameters cannot be optimized until a sufficient approximation of the extrinsic parameters is achieved. Third, since  $m$  corresponds to an approximation of the linear parameters, it cannot be the best solution if a significant distortion is presented.

With the aim of obtaining a good estimation of  $m$  with a non-linear optimization method, it is necessary to obtain an initial guess before iterating. Therefore, the initial guess is calculated supposing  $d = 0$ . Then, the model of Weng removing distortion, see Eqs. (74), is applied to the  $n$  calibrating points, obtaining  $2n$  equations of the form:

$$\begin{aligned}({}^lX_{u_i} - u_0) {}^wX_{w_i} r_{31} + ({}^lX_{u_i} - u_0) {}^wY_{w_i} r_{32} \\ + ({}^lX_{u_i} - u_0) {}^wZ_{w_i} r_{33} + ({}^lX_{u_i} - u_0) t_z - \alpha_u {}^wX_{w_i} r_{11} \\ - \alpha_u {}^wY_{w_i} r_{12} - \alpha_u {}^wZ_{w_i} r_{13} - \alpha_u t_x = 0, \\ ({}^lY_{u_i} - v_0) {}^wX_{w_i} r_{31} + ({}^lY_{u_i} - v_0) {}^wY_{w_i} r_{32} \\ + ({}^lY_{u_i} - v_0) {}^wZ_{w_i} r_{33} + ({}^lY_{u_i} - v_0) t_z - \alpha_v {}^wX_{w_i} r_{21} \\ - \alpha_v {}^wY_{w_i} r_{22} - \alpha_v {}^wZ_{w_i} r_{23} - \alpha_v t_y = 0.\end{aligned}\quad (80)$$

By using Eqs. (80), all the  $m$  parameters can be calculated. As the  $m$  vector has 10 unknowns, it is necessary to use at least 5 test points. Nevertheless, a large number of points is used in order to obtain a more accurate solution. The following parameters are then defined:

$$\begin{aligned}W_1 &= \alpha_u r_1 + u_0 r_3, & w_4 &= \alpha_u t_x + u_0 t_z, \\ W_2 &= \alpha_v r_2 + v_0 r_3, & w_5 &= \alpha_v t_y + v_0 t_z, \\ W_3 &= r_3, & w_6 &= t_z,\end{aligned}\quad (81)$$

where the vectors  $r_1$ ,  $r_2$  and  $r_3$  correspond to each row of the matrix  ${}^cR_{\mathcal{H}}$  respectively. Moreover, the set of Eqs. (80) is expressed in matricial form as

$$AW = 0, \quad (82)$$

where  $A$  is a matrix with  $2n$  rows and 12 columns.

$$A = \begin{pmatrix} -{}^wP_{w_1}^T & 0_{1 \times 3} & {}^lX_{u_1} & {}^wP_{w_1}^T & -1 & 0 & {}^lX_{u_1} \\ 0_{1 \times 3} & -{}^wP_{w_1}^T & {}^lY_{u_1} & {}^wP_{w_1}^T & 0 & -1 & {}^lY_{u_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -{}^wP_{w_n}^T & 0_{1 \times 3} & {}^lX_{u_n} & {}^wP_{w_n}^T & -1 & 0 & {}^lX_{u_n} \\ 0_{1 \times 3} & -{}^wP_{w_n}^T & {}^lY_{u_n} & {}^wP_{w_n}^T & 0 & -1 & {}^lY_{u_n} \end{pmatrix}. \quad (83)$$

However, the vector  $W = (W_1, W_2, W_3, w_4, w_5, w_6)^T$  cannot be directly calculated because of the homogeneity of the system, which deals with multiple solutions. However, only one of these potential solutions satisfies the following conditions: (a) The norm of the  $W_3$  vector has to be the unity because it is the third row of the rotation matrix; (b) The  $w_6$  sign has to coincide with the position of the optical center with respect to the image plane: positive if the  $z$ -axis intersects the image plane, and negative if otherwise.

With the aim of avoiding the homogeneity of the system of Eq. (82), it is necessary to impose the following temporary restriction:

$$w_6 = t_z = 1. \quad (84)$$

Hence, Eq. (82) is modified, obtaining

$$A'W' + B' = 0, \quad (85)$$

where  $A'$  is the first 11 columns of the  $A$  matrix,  $B'$  is the last column of  $A$  and  $W'$  is a vector of the 11 unknowns, i.e.  $W' = (W_1, W_2, W_3, w_4, w_5)$ . Then,  $W'$  is computed by using the pseudo-inverse,

$$W' = (A'^T A')^{-1} A'^T (-B'). \quad (86)$$

At this point,  $W'$  is the solution of the system shown in Eq. (85). However, in order to be a solution of Eq. (82) as well, it has to accomplish the two constraints. Therefore, the solution is divided by  $\|W_3\|$ , which forces the norm of  $W_3$  to be the unity, and replaces the  $w_6$  sign if necessary. See the equation.

$$S = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ s_4 \\ s_5 \\ s_6 \end{pmatrix} = \pm \frac{1}{\|W_3\|} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ w_4 \\ w_5 \\ w_6 \end{pmatrix}. \quad (87)$$

Moreover, knowing that the vectors  $r_1$ ,  $r_2$  and  $r_3$  are orthogonal, Eqs. (81) can be applied to obtain a first approximation of the  $m$  vector.

$$\begin{aligned} \bar{u}_0 &= S_1^T S_3, & \bar{v}_0 &= S_2^T S_3, \\ \bar{\alpha}_u &= -\|S_1 - \bar{u}_0 S_3\|, & \bar{\alpha}_v &= -\|S_2 - \bar{v}_0 S_3\|, \\ \bar{t}_x &= (s_4 - \bar{u}_0 s_6) / \bar{\alpha}_u, & \bar{r}_1 &= (S_1 - \bar{u}_0 S_3) / \bar{\alpha}_u, \end{aligned}$$

$$\begin{aligned} \bar{t}_y &= (s_5 - \bar{v}_0 s_6) / \bar{\alpha}_v, & \bar{r}_2 &= (S_2 - \bar{v}_0 S_3) / \bar{\alpha}_v, \\ \bar{t}_z &= s_6, & \bar{r}_3 &= S_3. \end{aligned} \quad (88)$$

However, this first approximation does not imply that the matrix  ${}^C\bar{R}_W$  is orthonormal. The next step consists of calculating the orthonormal matrix  ${}^C\bar{R}_W$ . The first step is to verify,

$$\|{}^C\bar{R}_W - {}^C\bar{R}_W\| = \min_{{}^C\bar{R}_W} \|{}^C\bar{R}_W - {}^C R_W\|. \quad (89)$$

With the aim of solving Eq. (89), it is rewritten including a  $3 \times 3$  identity matrix  $I$ .

$$\|{}^C\bar{R}_W I - {}^C\bar{R}_W\| = \min_{{}^C\bar{R}_W} \|{}^C\bar{R}_W - {}^C R_W\|. \quad (90)$$

A  $4 \times 4$  matrix  $B$  is then defined

$$B = \sum_{i=1}^3 B_i^T B_i, \quad (91)$$

where

$$B_i = \begin{pmatrix} 0 & (i_i - \bar{r}_i)^T \\ \bar{r}_i - i_i & (\bar{r}_i + i_i)_\times \end{pmatrix} \quad (92)$$

and where  $I = (i_1, i_2, i_3)^T$ , and  $(x, y, z)_\times$  is the antisymmetric matrix of the vector  $(x, y, z)$ , that is

$$(x, y, z)_\times = \begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix}. \quad (93)$$

The vector  $q = (q_0, q_1, q_2, q_3)^T$  is then obtained by calculating the eigenvalues associated with matrix  $B$ , where  $q_i$  is an eigenvalue and  $q_i \leq q_{i+1}$ . Finally, the solution of the matrix  ${}^C\bar{R}_W$  is shown in the following equation:

$${}^C\bar{R}_W = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_2 q_1 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_3 q_1 + q_0 q_2) & 2(q_3 q_2 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}. \quad (94)$$

With the orthonormal rotation matrix, the rest of the parameters are recalculated once more, obtaining:

$$\begin{aligned} \tilde{u}_0 &= S_1^T \tilde{r}_3, & \tilde{v}_0 &= S_2^T \tilde{r}_3, \\ \tilde{\alpha}_u &= -\|S_1 - \tilde{u}_0 \tilde{r}_3\|, & \tilde{\alpha}_v &= -\|S_2 - \tilde{v}_0 \tilde{r}_3\|, \\ \tilde{t}_x &= (s_4 - \tilde{u}_0 s_6) / \tilde{\alpha}_u, \\ \tilde{t}_y &= (s_5 - \tilde{v}_0 s_6) / \tilde{\alpha}_v, \\ \tilde{t}_z &= \tilde{t}_z. \end{aligned} \quad (95)$$

An iterative method is then used to recalculate, for the third time, the values of  $m$ , assuming zero distortion. Finally, a two-stage iterative method is used. In the first stage, the parameters of  $d$  are linearly obtained by using

least-squares. The second stage computes the values of  $m$  iteratively. These stages are repeated as many times as needed depending on the desired accuracy.

3.5.1. Stage of non-linear optimization of  $m$  by fixing  $d$ .

The camera model of Weng is expressed in Eq. (96), see also Eqs. (76).

$$\begin{aligned}
 U(x) &= \frac{r_{11} {}^wX_w + r_{12} {}^wY_w + r_{13} {}^wZ_w + t_x}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} \\
 &\quad - {}^c\hat{X}_d - (g_1 + g_3) {}^c\hat{X}_d^2 - g_4 {}^c\hat{X}_d {}^c\hat{Y}_d - g_1 {}^c\hat{Y}_d^2 \\
 &\quad - k_1 {}^c\hat{X}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2), \\
 V(x) &= \frac{r_{21} {}^wX_w + r_{22} {}^wY_w + r_{23} {}^wZ_w + t_y}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} \\
 &\quad - {}^c\hat{Y}_d - g_2 {}^c\hat{X}_d^2 - g_3 {}^c\hat{X}_d {}^c\hat{Y}_d - (g_2 + g_4) {}^c\hat{Y}_d^2 \\
 &\quad - k_1 {}^c\hat{Y}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2). \tag{96}
 \end{aligned}$$

Eq. (97) shows the function of minimization that has to be used in optimization.

$$\sum_{i=1}^n \{ ({}^lX_{di} - {}^lX_{di}(m, d))^2 + ({}^lY_{di} - {}^lY_{di}(m, d))^2 \}. \tag{97}$$

At this point any optimization algorithm such as Newton-Raphson or Levenberg-Marquardt can be used to optimize Eqs. (96).

3.5.2. Stage of linear optimization of  $d$  by fixing  $m$

Note that by arranging Eqs. (14) and (76), the equations which have to be optimized become linear. Therefore, they can be optimized by using the pseudo-inverse technique. The linear equations obtained are the following:

$$\begin{aligned}
 {}^lX_d(m, d) - {}^lX_d &= u_0 + \alpha_u {}^c\hat{X}_d - {}^lX_d \\
 &= u_0 + \alpha_u \left( \frac{r_{11} {}^wX_w + r_{12} {}^wY_w + r_{13} {}^wZ_w + t_x}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} \right. \\
 &\quad - (g_1 + g_3) {}^c\hat{X}_d^2 - g_4 {}^c\hat{X}_d {}^c\hat{Y}_d - g_1 {}^c\hat{Y}_d^2 \\
 &\quad \left. - k_1 {}^c\hat{X}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2) \right) - {}^lX_d, \\
 {}^lY_d(m, d) - {}^lY_d &= v_0 + \alpha_v {}^c\hat{Y}_d - {}^lY_d \\
 &= v_0 + \alpha_v \left( \frac{r_{21} {}^wX_w + r_{22} {}^wY_w + r_{23} {}^wZ_w + t_y}{r_{31} {}^wX_w + r_{32} {}^wY_w + r_{33} {}^wZ_w + t_z} \right.
 \end{aligned}$$

$$\begin{aligned}
 &\quad - g_2 {}^c\hat{X}_d^2 - g_3 {}^c\hat{X}_d {}^c\hat{Y}_d - (g_2 + g_4) {}^c\hat{Y}_d^2 \\
 &\quad \left. - k_1 {}^c\hat{Y}_d ({}^c\hat{X}_d^2 + {}^c\hat{Y}_d^2) \right) - {}^lY_d, \tag{98}
 \end{aligned}$$

where the function to minimize is expressed in the equation

$$\min_d \|Qd + C\|, \tag{99}$$

where

$$C = \begin{pmatrix} u_0 + \alpha_u \left( \frac{r_{11} {}^wX_{w_1} + r_{12} {}^wY_{w_1} + r_{13} {}^wZ_{w_1} + t_x}{r_{31} {}^wX_{w_1} + r_{32} {}^wY_{w_1} + r_{33} {}^wZ_{w_1} + t_z} \right) - {}^lX_{d_1} \\ v_0 + \alpha_v \left( \frac{r_{21} {}^wX_{w_1} + r_{22} {}^wY_{w_1} + r_{23} {}^wZ_{w_1} + t_y}{r_{31} {}^wX_{w_1} + r_{32} {}^wY_{w_1} + r_{33} {}^wZ_{w_1} + t_z} \right) - {}^lY_{d_1} \\ \vdots \\ u_0 + \alpha_u \left( \frac{r_{11} {}^wX_{w_n} + r_{12} {}^wY_{w_n} + r_{13} {}^wZ_{w_n} + t_x}{r_{31} {}^wX_{w_n} + r_{32} {}^wY_{w_n} + r_{33} {}^wZ_{w_n} + t_z} \right) - {}^lX_{d_n} \\ v_0 + \alpha_v \left( \frac{r_{21} {}^wX_{w_n} + r_{22} {}^wY_{w_n} + r_{23} {}^wZ_{w_n} + t_y}{r_{31} {}^wX_{w_n} + r_{32} {}^wY_{w_n} + r_{33} {}^wZ_{w_n} + t_z} \right) - {}^lY_{d_n} \end{pmatrix}. \tag{100}$$

$$Q = \begin{pmatrix} -\alpha_u {}^c\hat{X}_{d_1} ({}^c\hat{X}_{d_1}^2 + {}^c\hat{Y}_{d_1}^2) & -\alpha_u ({}^c\hat{X}_{d_1}^2 + {}^c\hat{Y}_{d_1}^2) \\ -\alpha_v {}^c\hat{Y}_{d_1} ({}^c\hat{X}_{d_1}^2 + {}^c\hat{Y}_{d_1}^2) & 0 \\ \vdots & \vdots \\ -\alpha_u {}^c\hat{X}_{d_n} ({}^c\hat{X}_{d_n}^2 + {}^c\hat{Y}_{d_n}^2) & -\alpha_u ({}^c\hat{X}_{d_n}^2 + {}^c\hat{Y}_{d_n}^2) \\ -\alpha_v {}^c\hat{Y}_{d_n} ({}^c\hat{X}_{d_n}^2 + {}^c\hat{Y}_{d_n}^2) & 0 \\ 0 & -\alpha_u {}^c\hat{X}_{d_1} & -\alpha_u {}^c\hat{X}_{d_1} {}^c\hat{Y}_{d_1} \\ -\alpha_v ({}^c\hat{X}_{d_1}^2 + {}^c\hat{Y}_{d_1}^2) & -\alpha_v {}^c\hat{X}_{d_1} {}^c\hat{Y}_{d_1} & -\alpha_v {}^c\hat{Y}_{d_1} \\ \vdots & \vdots & \vdots \\ 0 & -\alpha_u {}^c\hat{X}_{d_n} & -\alpha_u {}^c\hat{X}_{d_n} {}^c\hat{Y}_{d_n} \\ -\alpha_v ({}^c\hat{X}_{d_n}^2 + {}^c\hat{Y}_{d_n}^2) & -\alpha_v {}^c\hat{X}_{d_n} {}^c\hat{Y}_{d_n} & -\alpha_v {}^c\hat{Y}_{d_n} \end{pmatrix}. \tag{101}$$

The solution for  $d$  can now be obtained by using the pseudo-inverse in the following way.

$$d = -(Q^T Q)^{-1} Q^T C. \tag{102}$$

4. Accuracy evaluation

The systems used to evaluate the accuracy of camera calibration can be classified in two groups. The first group is based on analyzing the discrepancy between the real position of the 3D object point with respect to the 3D position estimated from its 2D projection. The second group compares the real position in pixels of a 2D image point with the calculated projection of the 3D object point on the image plane. In the following text, some of the most frequently used methods of accuracy evaluation are described.

4.1. 3D measurement

1. *3D position obtained from stereo triangulation.* In the first step, two images are acquired from a set of 3D test points whose 3D coordinates are known. In the second, the estimated 3D coordinates of the same points are computed from their projections using the calibrated parameters. Finally, the discrepancy between real and estimated positions is compared.
2. *Radius of ambiguity in the calibrating plane.* First, a set of 3D test points, which lay on test plane and whose coordinates in the world coordinate system are known, is acquired. Second, for each image point, the calibrated model is used to project the optical ray back from the focal point through the 2D projection. The transverse of the optical ray with the test plane determines the intersection point. The distance from the 3D test point to this intersection point defines a radius of ambiguity around the 3D point.
3. *Distance with respect to the optical ray.* This method is a generalization of the previous method. In this case, the discrepancy to be measured is the distance of the 3D test points from the optical ray generated from their projections.
4. *Normalized Stereo Calibration Error (NSCE)* [33]. The array of pixels in an image is projected back to the scene so 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. The orientation of the surface has been fitted to a plane which is orthogonal to the optical axis. Let the depth of this plane be equal to  $cZ_w$ , and the row and column focal lengths be  $\alpha_u$  and  $\alpha_v$ . The back projection of the pixel on this plane is a rectangle of  $a \times b$  size. Let the real coordinates of the  $i_{th}$  3D object points ( $cX_{wi}, cY_{wi}, cZ_{wi}$ ) be represented in the camera coordinate system, and let its coordinates obtained by back-projecting the pixel and intersecting it with the surface plane ( $c\hat{X}_{wi}, c\hat{Y}_{wi}, c\hat{Z}_{wi}$ ) be also represented in the camera coordinate system. With these given, the NSCE is defined as

$$NSCE = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(c\hat{X}_{wi} - cX_{wi})^2 + (c\hat{Y}_{wi} - cY_{wi})^2}{c\hat{Z}_{wi}^2(\alpha_u^{-2} + \alpha_v^{-2})/12} \right]^{1/2} \quad (103)$$

4.2. 2D measurement

1. *Accuracy of distorted image coordinates.* First, take an image of a set of 3D test points. Then, calculate the 2D position of each 3D point on the image plane, taking into account lens distortion. Accuracy is obtained by measuring the discrepancy between the real 2D points (obtained from image segmentation) and the estimated ones (obtained by using the camera model).

Table 1  
Accuracy of 3D coordinate measurement

	3D position (mm)			NSCE
	Mean	$\sigma$	Max	
Hall	0.1615	0.1028	0.5634	n/a
Faugeras	0.1811	0.1357	0.8707	0.6555
Faugeras NR <sup>1</sup> without distortion	0.1404	0.9412	0.0116	0.6784
Faugeras NR with distortion	0.0566	0.0307	0.1694	0.2042
Tsai	0.1236	0.0684	0.4029	0.4468
Tsai optimized	0.0565	0.0306	0.1578	0.2037
Tsai with principal point of Tsai optimized	0.0593	0.0313	0.1545	0.2137
Tsai optimized with principal point of Tsai optimized	0.0564	0.0305	0.1626	0.2033
Weng	0.0570	0.0305	0.1696	0.2064

<sup>1</sup>Newton-Raphson.

2. *Accuracy of undistorted image coordinates.* First, take an image of a set of 3D test points. Calculate the linear projection of the 3D points on the image plane, without taking lens distortion into account. Continue by determining the real 2D points through image segmentation and remove the lens distortion by using the camera model to obtain a set of undistorted points. Finally, accuracy is obtained by measuring the discrepancy between the linear projections and the undistorted points.

5. Experimental results

Instead of using our own experimental setup, we decided to download a list of corresponding points from the well-known Tsai's Camera Calibration Software Webpage (<http://www.cs.cmu.edu/~rgw/TsaiCode.html>). Actually, results are always conditioned to the structure of the 3D points and the image processing tools used in segmentation and further points extraction. Hence, this decision was just taken to allow the scientific community to reproduce the same conditions. Then, the surveyed calibrating techniques have been implemented and their accuracy measured using the following criteria: (a) Distance with respect to the optical ray; (b) Normalized Stereo Calibration Error; (c) Accuracy of distorted image coordinates; and (d) Accuracy of undistorted image coordinates.

The first two criteria calculate the accuracy with respect to a world coordinate system. The other two calculate the discrepancy on the image plane. First, Table 1 shows the accuracy measured by using the first criteria and the second criteria, respectively. Note that the NSCE method is not applicable to Hall because the method of

Table 2  
Accuracy of 2D coordinate measurement

	2D distorted image (pix.)			2D undistorted image (pix.)		
	Mean	$\sigma$	Max	Mean	$\sigma$	Max
Hall	0.2676	0.1979	1.2701	0.2676	0.1979	1.2701
Faugeras	0.2689	0.1997	1.2377	0.2689	0.1997	1.2377
Faugeras NR without distortion	0.2770	0.2046	1.3692	0.2770	0.2046	1.3692
Faugeras NR with distortion	0.0840	0.0458	0.2603	0.0834	0.0454	0.2561
Tsai	0.1836	0.1022	0.6082	0.1824	0.1011	0.6011
Tsai optimized	0.0838	0.0457	0.2426	0.0832	0.0453	0.2386
Tsai with principal point of Tsai optimized	0.0879	0.0466	0.2277	0.0872	0.0463	0.2268
Tsai optimized with principal point of Tsai optimized	0.0836	0.0457	0.2500	0.0830	0.0454	0.2459
Weng	0.0845	0.0455	0.2680	0.0843	0.0443	0.2584

Hall does not provide the camera parameters. Second, Table 2 shows the results of calculating the accuracy by using the third and fourth criteria, respectively. Note that the first three calibrating methods which do not include the modelling of lens distortion (i.e. Hall, Faugeras–Toscani and iterative Faugeras–Toscani without distortion) obviously give the same accuracy with distorted and undistorted 2D points as has been considered  $P_d = P_u$ .

These tables show the accuracy obtained by each of the camera calibration techniques surveyed. It can be observed that the techniques, which do not model lens distortion (the first three rows in the tables) provide less accuracy than the others, which do model the lens. Moreover, the technique of Hall appears as the best undistorted lens method because it is based on computing the transformation matrix without including any constraint. The other two techniques are based on a model which imposes a determined form of the transformation matrix. This fact ill effects the calibration. However, the discrepancy between their accuracy is not significant. Furthermore, the results show that the use of an iterative algorithm does not improve the accuracy obtained by using the pseudo-inverse in the technique of Faugeras–Toscani without distortion. This fact demonstrates that pseudo-inverse is the best approximation in undistorted models. In order to improve accuracy it has to go to lens modelling.

It can be observed from the tables that the non-linear techniques, which model lens distortion (the last 6 rows of the tables), obviously obtain better results than the undistorted techniques. However, the improvement obtained by the method of Tsai without optimization (fifth row) is not very significant because only a few parameters are iteratively optimized (i.e.  $f$ ,  $t_z$  and  $k_1$ ). Nevertheless, when the whole set of parameters is optimized, the method of Tsai (sixth row) shows the best accuracy obtainable despite needing more computing time. Note that accuracy is limited due to image segmentation and also that the model used always approximates the real behavior of the image sensor. However, if a real principal

point is known instead of the image center approximation, the Tsai method without optimization is as accurate as any iterative method, and allows a rapid computation. Note that the use of the Tsai optimized method by using the real principal point in the initial guess does not suggest an important improvement in the obtained accuracy. Finally, the results show that any iterative method which models lens distortion provides the same accuracy without depending on the kind of modelled lens. That is, the complete method of Weng does not obtain a better accuracy than the simple iterative method of Faugeras modelling only radial distortion. Even so, the accuracy is slightly less due to the complexity of this model which ill effects the calibration. The modelling of a camera including a large quantity of parameters does not imply that the accuracy obtained will be better.

## 6. Conclusions

This article surveys some of the most frequently used calibrating techniques. Effort has been made to unify the notation among these different methods, and they have been presented in a way the reader can easily understand. We can see that the differences among these techniques are mainly in the step concerning lens modelling. Also, the transformation from camera to image coordinates is slightly different in the method proposed by Tsai.

Furthermore, a survey on accuracy evaluation has been done. The methods surveyed have been implemented and their accuracy has been analyzed. Results show that only non-linear methods obtain a 3D accuracy smaller than 0.1 mm with a reasonable standard deviation. Moreover, the accuracy of non-linear methods on the image plane is much better than linear methods. Results show moreover that the modelling of radial distortion is quite sufficient when high accuracy is required. The use of more complicated models does not improve the accuracy significantly. It should be kept in mind that segmentation introduces a discrepancy between observable and mod-



elled projections which poses conditions on the accuracy. Moreover, when a low accuracy is sufficient, the fast and simple method of Hall is sufficient for most applications.

When comparing the obtained results, it can be seen that a relationship exists between the different criteria. Accuracy measuring methods obtain similar results if they are relatively compared. That is, good calibrating algorithms obtain acceptable accuracy results independently from the accuracy evaluation method used. Obviously, the results only prove something already demonstrated by the authors. However, in this article the accuracy has been measured by using the same test points for all the methods so results can be reliably compared. Hence, the reader can choose one or another method depending on the accuracy required. Moreover, once the calibrating method is chosen, the reader can take equations directly from this article to use in the desired calibrating algorithm.

## 7. Summary

In this article, we present a comparative study of the most commonly used camera calibrating methods of the last few decades. These techniques cover a wide range of the classical hard calibration of image sensors which begin from a previous knowledge of a set of 3D points and their corresponding 2D projections on an image plane in order to estimate the camera parameters. Hence, this study is presented describing a total of 5 different camera calibrating techniques which include implicit vs. explicit calibration and linear vs. non-linear calibration.

A great deal of attention has been paid to use the same nomenclature and a standardized notation in the presentation of all the techniques. Actually, this is one of the greatest difficulties which appears when going into the details of any calibrating technique. This problem usually arises because each method defines a different set of coordinate systems and camera parameters. Therefore, all the techniques have been re-arranged so as to allow a comparative presentation. The reader is introduced to calibration with the implicit linear technique of the pseudo-inverse presented by Hall, afterwards the explicit linear calibration of Faugeras-Toscani is presented. Furthermore, the article describes an easy modification of the Faugeras method in order to include radial lens distortion, the well-known method of Tsai and Tsai optimized, and finally the complete method of Weng which models up to three different kinds of lens distortion.

In order to compare the accuracy provided by each technique surveyed, a brief description of accuracy evaluation is presented. Each calibrating technique has been implemented and its accuracy evaluated. The same set of test points has been used for all the techniques, which allows the results to be reliably compared. Hence, the reader can choose one or another method depending on

the required accuracy. Moreover, once the calibrating method is chosen, the reader can take the equations directly from this article and easily use them in the desired calibrating algorithm.

There are numerous advantages thanks to an accurate calibration. For instance, dense reconstruction of 3D objects and surfaces has applications in visual inspection and medical imaging, such as quality control in industrial manufacturing and reconstruction of human backs and skulls for the detection of deformations or surgery. Another problem is the 3D pose estimation of an object in a scene, which has many applications such as obstacle avoidance, landmark detection and industrial part assembly, among others.

## Appendix

This appendix synthesizes the nomenclature used to express coordinate systems and camera parameters in the article.

$\{H\}$  defines a coordinate system H, which is composed of an origin  $O_H$  and either two  $\{X_H, Y_H\}$  or three  $\{X_H, Y_H, Z_H\}$  axis, depending on the number of dimensions defined.

The article defines the following coordinate systems:

- $\{W\} = \{O_W, X_W, Y_W, Z_W\}$  defines the world coordinate system.
- $\{C\} = \{O_C, X_C, Y_C, Z_C\}$  defines the camera coordinate system located at the focal point  $O_C$ .
- $\{R\} = \{O_R, X_R, Y_R\}$  defines the retinal coordinate system located at the principal point  $O_R = (u_0, v_0)$ .
- $\{I\} = \{O_I, X_I, Y_I\}$  defines the computer image coordinate system located in the upper-left corner of the image plane.

Each point P is always related to a coordinate system. Hence,  ${}^H P$  relates the point P with respect to  $\{H\}$ , where  ${}^H P = ({}^H X, {}^H Y, {}^H Z)$ . Each point can be related to any coordinate system. However, the following notations are the only ones used:

- ${}^W P_w = ({}^W X_w, {}^W Y_w, {}^W Z_w)$  expresses a 3D test point from the world (scene) expressed with respect to  $\{W\}$ .
- ${}^C P_w = ({}^C X_w, {}^C Y_w, {}^C Z_w)$  expresses a 3D test point from the world (scene) expressed with respect to  $\{C\}$ .
- ${}^C P_u = ({}^C X_u, {}^C Y_u, f) = ({}^C X_u, {}^C Y_u)$  expresses the linear projection of a point  ${}^C P_w$  on the image plane related to  $\{C\}$ , without including lens distortion.
- ${}^C P_d = ({}^C X_d, {}^C Y_d, f) = ({}^C X_d, {}^C Y_d)$  expresses a 2D image point, including lens distortion, related to  $\{C\}$ .
- ${}^I P_d = ({}^I X_d, {}^I Y_d)$  expresses a 2D image point related to the image coordinate system  $\{I\}$ , in pixels. This point is the observable point from image acquisition.

In order to distinguish a single point from a set, i.e. the set of test points, a second sub-index is used. Then,  $P_{ii}$  indicates the  $i$ th point on a set, where  $i = 1, \dots, n$ .

A rigid transformation between a two coordinate system is expressed by a transformation matrix, i.e.  ${}^J K_H$  expresses the coordinate system  $\{H\}$  with respect to  $\{J\}$ . Moreover,

$${}^J K_H = \begin{pmatrix} {}^J R_H & {}^J T_H \\ \mathbf{0}_{1 \times 3} & 1 \end{pmatrix},$$

where  $R = (r_1, r_2, r_3)^T$  expresses the orientation of  $\{H\}$  measured with respect to the axis of  $\{J\}$ .  $R$  can also be given related to the three rotation angles, i.e.  $\alpha$ ,  $\beta$  and  $\gamma$ . Moreover,  $T = (t_x, t_y, t_z)^T$  expresses the position of the origin of  $\{H\}$  with respect to  $\{J\}$ .

Finally, the following camera parameters are used:

- $k_1$  is the first coefficient of a series which models the radial lens distortion.
- $g_1$  up to  $g_4$  are the coefficients which model the de-centering and thin prism lens distortion.
- $f$  is the focal distance, i.e. the distance from the focal point  $O_C$  to the image plane.
- $(u_0, v_0)$  are the two components of the principal point, i.e. the projection of  $O_C$  on the image plane.
- $k_u, k_v$  are the two components which permit to transform a point from metric coordinates to pixels.
- $\alpha_u, \alpha_v$  are defined as  $\alpha_u = f k_u$  and  $\alpha_v = f k_v$ .
- $s_x$  is the scale factor.
- $d'_x = d_x N_{cx} / N_{fx}$
- $d_x, d_y$  are the center to center distances between adjacent sensor elements with respect to  $X$  direction and  $Y$  direction of the CCD sensor, respectively.
- $N_{cx}$  is the number of sensor elements in the  $X$  direction of the CCD sensor.
- $N_{fx}$  is the number of pixels in an image row as sampled by the computer.

## References

- [1] R.I. Hartley, Euclidean reconstruction from uncalibrated views, Second European Workshop on Applications of Invariance in Computer Vision, 1993, pp. 237–257.
- [2] O.D. Faugeras, Three-Dimensional Computer Vision, The MIT Press, Cambridge, MA, 1993.
- [3] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Vol. 2, Addison-Wesley Publishing Company, Reading, MA, 1993.
- [4] R. Ahlers, J. Lu, Stereoscopic vision—an application oriented overview, SPIE-Opt. Illumination, Image Sensing Mach. Vision IV 1194 (1989) 298–307.
- [5] J. Batlle, E. Mouaddib, J. Salvi, A survey: recent progress in coded structured light as a technique to solve the correspondence problem, Int. J. Pattern Recognition 31 (1998) 963–982.
- [6] R.A. Jarvis, A perspective on range finding techniques for computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 5 (1983) 122–139.
- [7] Z. Zhang, The matching problem: the state of the art, Technical Report No. 2146, Institut National de Recherche en Informatique et en Automatique, 1993.
- [8] T.S. Newman, A survey of automated visual inspection, Image Understanding 61 (1995) 231–262.
- [9] A. Casals, Sensor Devices and Systems for Robotics, Vol. 52, NATO ASI Series, Springer, Berlin, Heidelberg, 1989.
- [10] A. Broggi, Vision-based driving assistance in vehicles of the future, IEEE Intell. Systems 13 (6) (1998) 22–23.
- [11] L. Charbonnier, A. Fournier, Heading guidance and obstacles localization for an indoor mobile robot, IEEE International Conference on Advanced Robotics, 1995, pp. 507–513.
- [12] D. Khadraoui, G. Motyl, P. Martinet, J. Gallice, F. Chaumette, Visual servoing in robotics scheme using a Camera/Laser-stripe sensor, IEEE Int. J. Robotics Automat. 12 (1996) 743–750.
- [13] R.K. Lenz, R.Y. Tsai, Calibrating a cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 916–928.
- [14] M. Li, Camera calibration of a head-eye system for active vision, European Conference on Computer Vision, 1994, pp. 543–554.
- [15] M. Ito, Robot vision modelling—camera modelling and camera calibration, Adv. Robotics 5 (1991) 321–335.
- [16] R.K. Lenz, R.Y. Tsai, Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology, IEEE Trans. Pattern Anal. Mach. Intell. 10 (1988) 713–720.
- [17] M. Penna, Camera calibration: a quick and easy way to detection the scale factor, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 1240–1245.
- [18] Y. Liu, T.S. Huang, O.D. Faugeras, Determination of camera location from 2-D to 3-D line and point correspondences, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 28–37.
- [19] C.C. Wang, Extrinsic calibration of a vision sensor mounted on a robot, IEEE Int. J. Robotics Automat. 8 (1992) 161–175.
- [20] E.L. Hall, J.B.K. Tio, C.A. McPherson, F.A. Sadjadi, Measuring curved surfaces for robot vision, Comput. J. 15 (1982) 42–54.
- [21] J. Batista, H. Araujo, A.T. de Almeida, Iterative multistep explicit camera calibration, IEEE Int. J. Robotics Automat. 15 (1999) 897–916.
- [22] G.-Q. Wei, S. De Ma, Implicit and explicit camera calibration: Theory and experiments, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1994) 469–480.
- [23] O.D. Faugeras, G. Toscani, The calibration problem for stereo, Proceedings of the IEEE Computer Vision and Pattern Recognition, 1986, pp. 15–20.
- [24] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the shelf TV cameras and lenses, IEEE Int. J. Robot. Automat. RA-3 (1987) 323–344.
- [25] Z. Hong, J. Yang, An algorithm for camera calibration using a three-dimensional reference point, Pattern Recognition 26 (1993) 1655–1660.

- [26] S. Kamata, R.O. Eason, M. Tsuji, E. Kawaguchi, A camera calibration using 4 points targets, *Int. Conf. Pattern Recognition 1* (1992) 550–553.
- [27] L.L. Wang, W. Tsai, Camera calibration by vanishing lines for 3-D computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1991) 370–376.
- [28] S. Chen, W. Tsai, A systematic approach to analytic determination of camera parameters by line features, *Pattern Recognition* 23 (1990) 859–877.
- [29] T. Echigo, A camera calibration technique using three sets of parallel lines, *Mach. Vision Appl.* 3 (1990) 159–167.
- [30] C.C. Stama, C. Theurer, S.W. Henriksen, *Manual of Photogrammetry*, 4th Edition, American Society of Photogrammetry, Falls Church, VA, 1980.
- [31] J. Salvi, An approach to coded structured light to obtain three dimensional information, Ph.D. Thesis, Universitat de Girona, Departament d'Electrònica, Informàtica i Automàtica, 1997.
- [32] G. Toscani, *Systèmes de calibration et perception du mouvement en vision artificielle*, Ph.D. Thesis, Université Paris Sud, 1987.
- [33] J. Weng, P. Cohen, M. Herniou, Camera calibration with distortion models and accuracy evaluation, *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (1992) 965–980.
- [34] J. Salvi, J. Battle, E. Mouaddib, A robust-coded pattern projection for dynamic 3D scene measurement, *Int. J. Pattern Recognition Lett.* 19 (1998) 1055–1065.
- [35] J.Z.C. Lai, On the sensitivity of camera calibration, *J. Image Vision Comput.* 11 (1993) 656–664.
- [36] J. Stoer, R. Bulirsch, *Introduction to Numerical Analysis*, Springer, Berlin, 1980.

**About the Author**—JOAQUIM SALVI graduated in Computer Science in the Polytechnical University of Catalunya in 1993. He joined the Computer Vision and Robotics Group in the University of Girona, where he received the M.S. degree in Computer Science in July 1996 and the Ph.D. in Industrial Engineering in January 1998. He received the best thesis award in Industrial Engineering of the University of Girona. At present, he is an associate professor in the Electronics, Computer Engineering and Automation Department of the University of Girona. His current interest are in the field of computer vision and mobile robotics, focusing on structured light, stereovision and camera calibration.

**About the Author**—XAVIER ARMANGUE received the B.S. degree in Computer Science in the University of Girona in 1999 before joining the Computer Vision and Robotics Group. At present he is engaged in the study of stereovision systems for mobile robotics and he is working for his Ph.D. in the Computer Vision and Robotics Group in the University of Girona and in the Institute of Systems and Robotics in the University of Coimbra.

**About the Author**—JOAN BATLLE graduated in Physics in the Autonomous University of Barcelona, received the Ph.D. in Computer Science in the Polytechnical University of Catalunya. At present, he is a professor in the Electronics, Computer Engineering and Automation Department; the leader of the Computer Vision and Robotics Group; and the director of the Institute of Informatics and Applications. His research activity is mainly focused on real-time vision and autonomous robots. He is actually involved in some governmental projects about underwater robots and technology transfer to industrial enterprises.