

**A low-cost pose-measuring system for robot calibration**

G.D. van Albada, J.M. Lagerberg, A. Visser, L.O. Hertzberger

University of Amsterdam

Faculty of Mathematics and Computer Science

Department of Computer Systems

Kruislaan 403, 1098 SJ Amsterdam

The Netherlands

Telephone: (31) 20 525 7534

Fax: (31) 20 525 7490

E-mail: dick@FWI.UVA.NL

E-mail: jose@FWI.UVA.NL

E-mail: arnoud@FWI.UVA.NL

E-mail: bob@FWI.UVA.NL

**1. Abstract**

To maintain robot accuracy, calibration equipment is needed. In this paper we present a self-calibrating measuring system based on a camera in the robot hand plus a known reference object in the robot workspace. A collection of images of the reference object is obtained. Using image-processing, image-recognition and photogrammetric techniques, the positions and orientations of the camera are computed. The essential geometrical and optical camera parameters can be derived from the redundancy in the measurements. From each image the positions of markers on the reference object are extracted and the individual markers are identified. The camera positions for all images plus the parameters of the camera are solved together in a non-linear least-squares fitting procedure. Experimental results for this low-cost measuring system are presented.

Keywords: robot calibration, camera calibration, photogrammetry.

**2. Introduction**

Current and future robot applications require accurate position control for the robot. Robot repeatability has become much better in recent years, making this possible, in principle. Yet, more is needed to realise such accurate control, namely, an accurate control model, describing all the robot properties that influence the robot's positioning. The development of such models is described by Schröer [13]. The parameters in the model may change over time and therefore need to be determined repeatedly to maintain the control accuracy. Therefore, an easy-to-operate measuring system for use on the work floor is desirable for incidental and periodic (partial) recalibration.

Depending on the aims of a calibration session and the circumstances under which it is performed, different requirements must be met by the measuring and data-analysis procedures. It is easier to attain a moderate improvement in the positioning accuracy over a limited part of the reachable work space than to accurately identify all model parameters for a robot. Calibrating a robot at first installation is different from recalibrating a robot in a production line.

At the University of Amsterdam, we have developed a simple measuring system based on a camera in the robot hand, plus a known reference object in the robot workspace. It can be used to measure the 6-D robot pose in a limited part of the robot operating volume. The measuring system itself has been described elsewhere [1]. In this paper we shall describe the various data-processing

procedures required. This work described was performed for CAR, ESPRIT project nr. 5220<sup>1</sup>.

### **3. A short description of the measuring system**

The design aims for our measuring system were that it should be able to provide position and orientation data with an accuracy in the order of 0.1 mm and 1' in a limited part of the robot work space and that it should be low cost, portable, easy to operate by non-expert personnel and sufficiently robust.

The selected solution is a system based on a single camera in the robot hand, plus a specially designed, passive, flat reference object (reference plate) positioned in the robot work space.

When a sufficiently large and varied collection of images is available, the camera positions and orientations, the camera parameters and any distortion of the reference object can all be computed together, i.e. the system is self-calibrating, except for the absolute scale, position and orientation.

We have implemented a prototype version of the measuring system using a simple off-the-shelf camera. The reference plate consists of a blank aluminium plate with a black pattern of ring shaped markers printed onto it. Results, presented in this paper, show that the accuracy of the camera system, due to its self-calibration capacity, is generally sufficient for robot calibration.

The measuring procedure begins with the selection of the model parameters of the robot that need to be redetermined. Using this set of model parameters, the pose generation program (Albright, [2]) will generate a set of measurable poses that will allow these parameters to be computed.

Using these poses, a robot program is generated which directs the robot along a path containing these positions. At each measuring pose, the robot stops and a (stacked) image of the reference plate is obtained. The actual joint parameters at the measuring poses may be recorded, if possible, but may otherwise be assumed to be equal to the commanded values.

---

<sup>1</sup> In CAR the following companies and institutes co-operated: Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik (IPK Berlin, prime contractor), Leica (UK) Ltd., University of Amsterdam, Dept. of Computer Systems, TGT (Ireland), KUKA Schweißanlagen und Roboter GmbH, Volkswagen AG. ESPRIT projects are 50% funded by the EEC.

Next, the images obtained can be processed off-line to obtain the positions of the camera relative to the reference plate, plus the parameters of the camera. This “photogrammetric procedure” is the principal subject of this paper.

Using the calibration procedure developed at IPK, Berlin (Schröder, [13], Albright, [2]), the unknown robot parameters, plus the position of the reference plate relative to the robot base, can be derived from these photogrammetrically derived poses.

The measuring procedure is illustrated in figure 1.

Here figure 1.

#### **4. The photogrammetric procedure**

In the photogrammetric procedure, we process the obtained images of the reference plate and compute the camera position for each image. When a sufficiently large collection of sufficiently different images is available, the properties of the camera and of the reference plate can also be derived, making the system self-calibrating<sup>1</sup>.

The entire photogrammetric procedure consists of two separate tasks:

(1) *image-processing procedure*

recognising and identifying the markers on the reference plate and determining their positions in the image,

(2) *pose and parameter estimation procedure*

fitting of the model that can predict the position of the markers in every image. This is achieved by the solving the camera positions for all images and the camera parameters in a non-linear least-squares procedure.

In the image acquisition phase the reference plate containing a pattern of filled and open circular markers (figure 2) is observed with a CCD through a lens; the data are transferred to a frame grabber, digitised and stored. For optimum results in the image-processing phase (1), the digital image thus produced should have, as far as is possible, uniform brightness, uniform contrast, maximum resolution and few image defects.

Here figure 2.

---

<sup>1</sup> The only parameters that cannot be derived in this way are the scale and the absolute position and orientation of the reference plate. The impossibility to derive these 7 parameters is fundamental to any measuring system.

Some authors [e.g. 7] combine the image-processing phase and the pose and parameter estimation phase into a single image-reconstruction phase (2) by attempting to fit a suitably distorted version of the reference plate to the observed intensities. The advantage of such a procedure is that all the information in the image can be utilised; the drawbacks are that the pose and parameter estimation procedure becomes quite computationally intensive, and that image defects away from the critical edges can affect the quality of the fit.

We have chosen to use a contour-tracing approach to determine the positions of the markers in the image, primarily for reasons of simplicity.

Our pose and parameter estimation procedure can be used in several modes. When the camera and reference plate parameters are known with sufficient accuracy, the 6-D pose for an individual image can be calculated. When enough points have been identified in a series of at least four images (but preferably far more), the pose and parameter estimation procedure can be used to compute one unique set of camera parameters plus the 6 M pose parameters for all M images in a series, using a non-linear least-squares fitting procedure. For still larger numbers of images, distortions of the reference plate and additional corrections to the camera model can be computed.

#### **4.1. The image-processing procedure**

In the image-processing procedure, we take the digitised image as produced by the frame-grabber, remove various image defects, and attempt to extract the positions of the markers in the image with the best possible accuracy by contour integration. Accuracies better than 0.1 pixel can routinely be obtained. Next, we try to identify the individual markers, so that the position of the camera relative to the plate can be reconstructed.

The entire procedure has been built using the SCIL-Image environment (Van Balen et al. [3]), to which we have added routines specifically designed for our purposes.

##### 4.1.1. Image preprocessing

The reference plate has a background with a uniform, high reflectivity. On this background are markers with a uniform, low reflectivity. Ideally, this should result in an image with a constant, bright, background, with uniformly dark images of the markers. This is never the case. Vignetting, non-uniform illumination, anisotropic reflection or emission of the reference plate, sensitivity variations across the CCD etc. all produce variations in the brightness of the

image. Gain variations and dark-current contributions by the CCD and some of the vignetting can be removed using “white” and dark images (eq. 1).

To increase the number of measurable contours, the image noise can be reduced by the careful application of an edge preserving noise-reduction filter; in our case the Sigma filter of Lee [9]<sup>1</sup>, applied with quite a small value for sigma. When multiple images from the same position are available, the noise can be reduced by stacking the images. This technique requires the camera to remain absolutely still, but has the advantage that the location of the contours is not affected. The application of a Sigma filter will shift the contours slightly, but symmetry considerations indicate that the filter should have no systematic effect on the derived positions.

Gradual variations in the illumination and reflectivity of the reference plate, and any remaining vignetting effects are removed by dividing by an upper envelope of the image brightness. Sharp shadows are very difficult to remove completely and should be avoided when obtaining the image. The upper-envelope function is implemented by first determining a local maximum, followed by a local minimum [e.g. 15]. The presence of the dark images of the markers complicates the procedure, as it forces us to use a large structuring element. A procedure incorporating more knowledge of the image properties should perform better.

Disregarding the noise reduction, the image pre-processing procedure thus becomes:

$$I'_{\text{obs}} = \frac{(I_{\text{obs}} - I_{\text{dark}})}{I_{\text{white}}} \quad (1)$$

$$I_{\text{upper}} = \text{upper\_envelope}(I'_{\text{obs}}), \quad (2)$$

$$I''_{\text{obs}} = \frac{I'_{\text{obs}}}{I_{\text{upper}}}, \quad (3)$$

with  $I_{\text{obs}}$  the observed image,  $I_{\text{white}}$  an image of a uniformly illuminated uniform white surface and  $I_{\text{dark}}$  a dark current image obtained with the lens-cap on the lens.  $I''_{\text{obs}}$  is the final corrected image.

Some further enhancement of the images can be accomplished by applying a linear deconvolution filter. Symmetric smearing, caused by various lens and

---

<sup>1</sup> A large number of different noise reduction filters can be found in the literature. The Sigma filter is probably one of the safest filters for the current application.

focusing defects, is somewhat reduced by this technique. But, even more than the noise reduction filter, deconvolution should be applied with care. Firstly, all deconvolution filters invariably amplify the noise in the image. Secondly, they will affect the positions of the measured contours, and thirdly, as the smearing is unknown, but variable, only a slight reduction can safely be accomplished.

#### 4.1.2. Recognition of markers in the image

Recognition of the markers is performed by measuring the position and other properties of ellipses in the image.

The positions of the markers are found by integration of and model fitting to the iso-intensity contours at various intensity levels in the image. The contours belonging to the markers are recognised by determining the accuracy with which they fit an ellipse, plus some other heuristically-derived properties. This is done for the grey-scale image and for a slightly sharpness-enhanced version of the image. In order to ensure that each edge is measured with a good accuracy and that as few edges as possible are missed, contours are measured at four different levels in both versions of the image. A first suitable contour-level is determined from the histogram of intensities in the image using an "iso-data" algorithm<sup>1</sup>. Subsequent contour levels are determined taking into account the area enclosed by the contours. In this way a large amount of information is extracted.

For each contour, a number of parameters are derived. These are helpful for deciding the quality of the measurement and the nature of the measured point. These include the position of the centre, the area, the circumference, parameters of a fit to an ellipse, the quality of that fit, a bounding box and whether an inside or outside contour was measured. Very small and very large ellipses are given a low weight or rejected. In this way possible error sources like residual noise in the image and irregularities on the reference plate are handled.

As a first approximation, the centre of an elliptical contour is measured in the image plane. Strictly, this is incorrect as this point will not coincide with the image position of the centre of the circle on the reference plate. The effect is negligible only for small ellipses. It increases proportionally to the area of the ellipse, and increases with the projection angle in such a way that it becomes

---

<sup>1</sup> In this algorithm, a threshold or contour level is found such that it is the mean of the intensity weighted mean of the brighter pixels and that of the darker pixels. For most images this defines a unique level.

significant before the circle becomes clearly foreshortened. The formula for the displacement  $\Delta_c$  is, by close approximation:

$$\Delta_c = \ell_{\text{major}} \times \ell_{\text{minor}} \times \sin(\phi) \quad (4),$$

with  $\ell_{\text{major}}$  and  $\ell_{\text{minor}}$  the full major and minor axis lengths for the projected ellipse, and  $\phi$  the angle between the image plane (CCD) and the plane of the reference plate. All units are radians.

For instance, for a circle of 1 cm diameter, observed from a distance of 30 cm at an angle of  $45^\circ$ , the error amounts to about 2'. Therefore, the measured contours are stored and the deprojected centre of each contour is recomputed after an initial pose estimate has been obtained.

When the position and other properties for each contour for each marker image have been obtained, the identification procedure is started. First, all contours are sorted on centre position, so that those belonging to the same marker can be identified and their positions averaged using weights based on an estimate of the reliability of each measurement. Positions which are clearly discrepant are rejected. For each marker, a maximum of 16 contours can be obtained (inner edge and outer edge, plain and sharpness enhanced, 4 contour levels). The position of the marker is computed as a weighted mean of the positions derived for each contour. In order to obtain an estimate of the accuracy of the measurements, the positions derived from the inside edge of the annular markers is compared to those derived from the outside edge. In good quality images, root-mean-square (rms.) discrepancies as small as 0.02 pixel can be obtained in the vertical direction, with slightly worse figures obtained for the horizontal direction (the direction of the scan). The average accuracy obtained is often significantly worse than these figures.

#### 4.1.3. Identification of the markers

The next, and in our experience one of the trickiest problems to be solved, is the identification of the markers. The procedure must work even when the reference plate is observed at a very skew angle, and the markers and grid are significantly distorted. Furthermore, at small distances only a few markers are visible, and at large distances, many of small markers cannot be reliably measured.

To identify the markers, the measured marker positions are corrected for radial image distortion (this requires an initial estimate of the camera and lens parameters). The observed markers are ranked in apparent size, and clusters of up to eight small markers (satellites) around the large markers are identified,

starting with the largest marker. In this way most of the satellites of the large markers can be identified. In the clusters, pairs of opposite satellites are found and, from these, the four principal directions at each cluster. Each cluster is checked for consistency (e.g. the central marker must be significantly larger than any of the satellites) and the satellites are provisionally numbered, taking into account that the central marker must lie on the line connecting two opposite satellites. Markers in inconsistent clusters are retained to be identified, if possible, but are not used for the identification procedure.

Next, the nearest neighbour clusters of each cluster are found; once again this identification is checked for consistency (e.g. the distances to opposite neighbours must not differ too much). One of the main problems in this procedure is to reliably distinguish the diagonals from the principal axes when there is a strong perspective distortion.

None of the identifications is foolproof by itself, but enough information can be obtained in (almost) all cases to find a partial grid, from which projection parameters can be derived that allow the grid to be extrapolated. This allows more clusters of markers to be included recursively. In this procedure, the local principal axis directions for each cluster are verified once again.

By shifting and rotating the observed grid and comparing the pattern of open and closed markers with the expected pattern, a best fit to the expected pattern is obtained and the observed markers are identified. The procedure has been designed in such a way that identification still is possible even when not all satellites of a cluster have been observed, or when some of the satellites appear solid when they should be open. This is realised by distinguishing three types of matches for a cluster: mismatch, possible match, perfect match.

Our experience shows that the combination of a large number of markers and repeated consistency checks leads to a procedure where the majority of the markers can be identified automatically in almost all images.

When an apparently correct identification has been obtained, initial crude projection parameters can be derived. Using these, and the stored contour points, the marker positions are recomputed, correcting for perspective distortion, and the identification is once again verified.

The derived positions and identifications for each marker in an image are stored, together with an estimate of the accuracy. They are the input for the pose and parameter estimation procedure.

#### 4.2. The pose and parameter estimation procedure

In this section we describe the numerical technique used to estimate the camera positions and the model for the calibration system with a camera and a reference plate. The parameters to be calculated are the position and orientation of the camera (extrinsic parameters) with respect to the reference plate and the geometric and optical parameters of the camera (intrinsic parameters).

The camera model which contains the linear perspective and radial distortion effects of the lens and the CCD will be described first. We then describe the iterative procedure used to compute the camera poses and the intrinsic camera parameters. The intrinsic parameters can be found by analysing a collection of images obtained from sufficiently different poses. (Obviously, some information about camera parameters can be obtained from a single image; in order to reliably determine all camera parameters, a set of images is needed.) Finally, we describe how to compute the camera position for each individual image, given a first approximation of the intrinsic parameters.

Considerable work has been done on camera calibration; the work of Tsai [10, 11,12] has gained wide recognition. He uses a two-stage camera model, which is simple and computationally efficient. But the distance from his reference object to the lens is only about 13 cm. Furthermore, Tsai uses a separate method for calibrating the image centre, for which a special set-up is required.

Chang and Liang [5] reformulated Tsai's two-stage model into state space form and solved it using Kalman filters. Their final accuracy is about 0.35 mm at a distance of more than 1 m, but this is obtained by using two CCD cameras. In our method we use only one CCD camera.

The advantage of our technique is that the calibration of the image centre is incorporated into the procedure. Lenz and Tsai [10] have shown that the image centre could be off by up to 25 pixels, and should be calibrated for high accuracy applications. In our experiments with the HTH MX CCD camera the centre actually is at 30 to 40 pixels from the middle of the image. Our experiments show that we have to calibrate the image centre each time we take a series of measurements. The mounting between the lens and CCD is not completely rigid, which means that the intrinsic camera parameters have to be calibrated each time.

Our technique is very robust because we don't have specific requirements for the camera and the lens and try to eliminate the hardware errors with our software.

The result is that our technique can be used for any of the shelf camera with or without distortion, with or without a misalignment of the CCD sensor.

#### 4.2.1. Camera model

We have opted for a basically linear camera model (pinhole) to which corrections for camera distortion in the image plane can be added as needed.

Figure 3 illustrates the geometry of the camera and the various co-ordinate systems:

- $(X_w, Y_w, Z_w)$  is the 3D co-ordinate of a reference marker in the world co-ordinate system, which is defined by the orthonormal vectors  $\{\mathbf{x}_w, \mathbf{y}_w, \mathbf{z}_w\}$ .
- $(X_c, Y_c, Z_c)$  is the 3D co-ordinate of a reference marker in the camera co-ordinate system, which is defined by the orthonormal vectors  $\{\mathbf{x}_c, \mathbf{y}_c, \mathbf{z}_c\}$ .
- $(u, v)$  is the 2D co-ordinate of an image point, defined by the vectors  $\{\mathbf{u}, \mathbf{v}\}$ .

Here figure 3.

The world co-ordinate system is related to the camera co-ordinate system by a translation followed by a rotation. The relation between the camera co-ordinate system and the image co-ordinate system is given by a matrix containing the geometric camera parameters. This means that the transformation from 3D world co-ordinates to undistorted 2D images co-ordinates is given by the product of a translation matrix, a rotation matrix, and a matrix containing the geometric camera parameters.

The camera co-ordinate system is defined in such a way that:

- the origin lies in the optical centre of the camera,
- the centre of the image plane ( $u=0, v=0$ ) lies on the co-ordinate  $(0, 0, d)$  in the camera co-ordinate system. This determines the direction of the Z-axis of the camera co-ordinate system,
- the X-axis of the camera co-ordinate system is perpendicular to the Z-axis and an image point  $(u, 0)$  lies in the plane formed by the X- and Z-axis of the camera co-ordinate system,
- the Y-axis is perpendicular to the X- and the Z-axis.

The optical axis of the camera is not necessarily parallel to the Z-axis of the camera co-ordinate system, i.e. the optical axis of the camera need not pass through the centre of the image plane. Further it means that vectors  $(u, 0)^T$  and

$(0, v)^T$  are not necessarily aligned with the X- and Y-axis of the camera co-ordinate system. We also assume that the axes of the image plane are not necessarily orthogonal (this yields a less complex set of equations; in practice the axes are found to be orthogonal, as they should be).

Transformation from 3D world co-ordinates to 2D image co-ordinates

We are looking for a transformation from a reference marker  $(X_w, Y_w, Z_w)$  in the world co-ordinate system to an undistorted image point  $(u_{und}, v_{und})$ . Then, using the optical camera parameters, we compute the distorted image point  $(u_{comp}, v_{comp})$ , which can be compared with the measured image point.

Assume that for image  $i$  the camera is positioned at  $(x_i, y_i, z_i)$  and is rotated over  $(\alpha_i, \beta_i, \gamma_i)$  with respect to the X-, Y-, and Z-axes of the world co-ordinate system (Euler angles). Then, the transformation from the world co-ordinate system to the camera co-ordinate system is given by the product of rotation matrix R and translation matrix T:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R.T \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \tag{5}$$

where translation matrix T is a homogeneous matrix which transforms the origin of the world co-ordinate system to the origin of the camera co-ordinate system:

$$T = \begin{pmatrix} 1 & 0 & 0 & -x_i \\ 0 & 1 & 0 & -y_i \\ 0 & 0 & 1 & -z_i \end{pmatrix} \tag{6}$$

(Because we want to put a translation in matrix form we need a homogeneous matrix.) and R the 3-3 concatenation matrix of the pure rotation matrices defined by  $\alpha_i, \beta_i$  and  $\gamma_i$  :

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_i) & \sin(\alpha_i) \\ 0 & -\sin(\alpha_i) & \cos(\alpha_i) \end{pmatrix} \cdot \begin{pmatrix} \cos(\beta_i) & 0 & -\sin(\beta_i) \\ 0 & 1 & 0 \\ \sin(\beta_i) & 0 & \cos(\beta_i) \end{pmatrix} \cdot \begin{pmatrix} \cos(\gamma_i) & \sin(\gamma_i) & 0 \\ -\sin(\gamma_i) & \cos(\gamma_i) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7}$$

The combined 3-4 rotation-translation matrix R.T is denoted by RT.

We can split the intrinsic camera parameters into geometric and optical parameters.

The transformation from the camera co-ordinate system (as defined before) to the image co-ordinate system is denoted by matrix A, which is a 3 by 3 matrix.

Because of our choice of the camera co-ordinate system, this matrix contains a number of zeros. First of all, a point along the X-axis of the camera co-ordinate system should be transformed into a point on the u-axis of the image co-ordinate system. This means that  $A \cdot (1, 0, 0)^T = (a_{11}, a_{21}, a_{31})^T$  should have a zero for  $a_{21}$ .

Then, because a point on the Z-axis is transformed to the point (u=0, v=0), the third row of matrix A should contain zeros for  $a_{13}$  and  $a_{23}$ . The elements on the diagonal of the A matrix are the scale factors. Because the A matrix is a homogeneous matrix, its elements are determined only relatively to each other. A scale factor has to be chosen, therefore we choose  $a_{33}$  equal to 1. This means that matrix A containing the geometric camera parameters, is given by:

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} \quad (8),$$

with 5 camera parameters  $a_{11}$ ,  $a_{12}$ ,  $a_{22}$ ,  $a_{31}$  and  $a_{32}$  to be determined. If the image plane is perpendicular to the Z-axis of the camera co-ordinate system,  $a_{31}$  and  $a_{32}$  are equal to zero. If the axes of the CCD plate are orthogonal and if the CCD plate is perpendicular to the optical axis, matrix A is a diagonal matrix. Parameters  $a_{11}$  and  $a_{22}$  are in the order of the quotient of the distance d (between the optical centre and the centre of the image) and the pixel length.

The transformation from a reference marker in the world co-ordinate system to an undistorted point in the image co-ordinate system is given by:

$$\begin{pmatrix} u_h \\ v_h \\ w_h \end{pmatrix} = A \cdot RT \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (9),$$

with  $(u_h, v_h, w_h)$  the homogeneous co-ordinates of an image point. To obtain the undistorted 2D image co-ordinates  $(u_{und}, v_{und})$  we have to divide  $(u_h, v_h)$  by the scale factor  $w_h$  :

$$\begin{pmatrix} u_{und} \\ v_{und} \end{pmatrix} = \begin{pmatrix} u_h / w_h \\ v_h / w_h \end{pmatrix}. \quad (10)$$

The optical camera parameters are used in the relation between the undistorted and distorted image positions. In the course of our research, we have gradually increased the complexity of our distortion model. At first, we included only the

third order term of the radial lens distortion as part of the optical camera model, but results showed that we had to include the fifth order term of the radial lens distortion and the centre of the distortion as optical parameters. The computation of the distorted image co-ordinates ( $u_{\text{comp}}, v_{\text{comp}}$ ) in the image plane (using the third and the fifth order coefficients  $k$  and  $k'$  of the radial lens distortion with ( $D_u, D_v$ ) the centre of the distortion) is given by the following equations:

$$u_{\text{comp}} = u_{\text{und}} - \frac{(k r^2 + k' r^4) \cdot (u_{\text{und}} - D_u)}{1 + k r^2 + k' r^4} \quad (11a),$$

$$v_{\text{comp}} = v_{\text{und}} - \frac{(k r^2 + k' r^4) \cdot (v_{\text{und}} - D_v)}{1 + k r^2 + k' r^4} \quad (11b),$$

with  $r^2 = (u_{\text{und}} - D_u)^2 + (v_{\text{und}} - D_v)^2$ .

Later, discussions with R.S. Le Poole of the Leiden Observatory made us realise that, due to a parallax effect, the distortion depends not only on the angular distance to the optical axis, but also on the inverse of the linear distance to the object. This is most clearly illustrated by the fact that with a symmetrical lens, a 1:1 macro image should be free of distortion. Consequently,  $k$  and  $k'$  in the above formulae were replaced by<sup>1</sup>:

$$k = k_{\text{im}} + \frac{k_{\text{ob}}}{R_{\text{ob}}} \quad (12a)$$

$$k' = k'_{\text{im}} + \frac{k'_{\text{ob}}}{R_{\text{ob}}} \quad (12b),$$

where  $k_{\text{im}}$  and  $k'_{\text{im}}$  describe the distortion of the image side,  $k_{\text{ob}}$  and  $k'_{\text{ob}}$  the parallax of the object side.  $R_{\text{ob}}$  is the linear distance from the optical centre of the camera to the marker on the reference plate. For a symmetrical lens system,  $k_{\text{ob}}$  is about equal to  $-f k_{\text{im}}$ , where  $f$  is the focal length of the lens. We found a  $k_{\text{ob}}$  value close to this.

Image distortion and parallax are functions of the physical lens parameters, such as lens thickness and placement of the limiting aperture (see e.g. [6], section 6.3.1). The construction of the lens also affects the position and orientation of the

---

<sup>1</sup> Here, as in equations (11a,b), the distortion is assumed to be rotationally symmetric, implying among others that the lens elements are well aligned, and the optical axis of the lens system is perpendicular to the detector.

optical centre of the camera relative to its fixture to the robot, which must be determined in a separate procedure.

This gives us the complete transformation from a reference marker in the 3D world co-ordinate system to a 2D distorted image co-ordinate:

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \xrightarrow{\text{RT} \begin{pmatrix} x, \dots, \gamma \end{pmatrix}} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \xrightarrow{\text{A} \begin{pmatrix} a_{11}, \dots, a_{32} \end{pmatrix}} \begin{pmatrix} u_h \\ v_h \\ w_h \end{pmatrix} \xrightarrow{\text{cartesian}} \begin{pmatrix} u_{\text{und}} \\ v_{\text{und}} \end{pmatrix} \xrightarrow{\text{distortion} \begin{pmatrix} k_{\text{im}}, k'_{\text{im}}, k_{\text{ob}}, k'_{\text{ob}}, D_u, D_v \end{pmatrix}} \begin{pmatrix} u_{\text{comp}} \\ v_{\text{comp}} \end{pmatrix}. \quad (13)$$

This means that for each reference marker we have two non-linear functions:

$$u_{\text{comp}} = f(X_w, Y_w, Z_w, \mathbf{x}) \quad (14a)$$

$$v_{\text{comp}} = g(X_w, Y_w, Z_w, \mathbf{x}) \quad (14b)$$

with  $\mathbf{x} \equiv (x, y, z, \alpha, \beta, \gamma, a_{11}, a_{12}, a_{22}, a_{31}, a_{32}, k_{\text{im}}, k'_{\text{im}}, k_{\text{ob}}, k'_{\text{ob}}, D_u, D_v)$  the unknown parameters. These functions compute for each marker  $(X_w, Y_w, Z_w)$  in the reference plane the corresponding co-ordinate  $(u_{\text{comp}}, v_{\text{comp}})$  in the image

plane depending on the position of the camera  $(x, y, z, \alpha, \beta, \gamma)$ , the geometric camera parameters  $(a_{11}, a_{12}, a_{22}, a_{31}, a_{32})$  and the optical parameters  $(k_{\text{im}}, k'_{\text{im}}, k_{\text{ob}}, k'_{\text{ob}}, D_u, D_v)$ . This means that these two functions are dependent on the known  $(X_w, Y_w)$  co-ordinates of the reference marker and 17 unknown

parameters that we have to compute. (Since the reference markers lie in a common plane, the world co-ordinate system is chosen such that this plane lies on  $Z_w = 0$ ).

Although the derivation of the functions  $f$  and  $g$  is quite straightforward, the computation of these non-linear functions and their derivatives is complicated. We used Mathematica<sup>R</sup> [16] to solve this problem.

Our problem of camera calibration is to compute the intrinsic and extrinsic (position) parameters based on a number of points  $i$  ( $i = 1, \dots, N$ ) for which the reference co-ordinates  $(X_w, Y_w)_i$  are known and of which the image co-ordinates  $(u_{\text{obs},i}, v_{\text{obs},i})$  are measured. (We denote the unknown parameters by  $\mathbf{x}$ ). This means for each point  $i$  in each image we have 2 non-linear error equations:

$$u_{\text{obs},i} - f_i(\mathbf{x}) = u_{\text{obs},i} - u_{\text{comp},i} = \Delta m_{u,i} \quad (15a)$$

$$v_{\text{obs},i} - g_i(\mathbf{x}) = v_{\text{obs},i} - v_{\text{comp},i} = \Delta m_{v,i} \quad (15b)$$

with  $\Delta m_{u,i}$  and  $\Delta m_{v,i}$  the errors we want to minimise in a sense still to be defined. When we have a sufficiently large number of images, then the number of

measurements is much larger than the number of unknowns. This results in an over-determined system of non-linear equations, which can be solved by the Gauss-Newton method (ref. Schwarz [14], Chapter 7). Based on initial approximations of the desired unknowns, an iterative procedure is used to improve the unknown parameters. This iterative method will be described first. We then describe how the initial approximations are obtained.

#### 4.2.2. Iterative procedure for improvement of parameters

In (15) we have an over-determined system of non-linear equations to which we want to apply the method of least squares. This means we want to minimise the sum of  $e_{u,i}$  and  $e_{v,i}$ , which are the squares of the weighted residuals ( $i = 1, 2, \dots, N$ ) of the error equations:

$$w_{u,i}^2 (u_{obs,i} - u_{comp,i})^2 = e_{u,i} \quad (16a)$$

$$w_{v,i}^2 (v_{obs,i} - v_{comp,i})^2 = e_{v,i} \quad (16b)$$

The  $w_u$  and  $w_v$  are the relative weights of the measurements for each point.

Because it is quite troublesome to solve this system of non-linear equations, the non-linear error equations (15) are first linearised and to this linear system of equations the least-squares method is applied. The linearisation is done by constructing the Jacobian  $J$  of the functions  $f$  and  $g$  with respect to the unknowns:

$$J = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dy_1} & \dots & \frac{df_1}{dD_v} \\ \frac{dg_1}{dx_1} & \frac{dg_1}{dy_1} & \dots & \frac{dg_1}{dD_v} \\ \dots & \dots & \dots & \dots \\ \frac{dg_N}{dx_1} & \frac{dg_N}{dy_1} & \dots & \frac{dg_N}{dD_v} \end{pmatrix} \quad (17).$$

Let  $\Delta \mathbf{m}$  be the vector of length  $2N$  containing the residuals from (15),  $\mathbf{x}^0$  a vector of length  $n$  (with  $n$  the number of unknowns), containing an approximation to the unknowns  $\mathbf{x}$  and  $\Delta \mathbf{x}$  an adjustment to  $\mathbf{x}^0$ . Then the multiplication  $(J \cdot \Delta \mathbf{x})$  gives the first-order adjustment to  $f$  and  $g$  and thus to  $(\Delta \mathbf{m})$ . This means that at every step of the iteration the solution of the following system of linear equations is obtained:

$$W \cdot J \cdot \Delta \mathbf{x} = W \cdot \Delta \mathbf{m} \quad (18),$$

where  $W$  is the  $(2N$  by  $2N)$  diagonal weighting matrix for the equations. For reasons of simplicity, we will often omit  $W$  in the subsequent equations. Because this system is over-determined, we can only solve it in the least-squares sense.

The correction vector  $\Delta \mathbf{x}$  which is thus obtained will give an adjustment of the unknowns  $\mathbf{x}$ , and not the solution of the system of non-linear equations.

A safe way of solving an over-determined system of linear equations is by using the singular value decomposition which gives the least-squares solution of previous linear equations. The singular value decomposition can be quite helpful for an ill-conditioned system of linear equations, but involves a lot of computations. The solution is given by the pseudo-inverse  $J^+$  of Jacobian  $J$ :

$$J^+ \Delta \mathbf{m} = \Delta \mathbf{x} \quad (19).$$

(The pseudo-inverse is the equivalent of the inverse for a rectangular matrix.  $J$  must have full column rank in order for  $J^+$  to exist.) The pseudo-inverse multiplied with  $\Delta \mathbf{m}$  gives us an improvement of the old camera positions and camera parameters:

$$\mathbf{x}^1 = \mathbf{x}^0 + J^+ \Delta \mathbf{m} \quad (20).$$

A new Jacobian is constructed with respect to the new camera positions and new camera parameters. The pseudo-inverse of this Jacobian is again used to improve the previous estimates. This step is repeated. So at every step, we compute the new estimates of the  $n$  unknowns by solving  $2N$  linear equations generated by the Jacobian. Then, in the next iteration step, the new estimates are used to construct a new Jacobian. With this Jacobian (and the newly computed predicted marker positions) a new system of  $2N$  linear equations is solved. This iteration is continued until a certain norm of the correction vector is sufficiently small.

The condition number, and thus the attainable numerical accuracy for the system of equations, can be improved by using a ( $n$  by  $n$  square, diagonal) column weighting matrix  $C$ , replacing (18) by:

$$(W . J . C) . (C^{-1} \Delta \mathbf{x}) = W . \Delta \mathbf{m} \quad (21),$$

In the first experiments with 16 image viewpoints, we used the singular value decomposition to invert  $J$ . With a suitable choice for  $C$ , we obtained very good condition numbers for  $J$ . This means the singular value decomposition was not needed to solve our system and that a computationally more efficient method, such as the Gaussian elimination method, could be used. This is a classical method in which the so-called normal equations are solved:

$$(W . J . C)^T (W . J . C) (C^{-1} \Delta \mathbf{x}) = (W . J . C)^T \Delta \mathbf{m} \quad (22).$$

These equations were solved by Gaussian elimination with backward substitution.

This change has resulted in a very significant speed-up of the computation, especially for large amounts of input data. Furthermore, the amount of memory required was reduced dramatically, making it possible to process up to 150 images in a single run.

### Weighting of the equations

The quality of the measurements varies from image to image and from marker to marker in each image. This quality must be reflected in the weights in matrix  $W$  in order to obtain an optimal solution. Ideally, the weights should be chosen so that the expectations for the  $e_{u,i}$  and  $e_{v,i}$  values in (16) should be the same for all measurements. However, reliable values for these expectations are not readily available. Therefore, at the start of the iterative procedure, weight values are derived from the quality measures produced by the image recognition procedure. In the course of the iterative procedure, it will be found that some images have a larger overall spread in the residuals than average. In order not to let such images unduly influence the derived camera parameters and hence the positions derived for the other images, the weights for all markers in these images are correspondingly lowered. As the quality of the measurements in  $v$  is significantly better than those in  $u$  (because of the properties of the camera and the frame grabber), the corresponding equations are weighted accordingly. Furthermore, the fit obtained for some markers in an image can be much worse than the average for that image. Such markers are effectively removed from the solution by assigning a very low weight. On averages about 5% of all measurements are rejected in this way.

### 4.2.3. Solving for additional model parameters

In the course of our experiments we found that, for certain cameras, our distortion model was still inadequate. We also found that the reference plate was not as flat as we had initially assumed. Rather than adding a large number of additional parameters to our solution by further enlarging the number of terms in the Jacobian, we compute additional corrections straight from the residuals, after the solution has nearly converged.

For the camera distortion, this procedure is straightforward: the focal plane is divided into a grid of e.g. 14 by 14 cells, and the residuals in each cell are averaged. The averages are adjusted to remove skewing effects, rotation effects and any systematic scaling in  $u$  and  $v$ . These adjusted averages are then added to any previously computed corrections for each cell, after multiplication by a gain

factor less than 1 to ensure stability. Between two computations of these corrections, a few iterations of the least-squares procedure are run.

A similar approach is used for the distortions in the reference plate. Here, however, the  $(u, v)$  residuals in the focal plane must be converted into  $Z$  (or alternatively  $X, Y$  and  $Z$ ) corrections for each of the markers. This is accomplished using the derivatives already available in the Jacobian. For distortions in the reference plate only scaling and rotation effects are removed from the transformed and summed residuals.

Both procedures proved to be stable, but should only be applied when very large numbers of measurements are available, as some 1800 additional parameters can be added to the system. The most significant effect proved to be a non-flatness of the reference plate with an rms. amplitude of 0.1 mm, confirmed by physically re-measuring the reference plate.

#### 4.2.4. Initial estimate of position of camera

The iterative procedure discussed above assumes that initial estimates are available of the intrinsic and extrinsic parameters. These initial estimates do not have to be very accurate, because at each step of the iterative procedure the old estimates are improved by better ones. However, good initial estimates make the procedure more stable and efficient.

To get an initial approximation of the intrinsic parameters  $(a_{11}, \dots, D_v)$  is not difficult, because we can assume an ideal camera model with all parameters equal to zero except for the  $a_{11}$  and  $a_{22}$ , which are chosen to the quotient of the focal length  $f$  and the pixel length.

Based on these initial values for the 11 intrinsic parameters, each camera position associated with an image can be computed in the following way:

In (13) the relation between a reference marker in the image plane and the corresponding image point is given. From this relation it is clear that the known intrinsic parameters are located in the second (A-matrix) and fourth part (distortion) of the relation, and the 6 unknown position parameters in the first part (in matrix  $RT$ ). The first step of the problem is to use all the observed image points in one image to estimate the unknown  $RT$  matrix based on the approximately-known intrinsic parameters. In the next step, the 6 position parameters are computed from the obtained  $RT$  matrix.

If we apply the inverse of the fourth, third and second steps of the transformation in (13) to each observed image point in an image, we get an "observed" point in the camera co-ordinate system:

$$\begin{pmatrix} \mathbf{u}_{\text{obs}} \\ \mathbf{v}_{\text{obs}} \end{pmatrix} \xrightarrow{(k_{\text{im}}, k_{\text{im}}, k_{\text{ob}}, k_{\text{ob}}, D_u, D_v)} \begin{pmatrix} \mathbf{u}_{\text{und}} \\ \mathbf{v}_{\text{und}} \end{pmatrix}_{\text{obs}} \xrightarrow{\text{hom.}} \begin{pmatrix} \mathbf{u}_{\text{h}} \\ \mathbf{v}_{\text{h}} \\ 1 \end{pmatrix}_{\text{obs}} \xrightarrow{A^{-1} (a_{11}, \dots, a_{32})} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}_{\text{obs}}. \quad (23)$$

So we have a set of co-ordinates in the camera co-ordinate system corresponding to set of reference markers for which we want to find the linear relation RT:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}_{\text{obs}} = \text{RT} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (24)$$

with  $r_i$  and  $t_j$  to be computed. Because  $Z_w = 0$  we have for each image point the following 3 linear relations in 9 unknowns:

$$X_c^{\text{obs}} = r_1 X_w + r_2 Y_w + t_1 \quad (25a)$$

$$Y_c^{\text{obs}} = r_4 X_w + r_5 Y_w + t_2 \quad (25b)$$

$$Z_c^{\text{obs}} = r_7 X_w + r_8 Y_w + t_3 \quad (25c).$$

(It is clear that the third column of RT cannot be deduced.) For each image point we could rearrange this:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}_{\text{obs}} = \begin{pmatrix} X_w & Y_w & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & X_w & Y_w & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & X_w & Y_w & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_4 \\ r_5 \\ r_7 \\ r_8 \\ t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (26).$$

We use the method of least squares to compute RT for all the points in an image.

The next step is to compute the 6 position parameters ( $x, y, z, \alpha, \beta, \gamma$ ) from the derived RT matrix. The first 2 columns of matrix RT form the first 2 columns of a rotation matrix (3 by 3), and the last column contains the translation vector.

Because we cannot deduce the third column of the rotation matrix, we have to find a way to decompose matrix  $RT$ . Matrix  $RT$  has the following form:

$$RT = \begin{pmatrix} r_1 & r_2 & \cdot & t_1 \\ r_4 & r_5 & \cdot & t_2 \\ r_7 & r_8 & \cdot & t_3 \end{pmatrix} \quad (27),$$

with vector  $\mathbf{r}^1$  and  $\mathbf{r}^2$  the first and second column of the rotation matrix and vector  $\mathbf{t}$  containing the involved translation. We first compute vector  $\mathbf{r}^3$  perpendicular to the plane spanned by vector  $\mathbf{r}^1$  and  $\mathbf{r}^2$ . Then we apply the Modified Gram-Schmidt method [8] to the 3-3 matrix consisting of vector  $\mathbf{r}^1$ ,  $\mathbf{r}^2$  and  $\mathbf{r}^3$ , which decomposes this 3-3 matrix into the product of an orthogonal matrix and an upper triangular matrix. From the orthogonal matrix, the orientation ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) of the camera can be obtained. The inverse of matrix  $(\mathbf{r}^1 \mathbf{r}^2 \mathbf{r}^3)$  is used to compute the Cartesian position of the camera from vector  $\mathbf{t}$ .

So, if we have an initial estimate of the intrinsic camera parameters, we get a first approximation of each position of the camera based on all the points in the corresponding image.

## **5. Experimental results**

For the CAR project we developed and tested a prototype system based on the principles set out in the preceding sections. The prototype demonstrates the viability of the approach. For most of our measurements we used the following set-up on the "OSCAR" robot at our institute:

- A fixed focus, variable aperture,  $f=4.8$  mm lens at aperture ratios between  $f/4$  and  $f/8$ . No monochromatic filter was used.
- A HTH MX CCD camera with a 604(H) by 575(V)  $10\mu$  (H) by  $15\mu$  (V) pixels in a normal video mode (i.e. not pixel synchronous). The vertical line separation is  $7.5\mu$ , i.e. half the vertical pixel size.
- The images were digitised to 604 by 576 pixels, so that the horizontal pixel size corresponds approximately to the camera pixel size.

In this section we will discuss the results obtained for three sets of measurements that have been obtained using the "OSCAR" robot at the University of Amsterdam. We will primarily be concerned with obtaining estimates of the accuracy of these measurements.

The three sets are described in table I. They are identified by the positioning of the reference plate in the robot workspace.

Solutions were computed for various subsets of the data. First, a solution was computed for all 126 images together. This solution was used to compute the Z-displacements of the markers and residual image distortions as described in section 4.2.3. The derived values were used as fixed corrections for the reduction of each individual set of images.

Next, solutions were computed for each of the sets Left, Middle and Right. For each image we compute a number of error quantifiers. One significant quantity is the accuracy to which the measured points can be fitted by the model. The weighted rms. residuals  $\sigma(u)$  and  $\sigma(v)$  in  $u$  and  $v$  were computed separately for each image. Weighted averages of these residuals were computed for each solution, thus characterising the quality of the solution as a whole. In Table I these weighted averages are given. Table I also shows the number of points used for each solution, as well as the number of points discarded. Interestingly, the large central markers of the clusters have a five times higher probability of being discarded than the smaller peripheral markers<sup>1</sup>.

Set identifier	Number of images	Input number of points	Number of points discarded	weighted average of $\sigma(u)$ (pixels)	weighted average of $\sigma(v)$ (pixels)
All	126	26633	1264	0.068	0.050
Left	53	9807	442	0.061	0.046
Middle	48	10789	522	0.067	0.048
Right	25	6037	296	0.063	0.045

Table I. Characterisations of the three measurement sets obtained for the OSCAR robot and the solution obtained for the combined sets.

In order to obtain an estimate of the accuracy of the computed poses, we computed the expected error covariance matrices for the position and orientation and for the camera parameters from the  $\sigma(u)$  and  $\sigma(v)$  per image. By taking the square roots

---

<sup>1</sup> A possible explanation is that for the larger marker images any residual intensity gradients, perspective and distortion effects have a larger influence on the measured position. Even so, the large markers are useful in the identification procedure and at large distances when the small markers cannot be reliably measured, but the images of the large markers are small enough to be reliably measured.

of the diagonal elements of these covariance matrices, formal estimates of the errors in position ( $\Delta_x, \Delta_y, \Delta_z$ ), orientation ( $\Delta_\alpha, \Delta_\beta, \Delta_\gamma$ ) and camera parameters, are obtained. Assumptions essential for this computation are that the remaining errors per measured point are uncorrelated and that all systematic errors are accounted for. This corresponds to the assumption that our model contains all the significant effects affecting the positions of the marker images. Therefore, the (formal) errors computed in this manner are effectively lower bounds for the actual errors.

The results show a formal rms. accuracy of 0.10 mm and 1.0 minutes of arc, with a number of significantly worse points (figures 4 and 5).

Here figures 4 and 5.

The residuals in Table I show that the fit obtained for all images taken together is somewhat worse than that for the smaller sets. A comparison of the camera parameters derived for the various sets and subsets (see below) shows that the derived camera parameters have changed much more between the three sets of measurements than within each set. Therefore, the solution obtained for all images together is less reliable than the solution obtained for the individual sets.

As correlated errors and model defects could still be present, some experiments were performed to evaluate their effects. This was done by splitting the two larger sets of measurements (Left and Middle) into two smaller subsets. The split was done in two ways: by making subsets of the first and last 50% of the images, and by taking the even- and odd-numbered images. In other words, for the images in the Left and Middle sets, the camera positions were computed four times, each time using a different combination of images. We did not further use the positions derived using all sets together. This left three solutions per image: one from the complete set and two from subsets. The characterisations of these solutions are given in Table II.

Subset identifier	Number of images	Input number of points	Number of points discarded	weighted average of $\sigma(u)$ (pixels)	weighted average of $\sigma(v)$ (pixels)
Left all	53	9807	442	0.061	0.046
Left first	26	5291	244	0.060	0.046
Left last	27	4516	204	0.060	0.045
Left even	27	4685	214	0.058	0.045
Left odd	26	5122	235	0.061	0.046
Middle all	48	10789	522	0.067	0.048
Middle first	24	5670	292	0.067	0.050
Middle last	24	5119	241	0.066	0.045
Middle even	24	5096	247	0.067	0.048
Middle odd	24	5693	296	0.066	0.048

Table II. Characterisation of solutions obtained for subsets of the measurements.

The first/last split would be sensitive to time-dependent parameter variations, the even/odd split would be insensitive to those. The solutions obtained for these smaller sets were compared with those obtained by processing all 53 "Left" images together (all 48 "Middle" images together). The three solutions for each image were found to differ by a small but significant amount. These differences are entirely the result of small variations in the derived camera parameters between subsets, as the measurement data for each individual image are not changed. In figure 6, the spread in the positions obtained for the various solutions is illustrated.

Here figure 6.

The following conclusions were drawn from this comparison:

1. Differences between the positions computed in the different solutions were found to increase with the z-height to values about twice as large as the formal error. Orientation differences were slightly smaller than the formal error.
2. The internal consistency as evident from the  $\sigma(u)$  and  $\sigma(v)$  values of the first/last sets was not better than that of the even/odd sets.
3. The differences between the positions obtained with the first/last sets and those obtained for all 53 (all 48) images together were not larger or smaller than those obtained for the even/odd sets.

The last two conclusions show that any additional errors are not caused by a slow parameter drift, as this would cause the first/last sets to be more internally consistent than the even/odd sets and to differ more from the solution using all images.

The formal error estimate thus clearly underestimates the actual position error. The fact that  $\sigma(u)$  is consistently much larger and more variable than  $\sigma(v)$  suggests that the effect operates mainly in the horizontal pixel co-ordinate  $u$ . The pixel size in  $u$  is some 30% larger than in  $v$ , making the difference in accuracy even larger. The patterns in the errors are consistent with an assumption that the effect works mainly through the scale of the image. Very probably the imperfect synchronisation of the camera and the frame grabber in the scan-line direction ( $u$ ) is the principal contributor to this error. Synchronisation errors will cause small horizontal shifts of parts of scan-lines or even groups of adjacent scan-lines, causing correlated errors in the measured positions of all marker images intersected by that scan-line.

These experiments indicate that the formal error underestimates the actual errors, especially in the derived positions. The actual errors probably are in the order of 0.2 mm and 1 arc-minute. The use of a pixel-synchronous camera should significantly improve the performance of the system.

## **6. Discussion**

In this section we will discuss the possible sources of errors that should be taken into account when implementing a photogrammetric measurement system similar to ours. We have been able to reduce many of these error contributions by appropriate measures. Others, we have been able to model. Some of the principal remaining error sources have been identified and can be corrected in the future.

In our procedure, images of the reference plate are obtained with a camera consisting of a lens, a mounting and a CCD. Each of the components in this procedure contributes its own set of errors. These errors either have to be minimised by adopting a suitable measuring procedure, or have to be modelled in order to remove their contribution.

The contribution of each component in turn will be discussed.

### **6.1. The reference plate**

The measurement reference plate consists of a white flat plate with a large number of black, ring-shaped markers in a regular, grid-shaped pattern as illustrated in figure 2. Some of the markers are filled to yield solid black circles.

The pattern in these markers is used to recognise which part of the reference plate is observed in any one image. The combination of large and small markers makes it possible to recognise the reference plate across a large range of distances.

The reference plate must be made so that the following errors are avoided:

- mis-identification of the measured markers,
- erroneous reference positions for the markers
- inconsistent apparent marker positions when viewed from different directions.

For marker identification purposes, the current reference plate has been designed on the following basis. The markers are grouped in clusters of one large marker surrounded by eight small satellite markers. Clusters are characterised by open or filled central markers, and by having zero, one, or two filled satellites. For those having two filled satellites, the number of intervening open markers is determined, resulting in a total of twelve distinct patterns. The clusters are arranged in a square grid of 7 by 8 clusters. Each group of three clusters in each square of two by two clusters is unique in the pattern. Each group of three by three clusters can be uniquely identified by its pattern of open and closed central markers. In this way, a large range of views of the reference plate can be reliably identified. Designs have been made since then for reference plates of up to 11 by 12 clusters meeting even stricter uniqueness constraints.

Ideally, the accuracy of the reference plate must exceed the desired measuring accuracy by a wide margin. When no corrections to the positions of the markers can be derived from the measurements, the following properties are critical: the flatness (better than 0.02 mm), the smoothness of the edges of the markers (e.g. their freedom from smudges and scratches), and the positions of the centres of these markers (also better than 0.02 mm). As the reference plate may be observed from a very skew angle, the markings should not be raised above the surface of the plate.

Our reference plate was printed on a 4 mm aluminium plate, which was glued onto a 12 mm plate for stiffness. It shows variations in height exceeding 0.1 mm. We measure these variations in our photogrammetric procedure. Better plates can easily be manufactured, e.g. using carbon-fibre based materials. This should also reduce the risk of thermal expansion and deformation. For aluminium plates, thermal effects comparable to the desired measurement accuracy can easily occur.

A flat reference object (a plate) was selected because it can be more easily constructed and maintained than a 3-dimensional object. A 3-D object is, in principle, better for photogrammetric applications. The use of a plate is possible in combination with a wide-angle camera.

## 6.2. The illumination of the reference plate

Preferably, the reference plate should be uniformly illuminated and lack reflected highlights or shadows. Highlights visible to the camera from any of the measuring poses should also be avoided. As the camera will be used with quite a small aperture, a brightly back-lit plate should yield the best results. Any other light sources should be diffuse, so as to prevent sharp shadows or bright highlights that would complicate the image-processing.

## 6.3. The lens

The camera lens will be used to produce images of the reference plate markers over a range of object distances  $d_{ob;near}$  to  $d_{ob;far}$ , which we have put at 0.2 m and 1.4 m. The images should be as sharp as possible to obtain good measurements. Their measured positions should not depend on uncontrollable, variable effects. Lenses are known to display a large variety of imaging errors, affecting the quality of the image. The principal error types are the following:

- Defocusing and depth-of-field. A sharp image is produced only for an object at a precise distance. The further the object is removed from that distance, the more the image is smeared. Light from a point on the image is uniformly distributed in a circle centred on the desired image position. The distance range over which this effect stays within acceptable bounds is referred to as the depth-of-field. The effect is always present, but its effect on the image can be reduced by using short focal lengths and stopping down the aperture. Both increase the depth-of-field. As the effect is symmetrical, it need not strongly affect the measured position of the image.

The magnitude of the effect for an object at  $d_{ob;near}$  for an optical system focused at infinity is described by the following formula:

$$\Delta\alpha_{focus} = \frac{D}{d_{ob;near}} \quad (26),$$

where  $D$  is the diameter of the aperture,  $d_{ob;near}$  the (smallest) object distance, and  $\Delta\alpha_{focus}$  the angular diameter of the unsharp image. Focusing the system at  $2 d_{ob;near}$  instead of infinity will halve  $\Delta\alpha_{focus}$ .

- Diffraction. Due to the wave nature of light, the image of a point will be an Airy disk, if all other imaging errors are negligible. The diameter of this diffraction pattern is inversely proportional to the aperture diameter, thus limiting the degree to which the lens can be stopped down.

The diameter of the Airy disk is given by the following formula:

$$\Delta\alpha_{\text{airy}} = 1.22 \frac{\lambda}{D} \quad (27),$$

where  $D$  is the diameter of the aperture,  $\lambda$  the wavelength of the light, and  $\Delta\alpha_{\text{airy}}$  the radius of the Airy disk. We can now estimate an optimum aperture for the lens<sup>1</sup>, giving the sharpest image for an object at  $d_{\text{ob};\text{near}}$ . We want a depth of field ranging from an object distance  $d_{\text{ob};\text{near}}$  to infinity (i.e. we will use one single fixed focus over the entire distance range) with the lens focused at about  $2 d_{\text{ob};\text{near}}$ :

$$D^2 \approx 2.44 \lambda d_{\text{ob};\text{near}} \quad (28),$$

Depending on the illumination used,  $\lambda$  will be about  $6000 \text{ \AA} = 6 \cdot 10^{-7} \text{ m}$ ; with  $d_{\text{ob};\text{near}}$  about  $0.2 \text{ m}$ ,  $D$  should be about  $0.55 \text{ mm}$  - virtually independent of the focal length of the lens! The maximum angular resolution at infinity of the lens now becomes:

$$\Delta\alpha_{\text{airy}} = 1.22 \frac{\lambda}{D} \approx \sqrt{\frac{\lambda}{d_{\text{ob};\text{near}}}} \quad (29)$$

For our choice of parameters,  $\Delta\alpha_{\text{airy}} \approx 0.0017 \text{ radians} \approx 6'$ . For maximal resolution, the Nyquist sampling theorem indicates that a matching CCD should have a linear pixel size  $d_{\text{pix}}$  no larger than:

---

<sup>1</sup> A different result ( $D^2 \approx \lambda d_{\text{ob};\text{near}}$ ) is obtained when an approach based on the rms. phase difference in the image point is taken, analogous to that in chapter 9, section 3 "Tolerance conditions for primary aberrations" in [4]. However, this approach addresses the admissible aberration for a given aperture, i.e. the object distance at which the image sharpness begins to deteriorate measurably, whereas we want to derive the optimum aperture, leading to a different optimisation criterion. The difference is relatively minor when we take into account that the difference in  $D$  itself is 50% and that the image quality will only gradually deteriorate near the optimum. Furthermore, most images will be obtained at distances greater than  $d_{\text{ob};\text{near}}$ .

$$d_{\text{pix,max}} = \frac{\lambda d_{\text{im}}}{2D} \quad (30),$$

or about 0.0006  $d_{\text{im}}$ , where  $d_{\text{im}}$  is the image distance used (very nearly the focal length  $f$  in most cases). For our system, with 10 $\mu\text{m}$  pixels and an  $f = 4.8\text{mm}$  lens, we are still a factor three away from the desired resolution. As pointed out by the anonymous referee, aliasing will occur when the highest spatial frequency in the image exceeds the resolution of the CCD. This effect is already significantly reduced due to the fact that light-sensitive elements in a CCD are nearly as large as their spacing. It can be further reduced by intentional defocusing, or by further reducing the aperture.

Note that the desired (and achieved) angle measuring accuracy of the system significantly exceeds the angular resolution. This is achieved by a combination of sub-pixel interpolation in the grey-scale image in the computation of the position for each marker, the use of information from a large number of pixels for each marker, and the use of information derived from up to about 500 markers in each image.

- Other important image defects include:
  - Chromatic aberration. The images formed at different wavelengths can be displaced significantly relative to each other<sup>1</sup>. This can be reduced by using a band-pass filter, transmitting a band in the order of 100 Å. Choosing a band pass in the blue region of the spectrum will allow a smaller aperture and thus a sharper image to be obtained. However, the sensitivity of the system will be adversely affected and with too narrow a band-pass, speckling of the image may occur.
  - Barrel or pin-cushion distortion. Radial distortion is modelled in the 3rd and 5th order term in the least-square fit of the image-processing procedure. More complex, distance independent, distortion effects are modelled by averaging the residuals in the focal plane.
  - Parallax, or object distance dependent distortion, can be significant, but is usually much smaller than the distance independent effect. It has been included in our camera model.
  - Astigmatism and image-plane curvature, coma and spherical aberration. These are all reduced by using a small aperture. Their effects are not easily

---

<sup>1</sup> It may be assumed that lens manufacturers will consider displacements of less than 0.5 pixel insignificant for most applications, while we are interested in at least a 10 times better accuracy.

modelled.

- Vignetting. The brightness of the image falls off sharply with the distance to the centre of the image. This effect is, to a large extent, attributable to a simple geometric projection effect. Some reduction is possible by using a specially designed lens. However, vignetting can be modelled, or measured using a "white" image as in our procedure, and thus is mostly corrected in the image-processing procedure. The intensity gradient across a single pixel can have a very small effect on the measured position for the image contours, but vignetting is mainly a problem for the identification process. The combination of vignetting and unsharp images will lead to errors in the position which are proportional to the square of a marker's diameter in the image.
- Ghost images due to internal reflections in the lens can mostly be removed by coatings. This effect cannot easily be modelled.
- Dirt and scratches on the lens surfaces. Each speck of dirt will produce a diffraction image superposed on the desired image. The amount of light in this image is proportional to the area of the blemish.

Off-the-shelf lenses, such as the lenses used in our experiments, are presumably optimised to yield an image that is pleasing to the eye. They should have a reasonable overall sharpness over the whole image and colour range and a not-too-extreme distortion of the image. Certain image defects that are harmful for photogrammetric applications, like coma and chromatic aberration, are likely to be present, but can be significantly reduced by appropriate measures. For ultimate performance, a specially designed lens should be used, making use of the specific trade-offs allowed for photogrammetry.

As the inner and outer edges of our ring-like markers provide us with two independent measures of the marker position, we have a probe of the measuring accuracy. As the dark-to-light transitions at these edges go in opposite directions, certain asymmetric shifts can be detected, in principle.

#### **6.4. The mounting**

The mounting is important as it fixes the position of the lens relative to the detector, in our case a CCD. The mounting can allow the distance of the lens to the detector to be changed (focusing), the aperture to be changed and the lens to be removed from the camera. Each of these options implies a mechanical change to the optical system, leading to non-reproducible variations in its properties. For that reason, a fixed focus, fixed aperture lens is preferred.

### 6.5. The CCD and the frame grabber

The image produced by the lens must be detected using a CCD or a similar (rectangular) array of detector elements. In this procedure, distortions in the image positions and in the intensities may be caused.

The output of the detector elements is usually converted to a standard video signal, which is digitised using a frame grabber. A problem with this procedure is that the outputs of adjacent detector elements can be mixed in an unpredictable fashion in the output signal. Synchronisation errors between the camera and the frame grabber can also lead to geometric distortions that vary from line to line in the image. Though much less than one pixel width, these appear to be the principal remaining error source in our system. For these reasons, the use of a "pixel synchronous" detection system is preferred.

The sensitivity to light will vary across the surface of each detector element, leading to variations in average gain and effective position of the detector elements.

Each element of the detector array will also have a dark-current - i.e. a non-zero output even when there is no incident light, and a noise contribution. CCD detectors have the advantage that they are highly linear. Dark-current and (read-out) noise are strongly temperature dependent.

Dark current contributions can be subtracted, and gain variations can be divided out in the image-processing step, provided that dark and white images have been obtained, preferably immediately before and after a measuring sequence. The white image is also useful for the removal of the vignetting contribution. Automatic gain correction in the camera can interfere with these corrective measures.

The noise contribution of the CCD can be reduced by stacking images obtained without moving the camera. This may be advisable as the small lens aperture may result in relatively low contrast, noisy images.

The choice of a correct illumination intensity is important to avoid saturation, and to avoid excessive digitisation noise. Preferably, the illumination should be adjusted on basis of the image obtained. Automatic adjustment of the camera aperture or camera gain is not advisable for the reasons stated before. But, in adjusting the illumination, care should be taken not to change the colour of the light.

Variations in the (effective) positions of the individual detector elements are difficult to model reliably. Displacements affecting clusters of detectors can be found and removed in much the same way that lens defects are removed. The higher the desired accuracy, the more images are needed to obtain the model parameters.

## **7. Conclusions**

In this paper we have presented a low-cost method, based on photogrammetry, to obtain measurements for the calibration of robot systems.

The method has been implemented and tested and provides sufficiently good results for practical application, namely pose information, with an accuracy of about 0.2 mm and 1 arc-minute in a volume of 0.5 m<sup>3</sup>. The components used are relatively inexpensive, and can easily be combined to yield a portable system.

As most of the data processing has been highly automated, such a system will be usable by non-expert personnel.

In order to maximise the reliability of the measurements, each series of poses must be constructed so that a number of poses allowing an accurate calibration of the camera and the reference plate will be measured several times. Each series should contain a large number of different poses. We suggest to use at least 60.

Various improvements and extensions to the system are still possible. Better hardware combined with more sophisticated measuring procedures should lead to a significantly better accuracy. The software could be extended to allow a 3-dimensional reference consisting of two or more flat plates to be used. In this way, the measuring volume can be significantly extended.

By combining the video camera with a fast frame grabber + recording system, or alternatively with a video recorder, dynamic measurements can be obtained.

The relative locations and orientations of two robots in a workcell can be found by placing the reference plate between the robots and calibrating both robots with that common reference.

## **8. Acknowledgements**

The research described in this paper was partly funded by the EEC through ESPRIT II project 5220 "CAR". Stephen Kyle of Leica, and Steve Albright, Klaus Schröder and Michael Grethlein of IPK have contributed significantly to the development of the system through their expert advice and support. Aside from their scientific contribution, their support in other areas, such as Klaus Schröder's

help in obtaining a new camera when ours broke down, was also greatly appreciated. Discussions with R.S. Le Poole of the Leiden Observatory helped clarify the relationship between parallax and distortion.

## **9. References**

- [1] G.D. van Albada, J.M. Lagerberg and A. Visser, Eye in Hand Calibration, Industrial Robot, (MCB University Press), in press.
- [2] S.L. Albright, Calibration system for robot production control and accuracy, in R. Bernhardt and S. Albright, Robot Calibration, Eds. (Chapman & Hall, London, 1993), p. 37-56.
- [3] R. van Balen, D. Koelma, T.R. ten Kate, B. Mosterd, A.W.M. Smeulders, ScilImage: A Multi-layered Environment for Use and Development of Image Processing Software, in H.I. Christiansen and J.L. Crowley, Experimental Environments for Computer Vision and Image Processing, Eds. (World Scientific Publishing Co., Singapore, 1994) p. 107-126.
- [4] M. Born, E. Wolf, Principles of Optics, fourth edition, Pergamon Press, Oxford (1970).
- [5] Y. Chang and P. Liang, On Recursive Calibration of Cameras for Robot Hand-Eye Systems, Proceedings of the 1989 IEEE International Conference on Robotics and Automation, Arizona (1989) p. 838-843.
- [6] E. Hecht and A. Zajac, Optics, Addison-Wesley, Reading, Massachusetts (1974).
- [7] F.A. van de Heuvel, Automated 3-D measurement with the DCS200 digital camera, The Second Conference on 3D Optical Measurements Techniques, (Wichmann Verlag, Zurich, 1993) p.63-71.
- [8] W. Hoffmann, Basic Transformations in Linear Algebra for Vector Computing, PhD. Thesis, Amsterdam, May 1989.
- [9] J.S. Lee, Digital Image Smoothing and the Sigma Filter, Computer Vision, Graphics, and Image Processing 24, (1983) p. 255-269.
- [10] R.K. Lenz and R.Y. Tsai, Techniques for Calibration of the Scale Factor and Image Centre for High Accuracy 3D Machine Vision Metrology, Proceedings of the 1987 IEEE International Conference on Robotics and Automation, Raleigh, NC, March 31 - April 3 (1987) p. 68-75.

- [11] R.K. Lenz and R.Y. Tsai, A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration, *IEEE Transactions on Robotics and Automation*, Vol. 5, No. 3 (June 1989) p. 345-358.
- [12] R.Y. Tsai, A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses, *IBM Research Report RC 11413* (1985).
- [13] K. Schröer, Theory of kinematic modelling and numerical procedures for robot calibration, in R. Bernhardt and S. Albright, *Robot Calibration Eds.* (Chapman & Hall, London, 1993), p.157-193.
- [14] H.R. Schwarz, *Numerical Analysis, A Comprehensive Introduction*, (John Wiley & Sons, 1989).
- [15] P.W. Verbeek, H.A. Vrooman, L.J. van Vliet, Low-level Image Processing by Max-Min Filters, *Signal Processing* 15 (1988) p. 249-258.
- [16] S. Wolfram, *Mathematica Reference Guide*, Wolfram Research Inc., 1992.

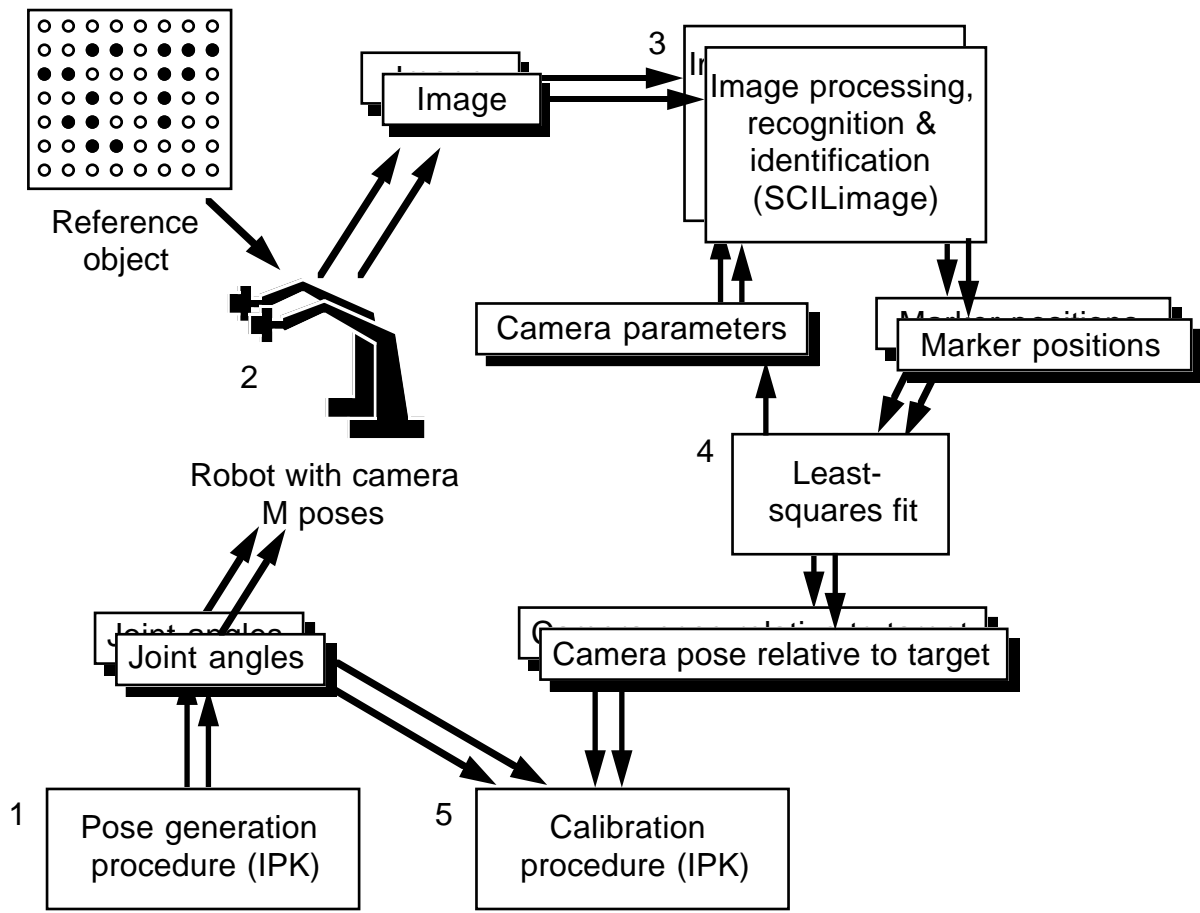
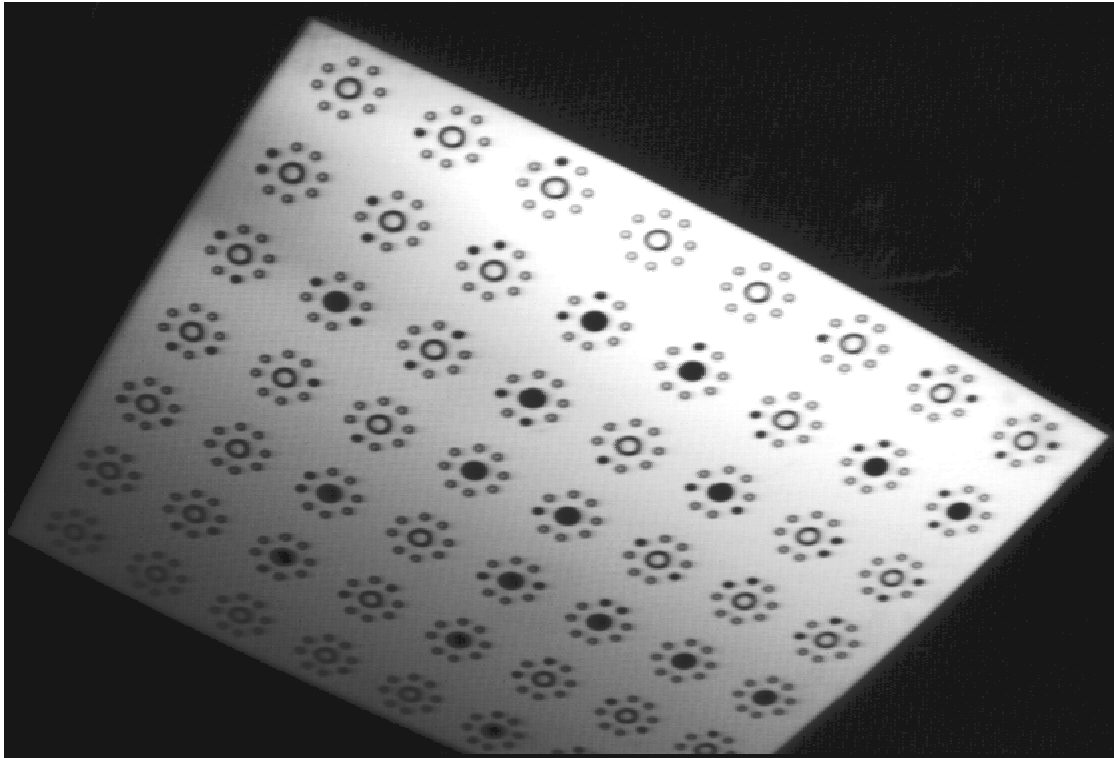
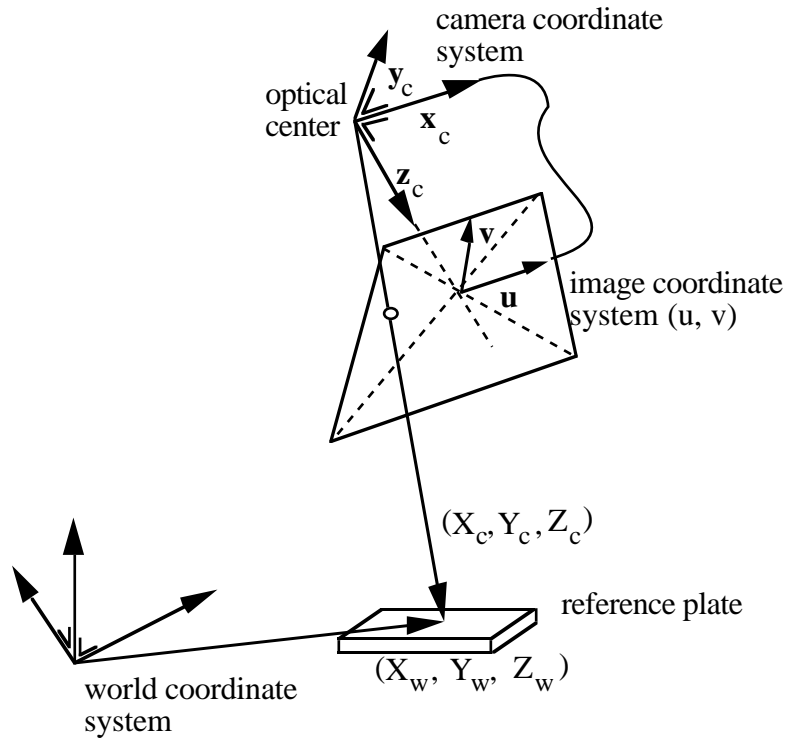


Figure 1. The relevant components and data used in the calibration procedure. The procedure starts with the generation of commanded joint angles for  $M$  measuring poses. At each of these  $M$  poses an image is obtained. The images are processed to compute the  $M$  poses actually attained, plus one common set of camera parameters.



*Figure 2. The reference plate as seen by the camera.*



*Figure 3. The geometry of the camera and the various coordinate systems. The  $z$ -axis of the camera coordinate system goes through the centre ( $u=0, v=0$ ) of the image. Vector  $\mathbf{u} = (u, 0)^T$  of the image coordinate system lies in the plane spanned by the  $X$ - and  $Z$ -axis of the camera coordinate system. Following common practice, the image plane is drawn between the object and the optical centre of the camera. All 3-D coordinate systems are right-handed.*

Formal position errors for 122 poses

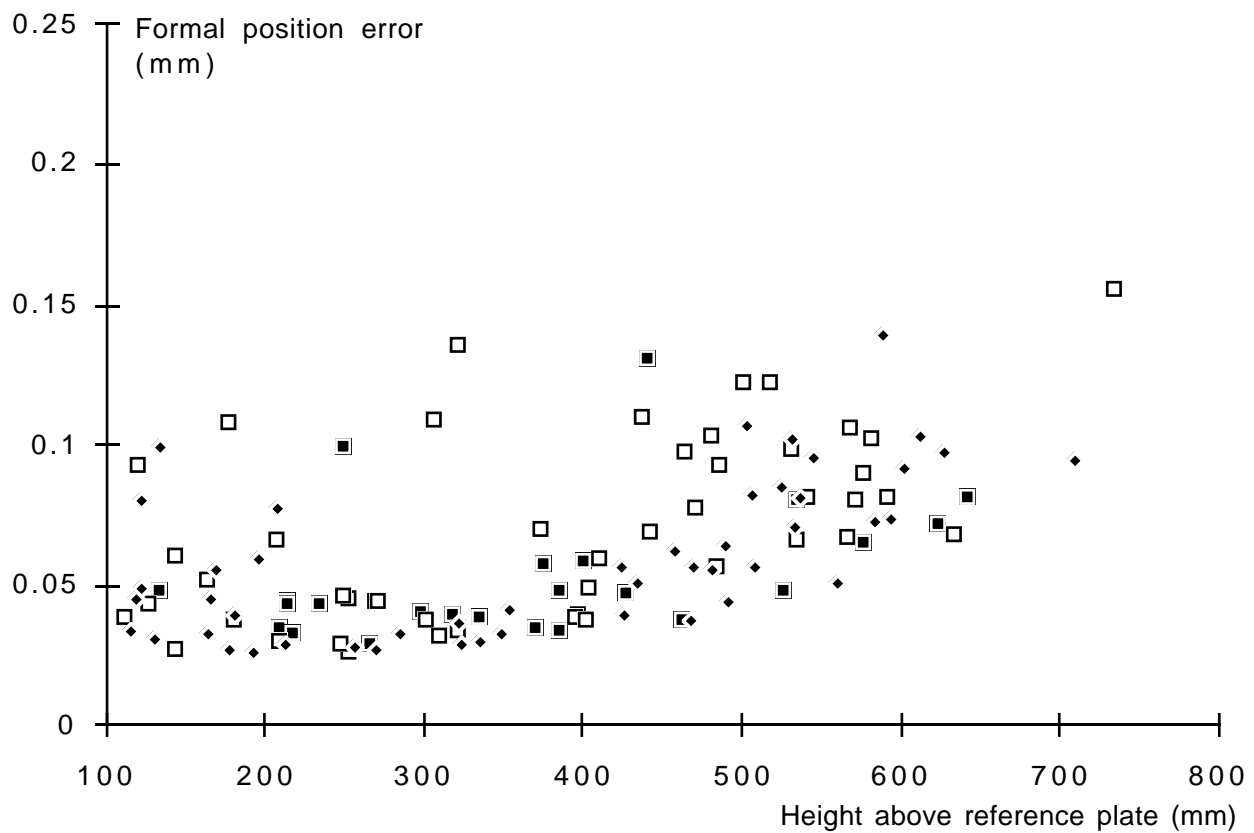


Figure 4. The estimated rms. position errors  $\sqrt{\Delta_x^2 + \Delta_y^2 + \Delta_z^2}$  for a collection of 122 images in three sets, as a function of the z-height of the camera. Notice that the position error increases with the height above the reference plate. For four images error estimates exceeding 0.25 mm were obtained.

Formal orientation errors for 122 poses

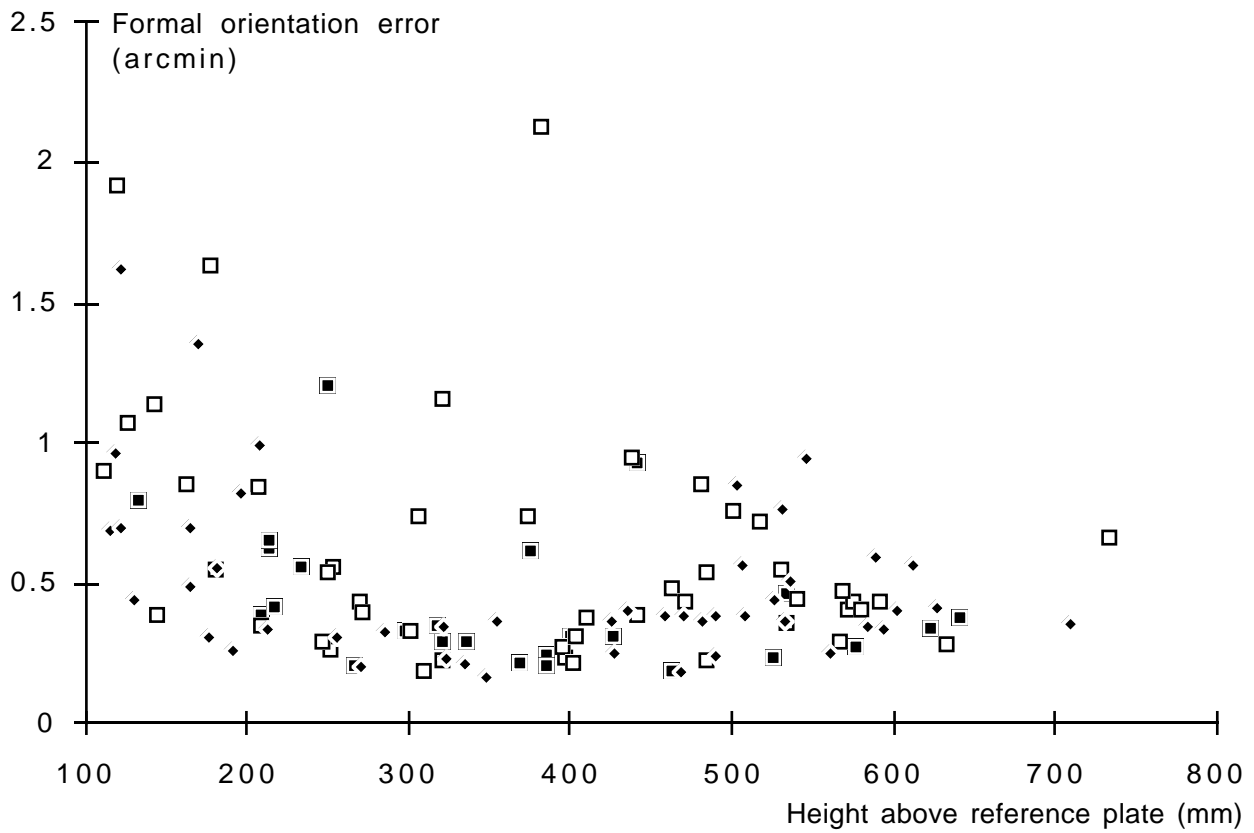


Figure 5. An estimate of the total orientation errors, computed as  $\sqrt{\Delta_{\alpha}^2 + \Delta_{\beta}^2 + \Delta_{\gamma}^2}$  for a series of 122 images in three sets, as a function of the z-height of the camera. For the range of z-heights shown, the orientation error is mostly independent of z. It tends to increase for small values of z, as fewer markers will be visible. For four images error estimates exceeding 2.5 arcminutes were obtained.

Position standard deviations for split groups

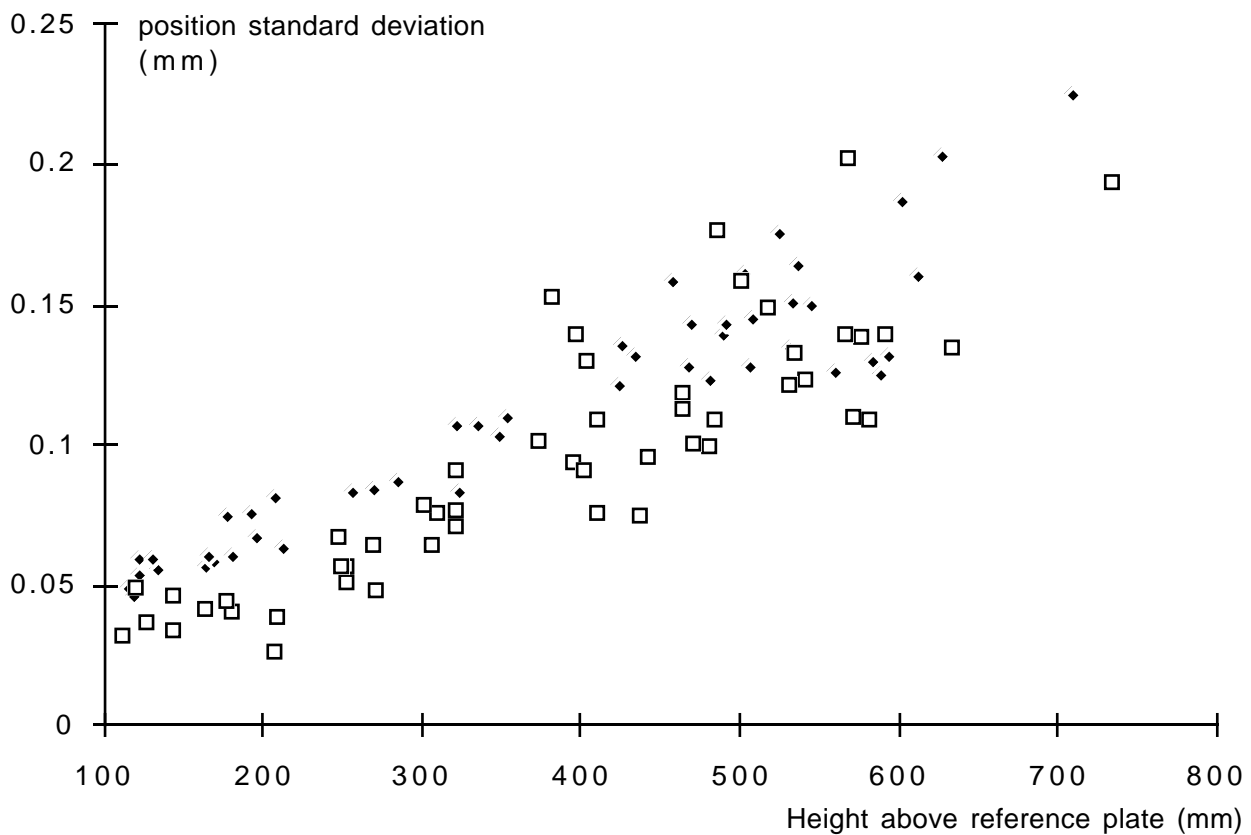


Figure 6. The standard deviations for the three positions computed from the images in the Left and Middle sets. Compare this figure with figure 4.