

# Simplified Intrinsic Camera Calibration and Hand-Eye Calibration for Robot Vision

Henrik Malm and Anders Heyden  
Centre for Mathematical Sciences  
Lund Institute of Technology  
S-221 00 Lund, SWEDEN  
email: henrik,heyden@maths.lth.se

**Abstract**—In this paper we investigate how intrinsic camera calibration and hand-eye calibration can be performed on a robot vision system using the simplest possible motions and a planar calibration object. The standard methods on plane-based camera calibration are extended with theory on how to use pure translational motions for the intrinsic calibration and we see how hand-eye calibration can be performed within the same framework. The calibration of two cameras in a stereo head configuration is shown to be an interesting application of the developed theory. Results of experiments on a real robot vision system are presented.

## I. INTRODUCTION

In the area of robot vision it is often the case that we want to attach a camera to the end-effector, or hand, of a robot so that the camera can work as visual sensor for the robot. By applying computer vision techniques this sensor can then be used to aid the robot in performing specific tasks such as finding and grasping objects. To be able to use the camera in this way we need to know the so-called intrinsic parameters of the camera. These include e.g. the focal length and the aspect ratio, i.e. the ratio between the width and the height of the imaging elements in the camera. The calculation of these parameters is called intrinsic camera calibration.

Other essential information that needs to be obtained is the orientation and position of the camera in relation to the robot hand. The transformation between the robot hand coordinate system and the camera coordinate system is called the hand-eye transformation and the task of calculating it is called hand-eye calibration.

Traditionally both the intrinsic camera calibration and the hand-eye calibration has been performed using a three-dimensional reference object, or calibration object, with known metric structure, often with some sort of grid pattern. See e.g. the book by Faugeras [2] for information on this kind of calibration. In recent years, so-called self-calibration methods has been common, cf. [4], [5]. These methods do not need a special calibration object and rely only on the rigidity of the scene in view. However, in applications requiring high precision measurements in performing their tasks, e.g. in most industrial vision systems, the use of a carefully constructed calibration grid is often most reliable. Since very accurate knowledge

of the relative 3D coordinates of points on the object is needed the construction of the grid is greatly simplified if a two dimensional planar object can be used. Zhang [13] and Sturm and Maybank [10] has independently developed principally identical algorithms for calibration from a planar object using two homogeneous linear constraints on the matrix describing the image of the so-called absolute conic, cf. [3]. These constraints arises from the estimated homography (2D projective transformation) from the object plane to the image plane at each position of the camera. By solving the linear system built up from these constraints and by a subsequent Cholesky factorization of the obtained matrix, the intrinsic parameters of the camera are obtained.

Concerning hand-eye calibration, early references on methods using a calibration object are Tsai and Lenz [12] and Shiu and Ahmad [9]. These methods are applicable both using a two- and three-dimensional calibration object since the calculation of the relative camera orientation and position, which is a subtask in these hand-eye calibration techniques, can be performed using any intrinsic camera calibration method which uses a calibration object. Methods for solving the hand-eye calibration problem without the use of a calibration object has also emerged. Three quite different approaches to this can be found in Andreff et.al. [1], Ma [6] and Malm et.al. [7]. The latter two of these papers also deals with the intrinsic calibration problem. However, as mentioned above, for high precision demands a method based on using a special calibration object is usually more reliable.

In this paper we will make extensions to the theory on plane-based intrinsic camera calibration and fit a hand-eye calibration algorithm into the same framework. Our ambition is to try to use as simple motions of the robot hand as possible so that the method can be applied to robot vision systems with a reduced number of degrees of freedom. For the intrinsic calibration and the orientation calculation in the hand-eye calibration we will show that it is sufficient to use images where the relative orientation between the object and camera does not change, i.e. the motions of the camera between obtaining the images are pure translations. This case is degenerate in the formula-

tions of Zhang [13] and Sturm and Maybank [10]. Using only the results in these two papers a new image does not give any new constraints on the calibration parameters if the motion has been purely translational. However, using the knowledge that the motion is a pure translation we can set up some additional constraints that fits nicely with the previously developed theory. It will be shown that two non-parallel translations will be sufficient to calculate the intrinsic parameters and the orientation of the camera relative to the robot hand.

The classic camera calibration paper by Tsai [11] also includes an algorithm for calibration from a planar object using translational motions. However, the technique presented here is simpler and more adaptable to different amounts of knowledge of the translational motion. Zhang gives a hint in his technical report [14] that an extension of the standard plane-based calibration technique to deal with translational motions is possible and that extension is fully developed here, producing new constraints that fits nicely with the constraints in the theory from [13] and [10]. The development of the theory in this paper was started in [8].

A special application of the calibration theory presented in this paper is the calibration of a stereo head configuration of two cameras. The fact that both cameras perform the same kind of translational motion can be easily used to make an accurate simultaneous calibration of the two cameras and to calculate the relative orientation and translation of the cameras.

For completion, we also discuss how to find the position part of the hand-eye calibration. To this end we have to use rotational motions of the robot hand. However, two simple rotations around e.g. the x- and z-axes in the robot hand coordinate system will be sufficient. This part of the calibration algorithm uses the pose of the camera in relation to the calibration object, which is obtainable from the estimated homographies after the camera has been calibrated.

We start in Section II by explaining the camera model in use and by briefly deriving the standard constraints for plane-based intrinsic calibration. Then, in Section III, the constraints on the intrinsic parameters appearing when using pure translational motions are presented. We continue in Section IV by discussing the special case of calibrating a stereo head using translational motions. In Section V we then present a simple and straightforward extension of the theory to cover also hand-eye calibration. In Section VI results of experiments on a real robot vision system are presented and the paper is closed in Section VII by some conclusions.

## II. PRELIMINARIES

The perspective pinhole camera is used as our projection model. That is,  $\lambda \vec{x} = P\vec{X}$ , where  $\vec{x} = (x, y, 1)$  is

the 2D homogeneous coordinates for the image point,  $\vec{X} = (X, Y, Z, 1)$  is the 3D homogeneous coordinates for the object point and  $\lambda$  is an arbitrary scale factor.  $P$  is the  $3 \times 4$  projection matrix that can be decomposed as

$$P = K[R|t] = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R|t] . \quad (1)$$

Here,  $\beta$  denotes the focal length,  $\frac{\alpha}{\beta}$  and  $\frac{\gamma}{\beta}$  the aspect ratio and the skew, respectively, and  $(u_0, v_0)$  the principal point. These are called the intrinsic parameters, and  $K$  is called the intrinsic camera matrix.  $R$  describes the orientation of the camera in relation to the world coordinate system and  $t$  the translation of the camera from the origin of this system, expressed in the coordinate system of the camera.

In this section we will concentrate on describing the constraints appearing when calibrating a camera from a planar object using general positions and orientations of the camera. The orientation and origin of the world coordinate system can be chosen so that the plane of the calibration object has  $Z = 0$ . We then get

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[r_1 \ r_2 \ r_3 \ t] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K[r_1 \ r_2 \ t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} , \quad (2)$$

where  $r_i$  is the  $i$ :th column of  $R$ . In this way the object point is related to the corresponding image point by a homography  $H$ :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} , \quad H = \frac{1}{\lambda} K[r_1 \ r_2 \ t] . \quad (3)$$

The homography  $H$  can be estimated for each image, cf. [13] for details. Let  $h_i$  be the  $i$ :th column in  $H$ . We then have

$$[h_1 \ h_2 \ h_3] = \frac{1}{\lambda} K[r_1 \ r_2 \ t] \quad (4)$$

and

$$r_1 = \lambda K^{-1} h_1 \quad (5)$$

$$r_2 = \lambda K^{-1} h_2 . \quad (6)$$

Introduce  $\omega = K^{-T} K^{-1}$ . Since  $r_1$  and  $r_2$  are orthonormal, the following constraints, involving  $h_1$ ,  $h_2$  and  $\omega$ , can be derived from (5) and (6).

$$r_1^T r_1 = \lambda^2 h_1^T K^{-T} K^{-1} h_1 = \lambda^2 h_1^T \omega h_1 = 1 \quad (7)$$

$$r_2^T r_2 = \lambda^2 h_2^T K^{-T} K^{-1} h_2 = \lambda^2 h_2^T \omega h_2 = 1 \quad (8)$$

$$r_1^T r_2 = \lambda^2 h_1^T K^{-T} K^{-1} h_2 = \lambda^2 h_1^T \omega h_2 = 0 \quad (9)$$

These equations could of course be simplified so that the unknown scale factor  $\lambda$  is excluded:

$$h_1^T \omega h_2 = 0 \quad (10)$$

$$h_1^T \omega h_1 = h_2^T \omega h_2 \quad (11)$$

The matrix  $\omega$  describes the image of the absolute conic, cf. [3], and we now have two linear constraints on this symmetric matrix from each different image of the plane. By using three different views of the plane, obtained using different orientations of the camera, we have enough constraints to solve for  $\omega$ . The intrinsic camera matrix  $K$  can then be obtained by Cholesky factorization and matrix inversion.

### III. INTRINSIC CALIBRATION FROM TRANSLATIONS

In this paper we concentrate on the case when the motion of the camera is pure translational between the images obtained. Let  $H'$  be the estimated homography from the object plane to the image plane after the translation. If  $H$ , in equation 3, is the estimated homography before the translation  $H'$  can be written as

$$H' = \frac{1}{\lambda'} K [r_1 \ r_2 \ t + R t'] = \quad (12)$$

$$= \frac{1}{\lambda'} K [r_1 \ r_2 \ t + t'_1 r_1 + t'_2 r_2 + t'_3 r_3] , \quad (13)$$

where  $t'$  is the translation vector expressed in the coordinate system of the calibration object, i.e. in the world coordinate system. The vector  $t$  is the translation between the camera and the calibration object in the initial position before the motion, expressed in the camera coordinate system. This vector will be cancelled out early in the calculations.

Note that  $K$  and  $R$  are unchanged from before the translation and that the first two columns in  $H'$  are parallel to the first two columns in  $H$ . If we have multiple images obtained with the same orientation of the camera we can estimate the homographies from these images simultaneously and set the first two columns equal for every estimated homography, i.e. the scale factors becomes equal,  $\lambda' = \lambda$ . This drastically reduces the number of parameters to be estimated and makes the preceding calculations more accurate. In the continuation there will thus only be one scale factor,  $\lambda$ , present.

Let us have a look at the third columns in the matrices  $H$  and  $H'$ , denoted by  $h_3$  and  $h'_3$ , respectively.

$$h_3 = \frac{1}{\lambda} K t , \quad (14)$$

$$h'_3 = \frac{1}{\lambda} K (t + t'_1 r_1 + t'_2 r_2 + t'_3 r_3) . \quad (15)$$

Let

$$\hat{h}_3 = h'_3 - h_3 . \quad (16)$$

Then, using (4),

$$\hat{h}_3 = \frac{1}{\lambda} K (t'_1 r_1 + t'_2 r_2 + t'_3 r_3) = t'_1 h_1 + t'_2 h_2 + \frac{1}{\lambda} K t'_3 r_3 \quad (17)$$

and subsequently

$$r_3 = \frac{\lambda}{t'_3} K^{-1} (\hat{h}_3 - t'_1 h_1 - t'_2 h_2) . \quad (18)$$

In search for new calibration constraints containing  $\hat{h}_3$  and  $t'$ , scalar products including  $r_3$ ,  $r_1$  and  $r_2$  are written down. Taking the scalar product of the orthonormal vectors  $r_1$  and  $r_3$  and using (7) and (9) gives

$$\begin{aligned} r_1^T r_3 &= \frac{\lambda^2}{t'_3} h_1^T K^{-T} K^{-1} (\hat{h}_3 - t'_1 h_1 - t'_2 h_2) \\ &= \frac{\lambda^2}{t'_3} (h_1^T \omega \hat{h}_3 - t'_1 h_1^T \omega h_1 - t'_2 h_1^T \omega h_2) \\ &= \frac{\lambda^2}{t'_3} (h_1^T \omega \hat{h}_3 - \frac{t'_1}{\lambda^2}) = 0 . \end{aligned}$$

So

$$h_1^T \omega \hat{h}_3 = \frac{t'_1}{\lambda^2} . \quad (19)$$

Similarly, the scalar product of  $r_2$  and  $r_3$  gives

$$h_2^T \omega \hat{h}_3 = \frac{t'_2}{\lambda^2} . \quad (20)$$

It remains to examine the scalar product of  $r_3$  with itself which should be equal to  $|r_3|^2 = 1$ ,

$$\begin{aligned} r_3^T r_3 &= \frac{\lambda^2}{t'^2_3} (\hat{h}_3^T - t'_1 h_1^T - t'_2 h_2^T) \omega (\hat{h}_3 - t'_1 h_1 - t'_2 h_2) \\ &= \frac{\lambda^2}{t'^2_3} (\hat{h}_3^T \omega \hat{h}_3 - \frac{t'^2_1}{\lambda^2} - \frac{t'^2_2}{\lambda^2}) = 1 . \end{aligned}$$

This gives

$$\hat{h}_3^T \omega \hat{h}_3 = \frac{|t'|^2}{\lambda^2} . \quad (21)$$

By letting  $\bar{\omega} = \lambda^2 \omega$ , the complete set of constraints arising from two images of a plane when the camera undergoes translation  $t'$  looks like

$$h_1^T \bar{\omega} h_2 = 0 \quad (22)$$

$$h_1^T \bar{\omega} h_1 = 1 \quad (23)$$

$$h_2^T \bar{\omega} h_2 = 1 \quad (24)$$

$$h_1^T \bar{\omega} \hat{h}_3 = t'_1 \quad (25)$$

$$h_2^T \bar{\omega} \hat{h}_3 = t'_2 \quad (26)$$

$$\hat{h}_3^T \bar{\omega} \hat{h}_3 = |t'|^2 \quad (27)$$

Since the first two columns in  $H$  are equal to the first two columns in  $H'$  and since the scale factor  $\lambda$  is unknown we get 11 known distinct elements from the two estimated homographies. There are 6 degrees of freedom for the pose of the camera in the first image. Therefore, there are  $11 - 6 = 5$  degrees of freedom left for the intrinsic parameters and the translational motion. That is, if we want to calculate all the 5 intrinsic parameters from one translational motion  $t'$  we need to know both the length

and the direction of the translation. If we e.g. set the skew  $s = 0$ , the camera can be calibrated when the length of the translation is unknown. If set  $s = 0$  and the aspect ratio  $\gamma = 1$ , we can calibrate the camera when the direction of the translation is unknown but the length is known.

After solving the system (22)-(27) of equations in the unknowns of  $\bar{\omega}$ ,  $t'_1$ ,  $t'_2$  and  $|t'|$  we perform a Cholesky factorization of  $\bar{\omega}$ . Inverting and scaling the resulting matrix gives us the intrinsic calibration matrix  $K$ . The scale factor  $\lambda$  is easily found since  $K$  should have a 1 in the bottom right position.

In Section V-A we are going to calculate the orientation of the camera in relation to the robot hand. To this end a minimum of two translations of the robot hand are needed. Let  $t''$  be a second translation vector. From this translation we get a new set of equations of the form (25)-(27). If  $\hat{h}_3$  is the equivalent to  $\tilde{h}_3$  for the second translation we, for example, get the additional constraint

$$\hat{h}_3^T \bar{\omega} \hat{h}_3 = |t''|^2 . \quad (28)$$

Consequently, knowing the length of two translations we get five constraints on  $\bar{\omega}$  and the camera can be calibrated up to skew  $s = 0$ . If we, in a robot vision system, can control the motion of the robot hand it is a reasonable assumption that the length of the translations in the robot hand coordinate system are known. These lengths are equal to the lengths of the translations,  $t'$  and  $t''$ , in the coordinate system of the calibration object.

#### IV. CALIBRATION OF A STEREO HEAD

One useful application of the theory presented in this paper is the calibration of a stereo head configuration of two cameras. If we translate a stereo head, the translation vector  $t'$  will be the same for both of the cameras in the coordinate system of the calibration object. This can be used to make a simultaneous calibration of the two cameras.

Let a tilde represent the entities corresponding to the second camera in the stereo head. Suppose we know the length of the translation of the stereo head. Then the following equations can be used to calibrate the cameras up to skew  $s = 0$ . Observe that we here will remove the bar over  $\omega$  to make the equations more readable, but  $\omega$  and  $\tilde{\omega}$  will still include two (generally different) unknown

scale factors.

$$h_1^T \omega h_2 = 0 \quad (29)$$

$$h_1^T \omega h_1 = 1 \quad (30)$$

$$h_2^T \omega h_2 = 1 \quad (31)$$

$$\tilde{h}_1^T \tilde{\omega} \tilde{h}_2 = 0 \quad (32)$$

$$\tilde{h}_1^T \tilde{\omega} \tilde{h}_1 = 1 \quad (33)$$

$$\tilde{h}_2^T \tilde{\omega} \tilde{h}_2 = 1 \quad (34)$$

$$h_1^T \omega \hat{h}_3 - \tilde{h}_1^T \tilde{\omega} \hat{h}_3 = 0 \quad (35)$$

$$h_2^T \omega \hat{h}_3 - \tilde{h}_2^T \tilde{\omega} \hat{h}_3 = 0 \quad (36)$$

$$\hat{h}_3^T \omega \hat{h}_3 = |t'|^2 \quad (37)$$

$$\hat{h}_3^T \tilde{\omega} \hat{h}_3 = |t'|^2 \quad (38)$$

Another situation could be that we know the direction of the translation but the length of the translation is unknown. For example, the calibration plane could be placed on the floor and the stereo head be translated orthogonal to the floor ( $t' = (0,0,1)$ ). The last four equations in the last listing should then be replaced by

$$h_1^T \omega \hat{h}_3 = t'_1 \quad (39)$$

$$\tilde{h}_1^T \tilde{\omega} \hat{h}_3 = t'_1 \quad (40)$$

$$h_2^T \omega \hat{h}_3 = t'_2 \quad (41)$$

$$\tilde{h}_2^T \tilde{\omega} \hat{h}_3 = t'_2 \quad (42)$$

$$\hat{h}_3^T \omega \hat{h}_3 - \hat{h}_3^T \tilde{\omega} \hat{h}_3 = 0 \quad (43)$$

After calculation of the camera matrices  $K$  and  $\tilde{K}$ , the pose of the cameras, i.e. the orientation and position of the cameras in relation to the calibration object, can easily be obtained from relation (4). (The scale factors  $\lambda$  and  $\tilde{\lambda}$  are obtained during the calculation of  $K$  and  $\tilde{K}$ .) From this the relative rotation and the relative translation between the two cameras in the stereo head can be obtained and the stereo head is then fully calibrated with respect to the intrinsic parameters and the relative pose. This can be done from one translation, minimally.

#### V. EXTENSION TO HAND-EYE CALIBRATION

##### A. Calculation of Camera Orientation

After calculating the intrinsic camera matrix  $K$  using two translations of the robot hand, it is an easy task to obtain the orientation of the the camera relative to the robot hand coordinate system. Since the directions of the translations in the robot hand coordinate system are known, we just have to find the corresponding directions in the camera coordinate system to be able to calculate the rotational transformation  $R_{he}$  between the robot hand and the camera. If  $t_c$  is a translation vector expressed in the camera coordinate system, the estimated homography after the translation,  $H'$ , has the following form

$$H' = \frac{1}{\lambda} K [r_1 \ r_2 \ t + t_c] . \quad (44)$$

Using the vector  $\hat{h}_3$  introduced earlier, cf. (16), we get

$$\hat{h}_3 = \frac{1}{\lambda} K t_c . \quad (45)$$

When calculating the translation directions in the camera coordinate system the length of the translation vectors are not of importance and  $t_c$  can be written as

$$t_c \sim K^{-1} \hat{h}_3 , \quad (46)$$

where  $\sim$  means ‘‘proportional to’’. Let  $t'_c$  and  $t''_c$  be two translation directions in the camera coordinate system and  $t'_r$  and  $t''_r$  be the corresponding translations in the robot hand coordinate system. By constructing a third corresponding pair using  $t'''_c = t'_c \times t''_c$  and  $t'''_r = t'_r \times t''_r$  and normalizing all the vectors we can calculate  $R_{he}$  for example using

$$R_{he} = [t'_c \quad t''_c \quad t'''_c] [t'_r \quad t''_r \quad t'''_r]^{-1} . \quad (47)$$

Due to errors in the calculation of  $t'_c$  and  $t''_c$ , the matrix  $R_{he}$  will not be exactly orthogonal. To find the nearest orthogonal matrix in the Frobenius norm we can do a singular value decomposition of  $R_{he}$ ,  $R_{he} = USV^T$ , and choose  $\bar{R}_{he} = UV^T$  as our rotation matrix.

### B. Calculation of Camera Position

This paper primarily deals with the problem of calibrating a single camera or a stereo head in a robot hand-eye configuration using translational motions. However, to calculate the position of the camera in relation to the robot hand motions also including a rotational part have to be used. To make the presentation on hand-eye calibration complete we here include a discussion on how the position of the camera can be obtained after the camera has been calibrated using the techniques described in the earlier sections. In our quest to use as simple motions and few degrees of freedom as possible, we choose to use pure rotations of the robot hand. The theory presented is the standard formulation of calculating the hand-eye translation from e.g. [12] and [9] applied to the special case of pure rotational motions.

Let

$$T_{he} = \begin{bmatrix} R_{he} & t_{he} \\ 0_{1 \times 3} & 1 \end{bmatrix} \quad (48)$$

be the hand-eye transformation. Here,  $t_{he}$  is the translation vector that determines the position of the camera in relation to the robot hand. It is this the vector that we want to calculate. Let

$$T_c = \begin{bmatrix} R_c & t_c \\ 0_{1 \times 3} & 1 \end{bmatrix} \quad (49)$$

be the transformation of the camera pose under the rotational motion of the robot hand. This transformation can be calculated since the pose of the camera in relation to the calibration plane can be obtained directly from the

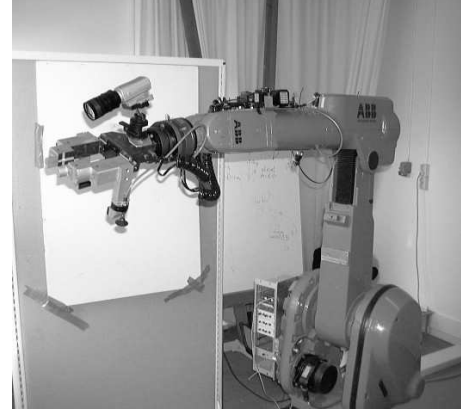


Fig. 1. The robot used in the experiments.

estimated homographies if the camera has been calibrated, cf. equations (4)-(6). Let

$$T_r = \begin{bmatrix} R_r & 0_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \quad (50)$$

be the known transformation of the robot hand. Note that  $T_r$  is a pure rotation. It is easy to convince oneself that the following relation must hold for the current motion of the robot hand, cf. [9].

$$T_r T_{he} = T_{he} T_c . \quad (51)$$

Performing the matrix multiplication and examining the rightmost columns, i.e. the translation part, in the transformation matrices on both sides of the equality gives the following relation

$$R_r t_{he} = R_{he} t_c + t_{he} . \quad (52)$$

Rewriting this as

$$(R_r - I) t_{he} = R_{he} t_c \quad (53)$$

gives a linear system in the unknown translation vector  $t_{he}$ . Unfortunately, since the  $3 \times 3$ -matrix  $R_r - I$  has at most rank 2, we cannot solve for  $t_{he}$  uniquely using one rotation but by intuition this was not really to expect. However, using two different pure rotations of the robot hand gives us enough equations to solve for  $t_{he}$ . For example, one rotation around the x-axis and one rotation around the z-axis in the robot hand coordinate system should be sufficient.

## VI. EXPERIMENTS

The calibration algorithms presented in this paper has been tested on a real robot vision system. The robot used is an ABB IRB2003, cf. Figure 1, which is capable of moving in all six degrees of freedom. The camera was mounted by a ball head camera holder on the gripper (or hand) of the robot so that the orientation of the camera

could be changed with respect to the orientation of the robot hand coordinate system. Figure 2 shows an image from one of the used image sequences picturing the calibration plane. In the following experiments approximately ten images were used in each sequence.

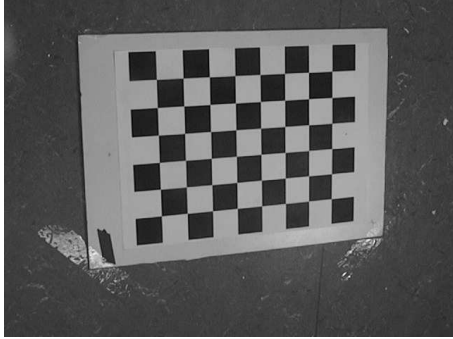


Fig. 2. An image from one of the sequences used.

Experiments were performed for both a single camera and a stereo head. In the stereo head case the head was simulated by making two identical translational motions starting from two different points in space using different orientations and settings of the camera. The different camera settings will, for simplicity, be referred to as camera 1 and camera 2 in this section.

As a reference the cameras was calibrated using the plane-based calibration technique by Zhang [13] for general motions. The intrinsic calibration matrices  $K$  and  $\tilde{K}$  for camera 1 and camera 2, respectively, calculated using this method was

$$K_{Zhang} = \begin{bmatrix} 1063.1 & -6.2 & 328.9 \\ 0 & 1094.0 & 247.9 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$\tilde{K}_{Zhang} = \begin{bmatrix} 1089.6 & -9.9 & 335.1 \\ 0 & 1126.4 & 231.9 \\ 0 & 0 & 1 \end{bmatrix}.$$

To test the practical performance of the theory presented in this paper camera 1 was first translated straight down orthogonally to the floor, i.e.  $t' = (0, 0, 1)$ . By using equations (22) to (26) the camera was calibrated with the skew set to zero,  $s = 0$ . The following intrinsic camera matrix was obtained

$$K_{(0,0,1)} = \begin{bmatrix} 1019.4 & 0 & 339.3 \\ 0 & 1047.8 & 235.5 \\ 0 & 0 & 1 \end{bmatrix}.$$

The value on the focal length seems to be little bit underestimated compared to  $K_{Zhang}$ . However, the result using Zhangs method isn't the ground truth and the result obtained here probably still falls within the error bounds of that method.

Camera 1 was also calibrated using a motion sequence of the camera where the translation direction was changed after each obtained image. Here we assumed that the translation directions were unknown and instead used a fixed length of the translations, i.e. equations (22) to (24) and multiple instances of (27) was used for the calculation. The following matrix was obtained

$$K_{change} = \begin{bmatrix} 1030.7 & 0 & 337.4 \\ 0 & 1076.5 & 236.9 \\ 0 & 0 & 1 \end{bmatrix}.$$

$K_{change}$  is a bit closer to  $K_{Zhang}$  than  $K_{(0,0,1)}$  which might indicate a slightly more accurate result.

We also calibrated the stereo head configuration using the theory presented in Section IV. The camera pair was translated orthogonal to the floor, but this direction was assumed to be unknown. Instead we used a known length of the translations and applied equations (29) to (38). The following two matrices were calculated:

$$K_{stereo} = \begin{bmatrix} 1058.7 & 0 & 353.2 \\ 0 & 1092.8 & 231.4 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$\tilde{K}_{stereo} = \begin{bmatrix} 1091.8 & 0 & 310.1 \\ 0 & 1125.1 & 252.9 \\ 0 & 0 & 1 \end{bmatrix}.$$

The result is amazingly close to the reference method, where general motions of the cameras were used, at least w.r.t the focal length and the aspect ratio. The principal point is a little bit further off, but still acceptable.

When, as in this last experiment, the length of the translations is known and it is known that the translation direction is unchanged through the whole image sequence additional constraints can be put on the homography estimation. Actually, by using the known translation lengths in the estimation only 11 parameters in total needs to be estimated for all the homographies together. By putting as much knowledge into the homography estimation as possible the calibration algorithm becomes more stable and accurate.

The relative orientation and translation of the cameras in the stereo head configuration was also calculated following the discussion at the end of Section IV. Also, the rotational part of the hand-eye transformation was calculated for camera 1, as discussed in Section V-A, by using the motion sequence where the translational direction was changed after each obtained image. The results has unfortunately not been compared to a reference method but seemed reasonable by visual inspection.

## VII. CONCLUSIONS

We have in this paper presented extensions to the standard theory on intrinsic camera calibration and hand-eye

calibration using a planar calibration object. The vision system considered is a robot with a camera attached to its end-effector, or hand, of the robot arm.

The main objective has been to develop algorithms that uses as simple motions of the robot hand as possible and put small demands on the degrees of freedom of the robot vision system. The resulting theory shows that we can calculate the intrinsic parameters of the camera using a single translational motion and a planar calibration object if we know that the motion is pure translational and have some additional information about the translation, such as its length or direction. If two translations are used, the intrinsic parameters and the orientation of the camera in relation to the robot hand can be calculated if we know the directions of the translations in the robot hand coordinate system.

Using two additional rotational motions of the robot hand the position of the camera in relation to the robot hand can also be calculated. These rotations can for example be chosen as simple rotations around the x- and z-axis in the robot hand coordinate system.

The two cameras in a stereo head configuration can be calibrated simultaneously up to skew  $s = 0$  and the relative translation and rotation of the two cameras can be obtained using a single translational motion if we either know the length or the direction of the translation.

An interesting setup is when the calibration plane is placed on the floor and the robot hand is translated orthogonally to the floor. Since we know the direction of the translation in this case we do not need the knowledge of the length of the translation for the calibration. This is one of the situations that has been successfully tested experimentally. Experimental results using a couple of different configurations and degrees of knowledge has shown that the presented algorithms for calibration from translational motions gives similar results to the standard calibration techniques using general motions of the camera.

Further studies on the topic of calibration using translational motions and a planar calibration object should include a complete discussion on critical configurations. The calibration plane could not, for example, be parallel to the image plane for the algorithm to work. Also, motion directions parallel to the plane will not give any additional information on the calibration since this can be viewed as only extending the area of the plane in space.

#### VIII. ACKNOWLEDGMENTS

This work was sponsored by the Swedish Research Council, project 2000-530, SSF Junior Individual Grant, project 99-241 and The Royal Academy of Science. The authors would like to thank Tomas Olsson for helping with the experiments.

#### IX. REFERENCES

- [1] N. Andreff, R. Horaud, B. Espiau, "Robot Hand-Eye Calibration using Structure from Motion" *Int. J. Robotics Research*, vol. 20, no. 3, pp. 228-248, March 2001.
- [2] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.
- [3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [4] A. Heyden and K. Åström, "Euclidean Reconstruction from Constant Intrinsic Parameters", *Proc. Int. Conf. Pattern Recognition*, vol. 1, pp. 339-343, August 1996.
- [5] Q.-T. Loung and O. Faugeras, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices", *In. J. Computer Vision*, vol. 22(3), pp. 261-289, 1997.
- [6] S.D. Ma, "A Self-Calibration Technique for Active Vision Systems", *IEEE Trans. Robotics and Automation*, Vol. 12, pp. 114-120, February 1996.
- [7] H. Malm and A. Heyden, "Self-calibration from Image Derivatives for Active Vision Systems", *Proc. Int. Conf. Automation, Robotics and Computer Vision*, Singapore, December 2002.
- [8] H. Malm and A. Heyden, "Plane-Based Calibration: The Case of Pure Translation", *Proc. Int. Workshop on Machine Vision Applications*, Nara, Japan, December 2002.
- [9] Y.C. Shiu and S. Ahmad, "Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form  $AX = XB$ ", *IEEE Trans. Robotics and Automation*, Vol 5., pp. 16-29, 1989.
- [10] P. Sturm and S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications", *Proc. Conf. Computer Vision and Patterns Recognition*, Vol 1., pp. 432-437, 1999.
- [11] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-self tv cameras and lenses", *J. Robotics and Automation*, vol. 3(4), pp. 323-344, August 1987.
- [12] R.Y. Tsai and R.K. Lenz, "A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration", *IEEE Trans. Robotics and Automation*, vol. 5, pp. 345-358, 1989.
- [13] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", *Proc. Int. Conf. Computer Vision*, vol. 1, pp. 666-679, September 1999.
- [14] Z. Zhang, "A flexible new technique for camera calibration", *Technical Report MSR-TR-98-71*, Microsoft Research, December 1998.