Kinematic Calibration for Saccadic Eye Movements

Kai Welke, Markus Przybylski, Tamim Asfour and Rüdiger Dillmann
Universität Karlsruhe (TH), Institute for Anthropomatics, HIS
P.O. Box 6980, 76128 Karlsruhe, Germany

{welke,przybyls,asfour,dillmann}@ira.uka.de

***Abstract*— The visual perception system of humanoid robots should provide sensorial information that fulfills the requirements imposed by perceptual tasks in natural environments. One prerequisite for such systems is the ability to observe the environment by actively moving its visual sensors. This ability allows to implement two essential behaviors for a cognitive visual system: smooth pursuit and saccadic eye movement.**

In the work presented in this paper we propose a kinematic calibration approach for the active camera system of the Karlsruhe Humanoid Head. The proposed method solves two fundamental problems when performing saccadic eye movements: the required kinematic model for open-loop control and the ability of stereo reconstruction with active cameras. We present experiments on the accuracy of the kinematic model, the stereo triangulation and the saccadic eye movement.

I. INTRODUCTION

Most current humanoid robots have simplified head-eye systems with a small number of degrees of freedom (DoF). The heads of ASIMO [1], HRP-3 [2] and HOAP-2 [3] have two DoF and fixed eyes. However, humanoid systems that are able to execute manipulation and grasping tasks, that interact with humans and learn from human observation require sophisticated perception systems, which are able to fulfill the therewith associated requirements. The Karlsruhe Humanoid Head [4] (see fig.1) used in this work offers an active vision system with two cameras, which can be moved independently. This ability is usually exploited to implement behaviors that are essential to common visual perception tasks, namely smooth pursuit and saccadic eye movements. A similar active head is used on the humanoid robot CB [5] for which a calibration procedure has been proposed in [6] to enable 3D vision and foveation.

The smooth pursuit behavior (also called fixation or tracking) consists in focusing on a known combination of visual stimuli, e.g. the face of a person during interaction. The common control strategy deployed in such behaviors is closed-loop control, where the prior knowledge of the visual stimuli is exploited (see e.g. [7]).

Saccadic eye movements play an important role in the serialisation of visual information processing within a perceived environment. Saccadic eye movements are usually initiated by attention mechanisms, in order to focus on salient parts of the scene. In such mechanisms, the target of the saccade is determined by the spike of a single neuron ([8],[9]). The information necessary for closed-loop control is not available. Consequently, open-loop control strategies have to be provided in order to execute saccadic eye movements.

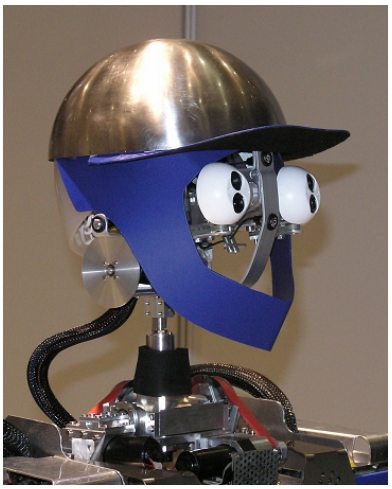

Fig. 1. The Karlsruhe Humanoid Head offers an active camera system.

Another problem arising from the application of active stereo camera systems is the ability of performing stereo reconstruction. When working with fixed eyes, the stereo calibration required for stereo triangulation is fixed and only has to be calculated once. In contrast the calibration for active stereo camera systems varies with each actuation of the eyes thus making static stereo calibration impossible.

In this work we present a new method for kinematic calibration of an active camera system which solves the problem of open-loop control for saccadic eye movements as well as the stereo calibration problem with actuated cameras. The calibration procedure results in a kinematic model which allows to solve the inverse kinematics problem for the eye system required for saccadic eye movements as well as the calculation of the stereo calibration required for stereo vision.

The paper is organized as follows. In Section II an overview of the different approaches for the problem classes of head-eye and hand-eye calibration in the literature is given and the approaches are analyzed according to their feasibility for our problem. In Section III a brief description of the system configuration is provided. Section IV describes the proposed approach for kinematic calibration. In Section V, the proposed method is evaluated on the Karlsruhe Humanoid Head. We provide experiments on the accuracy of the kinematic model itself, the accuracy of the stereo calibration and the accuracy of saccadic eye movements.

II. Related Work

The literature offers a variety of methods to solve the two related problems of head-eye and hand-eye calibration. Most of them are based on the traditional $AX = XB$ and $AX = ZB$ formulations of correspondences between coordinate frames in the system to be calibrated.

The $AX = XB$ formulation arises from the problem of head-eye calibration. In this case A denotes the coordinate transformation between two distinct camera positions when moving the camera by a transformation B of a joint. X denotes the unknown relationship from the actuated joint to the camera which is to be determined.

The $AX = ZB$ formulation has been proposed for the case of hand-eye calibration problems. Here A describes the transformation from the camera to the world frame. B denotes the transformation between the robot's hand coordinate frame and the base coordinate frame. The unknowns to be determined are the hand-to-camera transformation X and the base-to-world transformation Z.

Considering different methods which make use of the $AX = XB$ and $AX = ZB$ formulations, there are two essential decisions to be made when developing a method based on them. First, there are different possibilities to model the rotational parts of the involved coordinate transformations. The proposed models comprise Euler angles and quaternions as well as representations using a rotation axis and an angle or such based on Lie theory. The second decision concerns the mathematical method, which is used to actually find a solution for the unknown coordinate transformation, given its representation. Most solutions are based on linear or non-linear least squares optimization methods. Other approaches suggest the use of Lagrange multipliers or avoid any kind of optimization.

Some of the most important contributions in the field can be categorized in the following way. Tsai and Lenz [10] use an axis-and-angle representation for the rotation matrices. They solve separately for rotation and translation and present a linear least squares solution for both. Shiu and Ahmad [11] use a similar representation as Tsai and Lenz but develop a different linear solution. Li's method ([12],[13]) uses rotation matrices to model the problem. In his experiments, these matrices are based either on Euler angles or quaternions. He uses a non-linear optimization approach for the rotational part and a linear least squares approach for the translational part. The method by Neubert and Ferrier [14] uses Lie theory to model the problem and solves simultaneously for rotation and translation using a linear least squares approach. Horaud and Dornaika [15] present two methods, both of them based on quaternion representations. One is a closed-form approach using Lagrange multipliers. The other one is a non-linear least squares approach which solves for all unknowns at once. A completely different approach is presented by Young [16]. He does not refer to the $AX = XB$ and $AX = ZB$ formulations. Instead he uses a combination of a modified Denavit-Hartenberg (DH) convention and screw theory. No optimization is required. The method calibrates one joint at a time and can be used for any type of kinematic chain.

There are several papers which provide comparisons of the mentioned approaches to determine which method produces the most accurate results. Horaud and Dornaika [15] compared a linear and a non-linear least squares approach as well as an approach using Lagrange multipliers. Li [13] compared linear least squares approaches by Dornaika [17] and Tsai [10] with his own non-linear least squares approach. The results can be summarized as follows. The representation of the unknown rotation does not seem to be essential. In comparison the choice of the method to solve for the unknowns has a more significant impact on calibration precision. Non-linear least squares methods yield the most accurate results which is attributed to the degrading performance of linear least squares methods in the presence of noise. The downside of approaches that solve separately for rotation and translation is the fact that in these two-step methods the error propagates from the first part to the second part. Therefore it is reasonable to estimate all unknowns simultaneously.

Consequently, an advantage of the $AX = XB$ and $AX = ZB$ formulations is the fact that one does not have to resort to sophisticated representations of rotations. Simple representations like Euler angles provide very good results. However, there are some serious disadvantages. As proved several times ([10],[11],[18]), the equations $AX = XB$ and $AX = ZB$ have two degrees of freedom. A unique solution can only be found if two rotations around non-parallel axes of rotation are performed. This means that single joints with only one degree of freedom can not be calibrated. Each joint must have at least two degrees of freedom. An elegant solution to this issue is to combine two joints with one degree of freedom each and to treat them as one single joint with two degrees of freedom. Although most authors do not explicitly state it, this is only possible if the axes of the two respective joints intersect, because only this way a single common coordinate frame for both joints can be assigned to the point of intersection. Therefore the class of kinematic chains that can be calibrated using this approach is restricted. But even kinematic chains which according to the design schematics fulfill this condition may, due to production imprecisions, in practice not be accurate enough. In this case the simplifying assumption of intersecting joint axes is in fact a methodical error, resulting in inferior calibration accuracy.

The approach by Young [16] is more universal. Based on the Denavit-Hartenberg convention, it can be applied to any kind of kinematic chain. No simplifying assumptions are made. Therefore it should yield more accurate results than the approaches described above. Moreover, no optimization of any kind is used. However, the Denavit-Hartenberg convention always assigns the z axes to the axes of rotation or translation, which might not always be desirable.

In this paper an approach is suggested that combines the advantages of $AX = XB$ based methods and a DH-based approach. In contrast to the DH convention, the rotation axes are not necessarily assigned to the z axes which allows to choose

arbitrary coordinate frames for each joint. No simplifying assumptions are made concerning the relationships between adjacent joints. The proposed method avoids to consider two or more distinct joints as one joint with multiple degrees of freedom. Instead, every single joint is calibrated separately. That way the proposed approach can be applied to a wider class of kinematic chains. An $AX = XB$ based formulation is used to derive a non-linear target function that is minimized using the method by Levenberg and Marquardt [19]. For each joint to be calibrated, all necessary unknowns are estimated simultaneously, avoiding error propagation between them.

III. System Configuration

The Karlsruhe Humanoid Head used for our work has seven rotational degrees of freedom (see Fig.1). Four of them are used to move the head: neck roll, neck yaw, neck pitch and head tilt. The eyes are actuated by a common tilt joint and two independent pan joints. The vision system consists of two stereo camera pairs and mimics the foveated structure of the human eye. Therefore each eye contains one perspective camera with a wide angle of view and a foveal camera with a small angle of view. In both cases Dragonfly cameras from PointGrey are mounted which are accessed via an IEEE1394 interface and provide a maximum frame rate of 30 fps at a resolution of 640×480 pixels [1]. For the experiments presented in this paper the perspective cameras were used, which were outfitted with lenses with a focal length of 4 mm.

IV. Kinematic Calibration

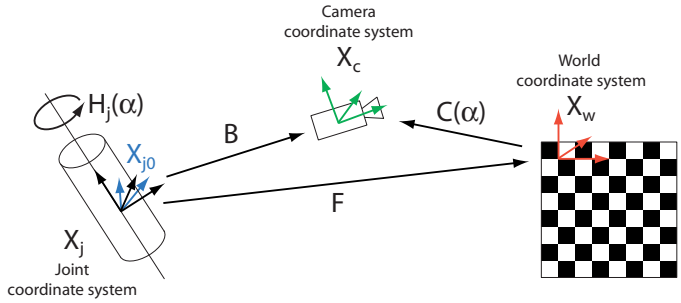

Fig. 2. Coordinate systems and transformations involved in the kinematic calibration procedure.

In this section, a detailed formal description of the proposed method is given. We proceed in the following way:

First, all involved coordinate systems and transformations are introduced. Then the acquisition of data for the necessary extrinsic camera calibration is explained. Based on this data the kinematic calibration problem is formulated as a non-linear least squares problem, which is solved using the method by Levenberg and Marquardt. Having determined the kinematic calibration of both the left and right eye pan joints, a stereo calibration for arbitrary angles of these joints can be calculated.

[1] http://www.ptgrey.com

A. Coordinate Systems and Transformations

Throughout this paper, the following conventions for coordinate systems and transformations will be used (see figure 2):

- X_w denotes the world coordinate system. It is a fixed system which is used as a reference to determine the extrinsic camera calibrations. The world coordinate system is defined by the calibration pattern.
- X_{j_0} is the coordinate system in the joint j at zero position. It is fixed.
- X_j denotes the rotated coordinate system of the joint j after actuation.
- X_c is the camera coordinate frame. As the camera is rigidly attached to the pan joint, the camera system moves when the joint is actuated.
- $H_j(\alpha)$ describes the coordinate transformation from the fixed joint coordinate frame X_{j_0} to the rotated joint coordinate frame X_j. As the considered joints have one degree of freedom, $H_j(\alpha)$ describes a rotation around the one rotation axis of the respective joint by an angle α which is obtained from the encoder readings.
- B denotes the transformation from the rotated joint coordinate system X_j to the camera coordinate system X_c. As the joint movement is already described by the transformation H_j, B remains constant, independent of the actual joint position. The goal of the calibration process is to determine this transformation.
- F denotes the transformation from the fixed joint coordinate frame X_{j_0} to the world coordinate system X_w. As none of these systems moves, F is constant.
- $C(\alpha)$ is the transformation from the world coordinate frame X_w to the camera coordinate frame X_c. It depends on the camera position and therefore on the joint angle α. C is usually called the extrinsic camera calibration.

B. Extrinsic Camera Calibration Data Acquisition

The kinematic calibration process is performed on the basis of a set of extrinsic camera calibrations $C(\alpha_1)...C(\alpha_n)$ at different rotations of the joint to be calibrated. The matrices C are stored together with the corresponding joint angles α. As prerequisite for the calculation of extrinsic calibration data, the intrinsic camera parameters for each camera have to be determined. For this purpose the intrinsic camera calibration procedure proposed by Zhang [20] is used.

C. Kinematic Calibration

Once extrinsic camera calibration data has been aquired at different angle positions of the joint to be calibrated, the goal of the approach consists in determining the kinematic calibration matrix B. In our approach, the matrix B is calculated using a non-linear least squares minimization technique. More precisely the Levenberg-Marquardt algorithm [19] is deployed to determine B from the set of extrinsic calibrations C and the corresponding joint angles α. The Levenberg-Marquardt algorithm minimizes a target function which is derived in

section IV-C.1. The representation and parameterization of the calibration matrix are discussed in section IV-C.2.

1) The Target Function: The target function which is minimized in order to determine the kinematic calibration matrix B is formulated using homogeneous matrices for all necessary coordinate transformations. A geometric interpretation of these transformations is given in figure 2.

The coordinate transformation from the fixed joint frame to the rotated joint frame is

$$X_j = H_j X_{j_0}. \tag{1}$$

A transformation from the rotated joint frame to the camera coordinate frame can be written as

$$X_c = B X_j. \tag{2}$$

The transformation from the world coordinate system to the camera coordinate system depends on the position of the camera and therefore on the angle α of the joint the camera is attached to. This can be formulated as

$$X_c = C(\alpha) X_w. \tag{3}$$

The transformation from the fixed joint coordinate system to the world reference system is

$$X_w = F X_j. \tag{4}$$

By combining equations (1), (2), (3) and (4), F can also be expressed as

$$F = C(\alpha)^{-1} B H_j(\alpha). \tag{5}$$

As the world coordinate frame X_w and the joint coordinate frame at zero position X_{j_0} never change, different transformations F_i and F_k with

$$F_i = C(\alpha_i)^{-1} B H_j(\alpha_i) \tag{6}$$

and

$$F_k = C(\alpha_k)^{-1} B H_j(\alpha_k) \tag{7}$$

can be calculated for different joint angles α_i and α_k, but the condition

$$F_i = F_k \tag{8}$$

always holds.

In practice however, the extrinsic camera calibrations and the joint encoder readings are not entirely accurate. Due to these and other errors it is impossible to find a B which satisfies equation (8). Instead it is the goal to find a B which minimizes the error

$$||F_i - F_k||_f = ||C(\alpha_i)^{-1} B H_j(\alpha_i) - C(\alpha_k)^{-1} B H_j(\alpha_k)||_f. \tag{9}$$

In this context $||.||_f$ denotes the Frobenius norm [21]. Let N be the number of extrinsic camera calibrations determined using the process described in section IV-B. Furthermore, let $\vec{x}$ be a parameterization of B and $G_r(\vec{x})$ and $G_t(\vec{x})$ be the minimization functionals which express the rotational and translational difference of two transformations F_k and F_{k+1} belonging to two external camera calibrations at two adjacent joint positions α_k and α_{k+1} , i.e,

$$\begin{aligned} G_r(\vec{x})_k &= angle(F_{k+1}, F_k) \\ G_t(\vec{x})_k &= translation(F_{k+1}, F_k) \end{aligned}$$

Each pair $G_r(\vec{x})$,$G_t(\vec{x})$ describes the rotational and translational difference between the homogeneous matrix F_k and F_{k+1}. To find a solution for B means to solve the minimization problem

$$\min_{\vec{x}} \sum_{k=1}^{N-1} ||w_r G_r(\vec{x})_k + w_t G_t(\vec{x})_k||, \tag{10}$$

where w_r and w_t are weighting factors for the rotational and translational parts of the error.

2) Representation and Parameterization of the Calibration Matrix: The goal of the kinematic calibration procedure is to assign a coordinate system to the rotation axis of the actuated joint. In this context two decisions have to be made:

- The representation of the calibration matrix B
- The parameterization of the calibration matrix B

The representation of B describes the mathematical means used to model the rotational part of the coordinate transformation B whereas the parameterization of B deals with the question which components of B actually have to be estimated in order to find a meaningful solution to the kinematic calibration problem. Both the choice of a representation and a parameterization of B is necessary to compute the minimization functionals $G_r(\vec{x})$ and $G_t(\vec{x})$ introduced in section IV-C.1.

For this work a three-angle representation was used for the rotational part of B. The three elementary rotations were concatenated using the Roll-Pitch-Yaw convention.

According to [10], [11] and [18], two independent axes of rotation are necessary to determine all six parameters of B. When using only one rotation around a single rotation axis, the problem is under-determined. When doing rotations around one axis and calculating the extrinsic camera calibrations at different angle positions, these extrinsic calibrations contain sufficient information to identify the rotation axis. However, there is not enough information to determine all components of the position and orientation of the joint coordinate frame on this axis. The origin of the coordinate frame on the axis is not uniquely determined. Furthermore the orientation of the coordinate frame is only restricted in a way that one coordinate axis points in the direction of the joint axis, while the other two coordinate axes can be chosen in a way that the resulting coordinate frame is a right-handed system.

In order to determine the stereo calibration the exact positions and orientations of the coordinate frames on the joint axes are not necessary. For this purpose the partial solution explained above is sufficient. It is even possible to calibrate all joints of the head-eye system using this type of partial solution. If the complete head-eye system is to be calibrated it is necessary to registrate the last coordinate system in the head with the world coordinate system. However, the registration

with the world coordinate system is always necessary, no matter if partial or complete solutions were determined for the individual joints' kinematic calibrations. Regarding the considerations above, arbitrary values can be used for the two undetermined components of the transformation, always resulting in a valid calibration matrix B.

As stated above, the parameterization of B also depends on which coordinate axis is assigned to the joint's axis of rotation. If its axis of rotation is the y axis, as shown in figure 3, the parameterization is $\vec{x} = (\alpha, \gamma, t_x, t_z)$, where α and γ denote elementary rotations around the x and z axes and t_x and t_z describe translations along the respective axes. The β and t_y components are not estimated and set to zero.

D. Stereo Calibration

The stereo calibration is required to enable methods of stereo vision on the Karlsruhe Humanoid Head. In order to use epipolar geometry to recover 3D positions of corresponding points from a stereo camera pair, the stereo calibration is required. For static cameras, the stereo calibration is usually calculated using the extrinsic camera calibrations of both cameras. The relative position of both camera coordinate systems H_{stereo} can be derived directly from the extrinsic calibrations. Having performed the kinematic calibration as described above H_{stereo} can be determined for arbitrary camera poses. This allows to perform stereo vision if the eye joints are actuated. First, in order to calculate H_{stereo} the transformation $H_{epl2epr}$ is calculated (see fig. 3) in the following way:

$$H_{epl2epr} = H_{epr}^{-1}(\alpha_R) \cdot B_R^{-1} \cdot C_R(\alpha_R) \cdot C_L^{-1}(\alpha_L) \cdot B_L \cdot H_{epl}(\alpha_L). \quad (11)$$

The matrices B_R and B_L represent the kinematic calibrations of both eye pan joints. $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ model the rotations of the respective joints by certain angles α_R and α_L. The external camera calibrations depend on the same angles. In theory the transformation $H_{epl2epr}$ could be determined at any position of the two joints. In practice however, modeling the joints' movements using $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ introduces errors. Therefore, the most accurate result can be obtained with both joints at their home positions at $\alpha_L = \alpha_R = 0$ where $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ become identity.

The stereo calibration H_{stereo} for arbitrary joint angles can then be calculated using the following equation:

$$H_{stereo} = B_L \cdot H_{epl}(\alpha_L) \cdot H_{epl2epr}^{-1} \cdot H_{epr}(\alpha_R)^{-1} \cdot B_R^{-1} \quad (12)$$

V. Experimental Results

Prior to the experiments, the optimal distance of the calibration pattern for the kinematic calibration was determined. Therefore different calibrations were performed with different positions of the calibration pattern. We used a pattern with 9×7 squares, side lenght 3.63 cm each, at distances of 0.50 m, 1.00 m, and 1.35 m from the eye system. For each calibration, the translational error Δt between the measured positions of the calibration pattern and the positions calculated using the calibrated kinematic model was measured. Table I shows the results of the experiments. As expected the accuracy of the calibration decreases with increasing distance of the calibration pattern. For the following experiments, we used a calibration distance of 0.80 m. With this distance the calibration pattern was still visible with eye pan actuations between -15 and 15 degrees and eye tilt actuations between -10 and 10 degrees. For the kinematic calibration we collected extrinsic data at steps of 1.5 degrees for the pan joints and 1.0 degrees for the tilt joint. Furthermore we evaluated the best weighting between rotational and translational error for the optimization. The best results could be achieved for the values $w_r = 0.5$ and $w_t = 1.0$. With these settings, a rotational error of 2 degrees corresponds to a translational error of 1 mm.

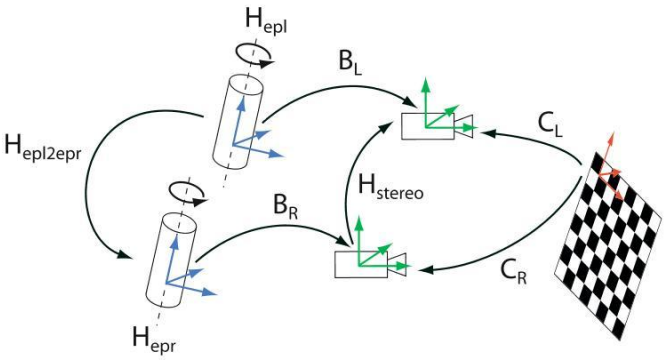

Fig. 3. The coordinate transformations necessary to determine the stereo calibration.

TABLE I
IMPACT OF THE DISTANCE TO THE CALIBRATION PATTERN ON THE MEAN MAXIMUM TRANSLATION ERROR Δt FOR DIFFERENT DISTANCES.

Calibration distance (m)	**Error** Δt (mm)
0.50	1.49
1.00	3.41
1.35	5.10

In the following we present experiments on the kinematic accuracy, the stereo accuracy and the accuracy of open-loop control.

A. Kinematic Calibration Accuracy

In a first series of experiments we investigated how accurate the kinematic model of the two pan joints is determined with the proposed method. Therefore we used a smaller calibration pattern with 5×4 squares, side length 4.5 cm. In the experiments, we performed arbitrary eye pan movments in the calibrated range of the eyes. The test pattern was positioned at distances ranging from 60 cm up to 140 cm from the eye system. For each distance 50 random test eye poses were recorded.

In order to determine the accuracy of the kinematic model, we located the 3D pose of the test pattern in the left and in the right camera using a model-based approach. Based on the calibrated kinematic model, both poses were transformed

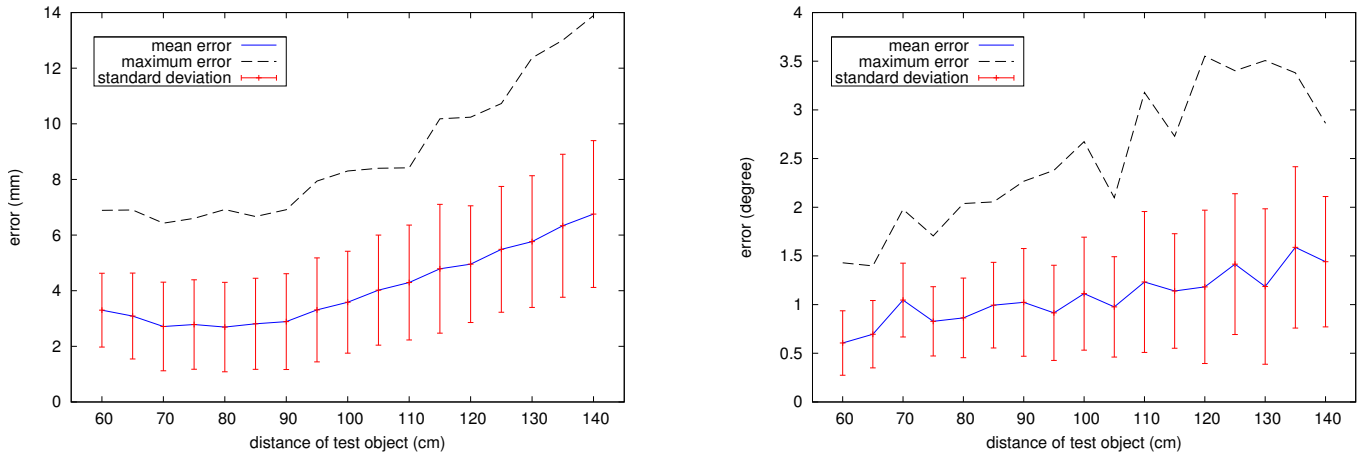

Fig. 4. Accuracy of the proposed kinematic calibration. The 3D pose of a test pattern was determined in the left and right perspective camera using a model-based approach. Both poses were transformed to a common coordinate frame using the calibrated model. The plots show the translational and rotational error for different distances of the test pattern.

into a common coordinate system and the translational and rotational errors were measured. Fig. 4 shows the results of this experiment. The plots illustrate the mean error, the standard deviation of the error and the maximum error for each distance and for the rotational and translational parts of the error. As can be seen, the major trend is a decreasing accuracy of the kinematic model with increasing distance of the test pattern. The plot for the mean of the translational part has its minimum of 2.34 mm at 80 cm - the distance where the kinematic calibration was calculated. The mean error reaches its maximum in the distance of 140 cm (6.85 mm mean error). The rotational error ranges from 0.58 up to 1.62 degrees.

B. Stereo Calibration Accuracy

The accuracy of the stereo calibration was tested in a similar way. Again the position of the test pattern was determined in the left camera using a model-based approach. Additionally, three corresponding corner points of the calibration pattern in the left and right image were determined and the 3D position of these points was calculated utilizing the epipolar geometry based on the calibrated transformation matrix H_{stereo}. From these three points, the pose of the test rig in the left camera was estimated. In that way the accuracy of the stereo triangulation could be evaluated with respect to the model based approach.

The experiments were performed for the test pattern located at distances from 60 cm up to 140 cm from the eye system, in each step 50 random test poses were recorded and evaluated. Fig. 5 shows the resulting translational and rotational error between the pose using the model-based approach and the pose calculated based on stereo triangulation. As can be seen the errors in the stereo triangulation accuracy show the same trend as the kinematic error, but are much larger. The increase of the error results from the fact, that small position errors (within the kinematic calibration) result in larger errors when performing stereo triangulation. The best results could again be achieved for small distances of the test pattern. Within a test pattern distance of 70 cm the mean translational error was measured with 8.7 mm and the minimum mean rotational error was measured with 1.72 degrees for the same distance.

C. Inverse Kinematics Accuracy

In the third series of experiments we tested the kinematic calibration in saccadic eye movement tasks. Therefore we calibrated the first three DoF of the head-eye system (namely eye pan left, eye pan right and eye tilt). For these joints, conventional differential inverse kinematics based on the inverse Jacobian could be deployed, since the kinematic system is not redundant (see also [4]). In the experiments in this section we again evaluated the accuracy at different distances of the test rig. For each distance, arbitrary camera poses were generated by moving three head joints (neck pitch, neck roll and neck yaw) to random positions in the interval of -10 to 10 degrees for each joint. The pose of the test pattern in the left camera was again determined using a model-based approach. Using the calibrated kinematic model, the inverse kinematic problem was solved and a saccade was performed in order to point the optical axes of the cameras towards the origin of the test pattern. In order to evaluate the accuracy, the distance between the corresponding corner of the test rig and the principal point of the camera in the image plane were measured, after performing the movement. Since left and right camera share a common tilt joint, the error in y direction will never be zero. The inverse kinematics module outputs the mean eye tilt actuation for left and right eye to minimize the overall error. In order to compensate for this effect, we used a modified error in y direction y_m to derive a more realistic result:

$$y_m = \frac{y_l + y_r}{2} \tag{13}$$

Using the modified y_m, the position error for left and right camera was calculated.

Fig. 6 shows the results of these experiments. The error in the left camera decreases with increasing distance of the test pattern. This effect is caused by the fixed range for eye pan

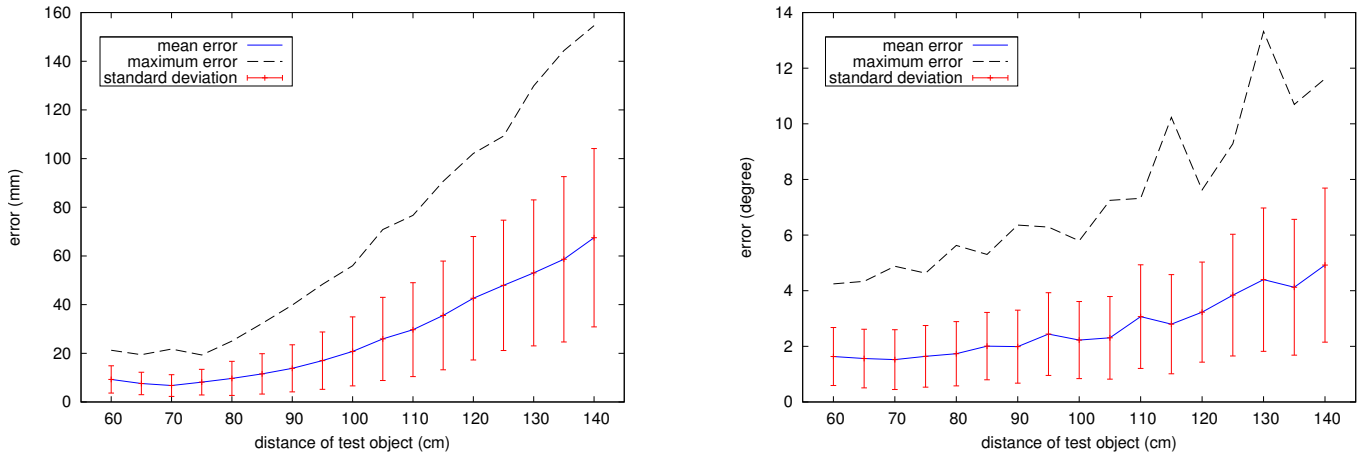

Fig. 5. Accuracy of the proposed stereo calibration. The 3D pose of a test pattern was determined in the left perspective camera using a model-based approach. Furthermore, the pose of the pattern was determined using stereo triangulation. The plots show the translational and rotational error between model based and triangulation based pose estimation for different distances of the test pattern.

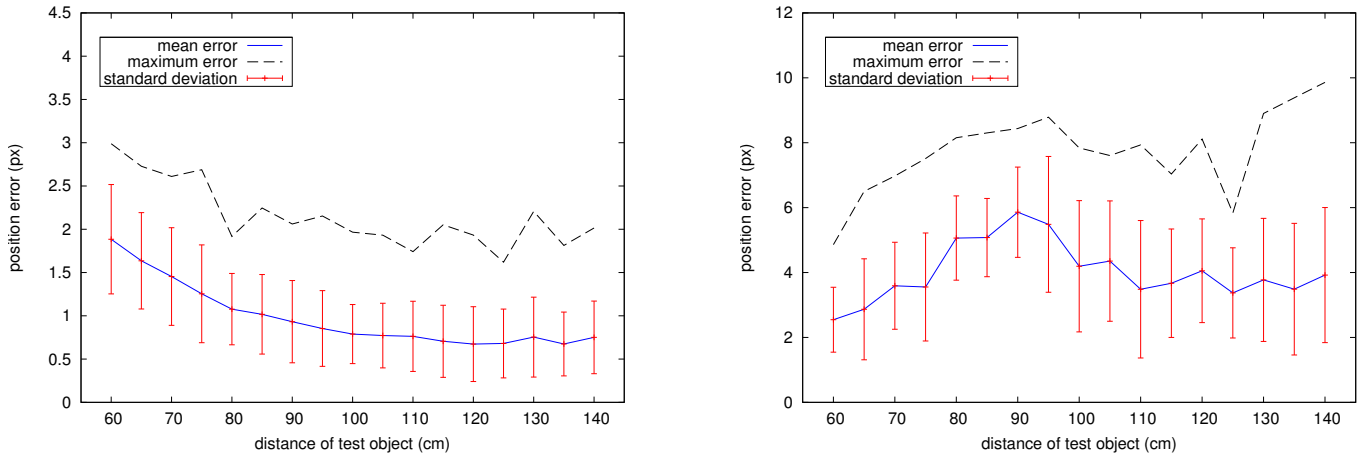

Fig. 6. Accuracy of saccadic eye movements. The 3D pose of a test pattern was determined in the left perspective camera using a model-based approach. The calibrated model was used for differential inverse kinematics in order to point the optical axes of the cameras towards the origin of the test pattern. The plots show the distance in pixels from principal point to the target corner of the test pattern in the image planes. Left: left camera. Right: right camera.

(-20 to 20 degrees) and tilt (-15 to 15 degrees) actuations. The maximum of the mean error for the left camera amounts to about 2 pixel for a distance of 60 cm. The plot for the right camera differs slightly from the results of the left camera. The increased error in the right camera is caused by the additional matrix H_{stereo} which has to be considered in the differential kinematics. In subsequent experiments, where the test pattern was located in the right camera, similar plots could be produced with more accurate results for the right camera and less accurate results for the left camera.

VI. Conclusions

In this paper we presented a new approach to solve the kinematic calibration problem for the Karlsruhe Humanoid Head's active camera system. The classical $AX = XB$ formulation of the head-eye calibration problem was combined with the benefits of a DH-based approach. The suggested method offers several advantages over common methods:

1) **Generality and accuracy:** As our method does not assume joint axes to intersect, it avoids a methodical error and allows for an improvement in calibration accuracy. Above that, it can be applied to a wider class of kinematic chains than most common methods.
2) **Robustness and accuracy:** In order to solve for the coordinate transformation from the joint to be calibrated to the camera coordinate frame, the desired calibration matrix is expressed as a non-linear least squares target function. Compared to solutions based on linear least squares, the non-linear approach presented here is less sensitive to noisy input data.
3) **Error propagation:** In contrast to many two-step approaches in the literature, the suggested method estimates all unknowns for one joint simultaneously.
4) **Verifiability:** Only one joint is calibrated at a time. This way the accuracy of an individual joint's kinematic calibration can be easily examined.

We presented experiments on the kinematic calibration accuracy, the stereo triangulation accuracy and the accuracy of inverse kinematics for saccadic eye movements. The experiments on stereo triangulation accuracy showed that at manipulation distance the pose of the test pattern could be determined with an error less then 1.5 cm. The experiments on the accuracy of the inverse kinematics showed that even at larger distances, the saccadic eye movement could be performed with a position error of less then 6 pixels in the image plane.

VII. ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Cognitive Systems project PACO-PLUS (FP6-2004-IST-4-027657) funded by the European Commission.

REFERENCES

[1] S. Sakagami, T. Watanabe, C. Aoyama, S. Matsunage, N. Higaki, and K. Fujimura, "The Intelligent ASIMO: System Overview and Integration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002, p. 24782483.

[2] K. Akachi, K. Kaneko, N. Kanehira, S. Ota, G. Miyamori, M. Hirata, S. Kajita, and F. Kanehiro, "Development of Humanoid Robot HRP-3," in *IEEE/RAS International Conference on Humanoid Robots*, 2005.

[3] "Fujitsu, Humanoid Robot HOAP-2," 2003. [Online]. Available: http://www.automation.fujitsu.com

[4] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, "The Karlsruhe Humanoid Head," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids2008)*, 2008.

[5] G. Cheng, S.-H. Hyon, A. Ude, J. Morimoto, J. G. Hale, J. Hart, J. Nakanishi, D. C. Bentivegna, J. K. Hodgins, C. G. Atkeson, M. Mistry, S. Schaal, and M. Kawato, "Cb: Exploring neuroscience with a humanoid research platform," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2008, pp. 1772–1773.

[6] A. Ude and E. Oztop, "Active 3-d vision on a humanoid robot," in *International Conference on Advanced Robotics*, 2007.

[7] A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye," *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 3457–3462, 15-19 2006.

[8] L. Itti and C. Koch, "Computational modelling of visual attention." *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, March 2001.

[9] J. Tsotsos, "Visual attention: from covert search to foveating saccades," in *Proc. Stockholm Workshop on Computational Vision*, 1993.

[10] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345–358, June 1989.

[11] Y. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 1, pp. 16–29, February 1989.

[12] M. Li, D. Betsis, and J.-M. Lavest, "Camera calibration of the KTH head-eye system," Computational Vision and Active Perception Lab., Dept. of Numerical Anaysis and Computing Science, Royal Institute of Technology (KTH), Tech. Rep., 1994.

[13] M. Li, "Kinematic calibration of an active head-eye system," *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 1, pp. 153–158, Feb. 1998.

[14] J. Neubert and N. Ferrier, "Robust active stereo calibration," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 3, 11-15 May 2002, pp. 2525–2531.

[15] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 4, pp. 617–622, Aug. 1998.

[16] G.-S. Young, T.-H. Hong, M. Herman, and J. Yang, "Kinematic calibration of an active camera system," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, 15-18 June 1992, pp. 748–751.

[17] R. Horaud and F. Dornaika, "Hand-eye calibration," *Int. Journal of Robotics Research*, vol. 14, No. 3, pp. 195–210, June 1995.

[18] F. Park and B. Martin, "Robot sensor calibration solving AX=XB on the euclidean group," *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 5, pp. 717–721, October 1994.

[19] D. Marquardt, "Algorithm for least-square estimation of non-linear parameters," *Journal of the SIAM*, vol. 11, pp. 431–441, 1963.

[20] Z. Zhang, "A flexible new technique for camera calibration," Microsoft Research, Tech. Rep., 1998.

[21] G. H. Golub and C. F. van Loan, *Matrix Computations*. North Oxford Academic, 1983.