Visual Self-Calibration of Pan-Tilt Kinematic Structures

Bartosz Tworek, Alexandre Bernardino and José Santos-Victor

Abstract— **With the increasing miniaturization of robotic devices, some actuators are not provided with absolute position sensing, thus making the state of the system unknown at startup. In this paper we present a vision based method for the automatic calibration of serial pan-tilt kinematic structures with a perspective camera on the end-effector. Examples of such systems are surveillance cameras and humanoid robot heads. The method is based on prospective motions of one joint in the kinematic chain to induce image motion in the camera. The analysis of the induced homography allows the computation of the angle of the other joint. The method can be iterated on more axes to calibrate longer serial chains composed of rotational joints. The method requires calibrated cameras, but runs completely automatic. We have implemented and validated the method in a small humanoid robot head.**

I. INTRODUCTION

Whenever the motors in robotic system are not provided with absolute sensors, it is not possible to know the state of the robot at startup. This happens frequently because most motors are equipped with incremental encoders as the main feedback sensor, lacking absolute position sensing. Therefore, a procedure is necessary to set the robot to a known initial state, also dented its home, or zero position.

To address this problem, it is usual to equip the robot with limit switches, or homing switches, that detect when the axes are in particular angular positions. However, due to miniaturization constrains, it may not be possible to install such sensors in the robot. Another possibility is to drive the axes to a mechanical stop and monitor the motor current. When the current exceeds a certain value, then the motor has reach the mechanical limit, whose angle can be known *a priori*. However, this procedure adds a source of physical stress in the system and may damage the mechanical components in the long term.

Even when the above strategies are feasible, they require the careful placement of limit and home switches, and a precise measurement of the mechanical limits. Additionally, when attaching the cameras to the end-effector, there are always some misalignments that may degrade the initial calibration procedure.

In this paper we propose a solution to this problem. We present a self-calibration procedure, at start time, that does not require absolute sensors or the need to drive the system to mechanical hard stops. Instead, it performs small prospective motions in the robot joints and observes the image motion induced in the camera. If the scene is static and scene objects are sufficiently distant from the camera, the induced image motion only depends on the prospective rotation and the angle between the camera's optical axis and the rotation axis. By iterating this procedure in the several robot axes, it is therefore possible to automatically determine the wake-up state of the system.

We address explicitly the case of pan-tilt kinematic structures because they are very common both on surveillance cameras and in robot heads. We implement the method and present results in a small humanoid robot head, calibrating its eyes and neck. Notwithstanding, the method can be easily extended for other serial kinematics structures.

This paper is organized as follows. In Section II we formulate the problem in terms of the system kinematics and an homography estimation problem. Then, in Section III we address the problem of estimating the particular homography arising in this problem. Section IV is devoted to the presentation of experimental results in the automatic calibration of a small robot head. Finally, Section V presents the conclusion of the work and directions for future developments.

II. PROBLEM FORMULATION

In this section we formulate our problem in terms of an homography estimation problem. An homography is a transformation that is able to explain the relationship between the points observed in an image before and after a rotation of the camera. From the homography it is often possible to recover the rotation angles. Therefore, we are going to analyze the homography arising from the prospective motions applied to the robot, as a function of the initial, unknown, joint angles.

Let us consider the tilt-pan kinematics structure presented in Fig. 1. A camera is attached first to a pan unit, and then the pan unit is attached to a tilt unit. A similar analysis can be made for other kinematics structures.

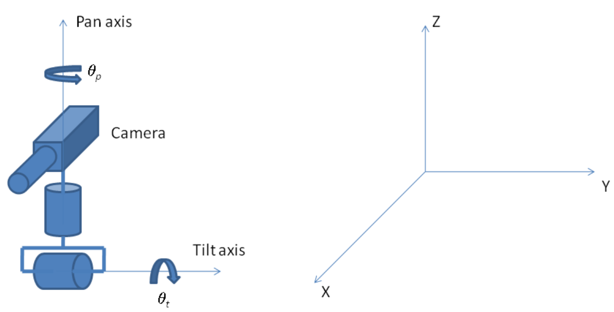

Fig. 1. Left: the kinematics structure of the pan-tilt system considered in this paper. Right: the adopted orientation for all the involved coordinate frames.

This work was supported by the European Commission, Project IST-004370 RobotCub, and by the Portuguese Government - Fundação para a Ciência e Tecnologia (ISR/IST plurianual funding) through the POS_Conhecimento Program that includes FEDER funds, and through project BIO-LOOK, PTDC / EEA-ACR / 71032 / 2006.

B. Tworek is with Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, AGH University of Technology and Science, Krakow, Poland bartosz.tworek@gmail.com

A. Bernardino and J. Santos-Victor are with the Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal {alex,jasv}@isr.ist.utl.pt

Considering identical reference coordinate frames for all joints in the canonical state, as shown in Fig. 1, the rotation matrix representing the camera's orientation with respect to the world reference frame depends on the pan and tilt angular displacements, and is given by:

$$
\begin{gathered}
R_c(\theta_t, \theta_p) = Rot_y(\theta_t) \cdot Rot_z(\theta_p) = \\
\begin{bmatrix} c_t & 0 & -s_t \\ 0 & 1 & 0 \\ s_t & 0 & c_t \end{bmatrix} \cdot \begin{bmatrix} c_p & -s_p & 0 \\ s_p & c_p & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} c_t c_p & -c_t s_p & -st \\ s_p & c_p & 0 \\ s_t c_p & -s_t s_p & c_t \end{bmatrix}
\end{gathered}
$$

with $c_t = \cos(\theta_t)$, $s_t = \sin(\theta_t)$, $c_p = \cos(\theta_p)$ and $s_p = \sin(\theta_p)$.

Fixed points in the world, at coordinates (X, Y, Z) can be expressed in the camera frame by:

$$
\begin{aligned}
\begin{bmatrix} X_c(\theta_t, \theta_p) \\ Y_c(\theta_t, \theta_p) \\ Z_c(\theta_t, \theta_p) \end{bmatrix} &= R_c(\theta_t, \theta_p)^T \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \\
&= \begin{bmatrix} c_t c_p X + s_p Y + s_t c_p Z \\ -c_t s_p X + c_p Y - s_t s_p Z \\ -s_t X + c_t Z \end{bmatrix}
\end{aligned}
$$

The perspective projection of such points in the image have the following normalized coordinates:

$$
x(\theta_t, \theta_p) = \frac{Y_c}{X_c} = \frac{-c_t s_p X + c_p Y - s_t s_p Z}{c_t c_p X + s_p Y + s_t c_p Z} \tag{1}
$$

$$
y(\theta_t, \theta_p) = \frac{Z_c}{X_c} = \frac{-s_t X + c_t Z}{c_t c_p X + s_p Y + s_t c_p Z} \tag{2}
$$

Let us consider that, at start-up, the system has initial angles θ_t^0 and θ_p^0. These angles are unknown when the system is turned on. Then, a prospective motion of the tilt unit is performed: the tilt angle is changed by θ_t. This process is illustrated in Fig. 2. For the sake of simplicity, and without loss of generality, we can consider a null initial tilt angle, $\theta_t^0 = 0$. This corresponds to set the world reference frame aligned with the initial robot's tilt frame.

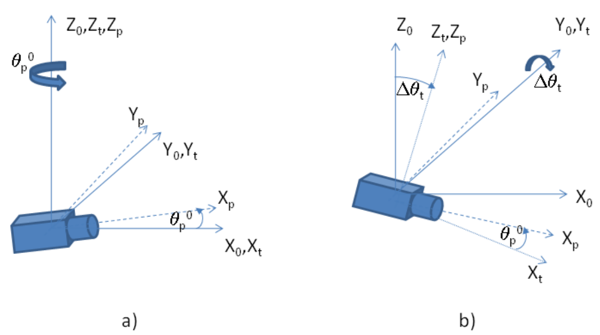

Fig. 2. The geometry of the system before (a) and after (b) the prospective motion.

In the above conditions, points observed by the camera at system start-up are at coordinates:

$$
x_0 = \frac{-s_{p0} X + c_{p0} Y}{c_{p0} X + s_{p0} Y} \tag{3}
$$

$$
y_0 = \frac{Z}{c_{p0} X + s_{p0} Y} \tag{4}
$$

where $s_{p0} = \sin(\theta_{p0})$ and $c_{p0} = \cos(\theta_{p0})$.

After the tilt's prospective motion, these points move the the new coordinates:

$$
x_1 = \frac{-c_t s_{p0} X + c_{p0} Y - s_t s_{p0} Z}{c_t c_{p0} X + s_{p0} Y + s_t c_{p0} Z} \tag{5}
$$

$$
y_1 = \frac{-s_t X + c_t Z}{c_t c_{p0} X + s_{p0} Y + s_t c_{p0} Z} \tag{6}
$$

Let us recall that we are willing to estimate θ_{p0}, the unknown pan angle at start-up. θ_t is a known, actuated angle. The image coordinates x_0, y_0, x_1 and y_1 can be measured from the images by suitable image feature detectors and trackers, and X, Y and Z are unknown 3D coordinates of world points. Our objective now is to eliminate these coordinates in the previous equations.

From equations (3) and (4) we can obtain the following constraints:

$$
\frac{Y}{X} = \frac{x_0 c_{p0} + s_{p0}}{c_{p0} - s_{p0} x_0} \tag{7}
$$

$$
\frac{Z}{X} = \frac{y_0}{c_{p0} - s_{p0} x_0} \tag{8}
$$

Now, dividing both the numerator and denominator of Eqs. (5) and (6) by X, and introducing the constraints in Eqs. (7) and (8), we obtain:

$$
x_1 = \frac{-c_t s_{p0} c_{p0} + c_t s_{p0}^2 x_0 + c_{p0}^2 x_0 + c_{p0} s_{p0} - s_T s_{p0} y_0}{c_t c_{p0}^2 - c_t c_{p0} s_{p0} x_0 + s_{p0} c_{p0} x_0 + s_{p0}^2 + s_t c_{p0} y_0} \tag{9}
$$

$$
y_1 = \frac{-s_t c_{p0} + s_t s_{p0} x_0 + c_t y_0}{c_t c_{p0}^2 - c_t c_{p0} s_{p0} x_0 + s_{p0} c_{p0} x_0 + s_{p0}^2 + s_t c_{p0} y_0} \tag{10}
$$

These equations can be rewritten in the homography form:

$$
\begin{bmatrix} x_1 \\ y_1 \\ \lambda_1 \end{bmatrix} = H \cdot \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{11}
$$

$$
H = \begin{bmatrix} c_t s_{p0}^2 + c_{p0}^2 & -s_t s_{p0} & c_{p0} s_{p0}(1 - c_t) \\ s_t s_{p0} & c_t & -s_t c_{p0} \\ c_{p0} s_{p0}(1 - c_t) & s_t c_{p0} & c_t c_{p0}^2 + s_{p0}^2 \end{bmatrix} \tag{12}
$$

A close inspection to the homography matrix shows that it has some repeated entries and only 6 of them are different. It has the form:

$$
H = \begin{bmatrix} h_1 & -h_2 & h_3 \\ h_2 & h_4 & -h_5 \\ h_3 & h_5 & h_6 \end{bmatrix} \tag{13}
$$

In the following section we will describe a method to estimate the entries of this matrix from the visual data. Once the homography is estimated, we can compute the unknown angle θ_{p0} by, e.g.:

$$
\theta_{p0} = atan2(h_2, h_5) \tag{14}
$$

III. COMPUTING THE HOMOGRAPHY

In this section we describe the method employed to estimate the particular homography arising in our formulation. The section is divided in three parts. The first part describes the methods employed to obtain point matches between the images before and after the prospective motion. The second part formulates the problem of estimating the homography from point matches, measured in the images. The third part presents the whole algorithm integrated in a robust estimation architecture (RANSAC).

A. FEATURE TRACKER

For the estimation of the homography we need first to extract from the images a set of points visible on both images (before and after the prospective motion) and their pairwise correspondences:

$$\{(x_0^i, y_0^i) \mapsto (x_1^i, y_1^i), i = 1 \cdots N\} \tag{15}$$

In the above equation, the lower index represents the image (0 for the image before and 1 for the image after the rotation), and the upper index represents the index of the point in the set, from 1 to N.

To obtain an adequate set of points in the first image, we use the corner detector in OpenCV [4], a open-source computer vision C library. The corner detector selects points in the image that are easy to track, and is based in [7]. It calculates the minimum eigenvalue of the Hessian matrix for each pixel of the image and then performs non-maxima suppression so that only local maxima in 3x3 neighborhood remains. Then it rejects the corners with the minimal eigenvalue less than a quality level determined by us. Finally it checks if all the corners found are separated enough one from another according to minimal distance determined by us.

The next step is to track the selected points. To do this we use the sparse iterative version of Lucas-Kanade optical flow method in pyramids [8], also implemented in the OpenCV library. It calculates the coordinates of the feature points on the current video frame given their coordinates on the previous frame. The function finds the coordinates with sub-pixel accuracy and rejects the points that cannot be reliably tracked in the second image.

B. HOMOGRAPHY ESTIMATION

There are several off-the-shelf routines for estimating homographies from point matches [5]. However, the homography arising in our problem has a special structure and we want to exploit this structure to improve its estimation. For each point match obtained in the tracking procedure, we know from Eqs. (12) and (13) that:

$$x_1^i = \frac{h_1 x_0^i - h_2 y_0^i + h_3}{h_3 x_0^i + h_5 y_0^i + h_6} \tag{16}$$

$$y_1^i = \frac{h_2 x_0^i - h_4 y_0^i - h_5}{h_3 x_0^i + h_5 y_0^i + h_6} \tag{17}$$

Rearranging the previous equations we get the system:

$$\begin{cases} -x_0^i h_1 + y_0^i h_2 + x_0^i x_1^i h_3 - h_3 + y_0^i x_1^i h_5 + x_1^i h_6 = 0 \\ -x_0^i h_2 + x_0^i y_1^i h_3 + y_0^i h_4 + y_0^i y_1^i h_5 + h_5 + y_1^i h_6 = 0 \end{cases} \tag{18}$$

This can be put in vector form:

$$\begin{cases} a_x^i \cdot h = 0 \\ a_y^i \cdot h = 0 \end{cases} \tag{19}$$

with

$$\begin{aligned} h &= \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 \end{bmatrix}^T \\ a_x^i &= \begin{bmatrix} -x_0^i & y_0^i & x_0^i x_1^i - 1 & 0 & y_0^i x_1^i & x_1^i \end{bmatrix} \\ a_y^i &= \begin{bmatrix} 0 & -x_0^i & x_0^i y_1^i & y_0^i & y_0^i y_1^i + 1 & y_1^i \end{bmatrix} \end{aligned}$$

Given a set of N corresponding points, we can form the following linear system of equations:

$$Ah = 0 \tag{20}$$

where:

$$A = \begin{bmatrix} a_x^1 \\ a_y^1 \\ \vdots \\ a_x^N \\ a_y^N \end{bmatrix}$$

Since the homography has 6 different entries we need at least 3 points to estimate it (each point contributes with two equations). We can compute h through the Singular Value Decomposition (SVD) of A. From the SVD we take right singular vector which corresponds to the smallest singular value. Finally we can reshape the entries of h into the homography matrix H.

C. ROBUST ESTIMATION

The above homography estimation method works well when there are no erroneous correspondences between the points in both images. Unfortunately, the tracking method sometimes provides false point matches that will degrade the results of the homography estimation. In order to address this problem, we use a well known robust estimation method that is able to eliminate the false matches (outliers) from the estimation process. The RANdom SAmple Consensus (RANSAC) [6] is an algorithm for robust estimation of models in the presence of many data outliers. We apply the RANSAC algorithm to our problem as follows:

- Repeat for L times:
  1) Select randomly a set of 3 feature pairs,
  2) Compute homography H,
  3) Calculate the distance d for each putative correspondence,
  4) Compute the number K of inliers consistent with H – the number of correspondences for which $d < t$.
- Keep H whose set of inliers is the largest.

In our case the RANSAC algorithm is applied to the putative correspondence set to estimate the homography and the correspondences which are consistent with this estimate. Distance in our case is computed as the symmetric transfer error:

$$d^2 = ||x - H^{-1}x'||^2 + ||x' - Hx||^2$$

where $x \leftrightarrow x'$ is the point correspondence.

The number of iterations L is set adaptively with following algorithm:

- $L = \infty, sample\_count = 0$
- While $L > sample\_count$ repeat:
  – Choose randomly samples and count the number of inliers
  – Set $e = 1 - \frac{(numberofinliers)}{(totalnumberofpoints})$
  – Set $L = \frac{log(1-p)}{log(1-(1-e)^s)}$ where $p = 0.99$ and $s = 4$
  – Increment $sample_count$ by 1.
- Terminate

This method is presented in [5].

IV. EXPERIMENTAL RESULTS

The proposed method was written in C++ and implemented as a module in the YARP framework [2]. YARP (Yet Another Robot Platform) is a middleware that facilitates the distributed processing and communication among different computers and provides operating system independence. We have also used the following additional libraries for the image based measurements and homography estimation:

- OpenCV (Open Source Computer Vision) which is used to create image processing part of the project [4].
- GSL (Gnu Scientific Library) which is used to solve complex mathematical problems [3].

All used libraries are free open source software.

The method was then tested with the iCub humanoid robot's head [1]. It contains 6 DOFs: neck pan, tilt and swing and eye pan and tilt as shown in Fig. 3.

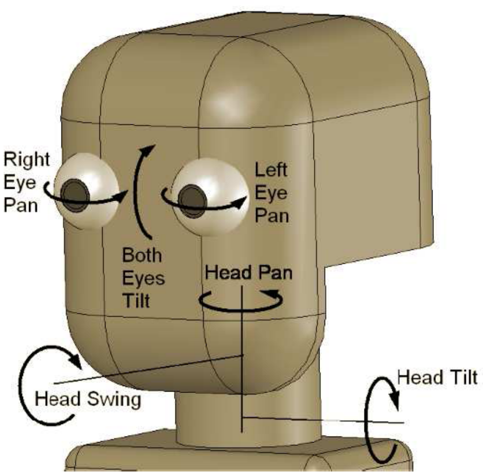

Fig. 3. Kinematics of the iCub robot head used in the tests

We have performed the calibration of the right eye and head pan, which follow directly the formulation described in this paper. Whereas for calibrating the left eye the procedure is identical, for calibrating the other joints, a different set of equation is required, but can be derived in a straightforward manner using the same principles.

The testing procedure was the following:

1) Find points of interest to track,
2) Move head or eye tilt,
3) Track points,
4) Compute homography matrix for correspondences,
5) Compute the pan angle.

Fig. 4 shows the sequence of actions required in the calibration procedure. Notice that the right eye, at startup, was offset from its canonical position. Then we have applied a tilt motion 5 degrees up to the eyes. After the above steps, if we have a sufficient number of point matches (3 is the minimum) we are able to compute the homography and the initial pan angle. Fig. 5 shows the images and the tracked points used to compute the homography. Then we move the pan axis with the symmetrical of the computed value in order to set the eye to its canonical position (zero pan).

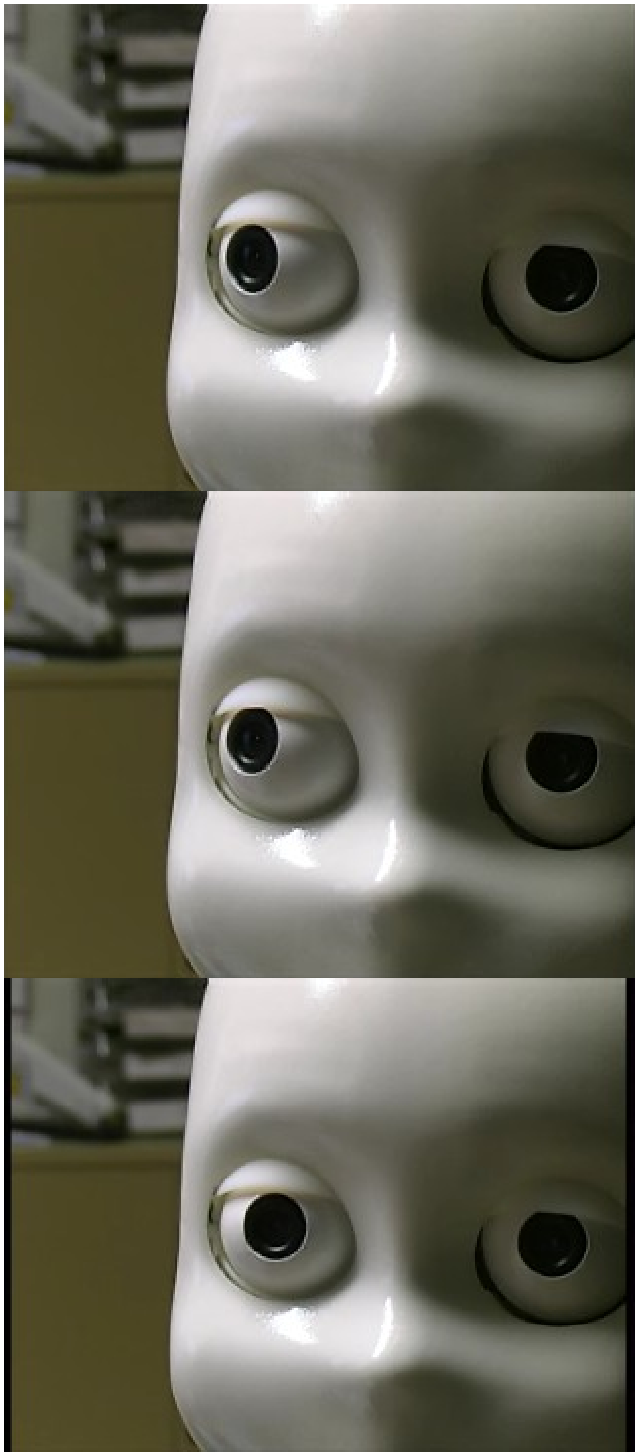

Fig. 4. Sequence of motions required to calibrate the eye. Top: right eye position at startup. Middle: right eye position after prospective tilt motion. Bottom: right eye position after calibration

The same procedure can be applied to the left eye, so that both eyes are calibrated. Once the eyes are calibrated, then we can calibrate the neck pan angle using the same method. This is illustrated in Fig. IV. Notice that in the beginning the neck pan is largely offset to its canonical position. Then we have applied a tilt motion of 10 degrees up. Then the neck pan angle was calculated and the head set to its canonical position. In the end, the eyes are still a bit tilted, because the tilt joint was not calibrated in this test.

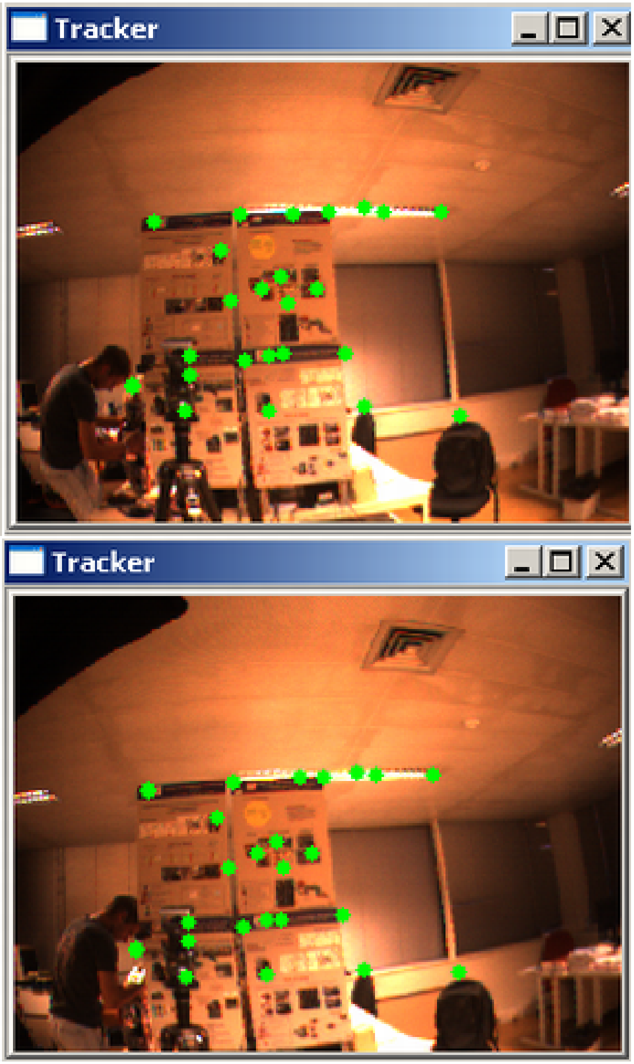

Fig. 5. Right eye image grabbed with tracked points, before (top) and after (bottom) tilt move

For a preliminary quantitative evaluation of the method's performance, we have initialized the system in a known position, and measured the pan angle using the proposed method, using different amplitudes in the tilt motions. Then, the pan was set to its assumed zero position, and the process iterated in a series of steps. The results are represented in Fig. IV, and it can be observed that the axis reaches a close vicinity of the correct position immediately after the first iteration. Then, in the remaining steps, the estimated positions are always kept within a 5 degrees range. In any case we had always good systematic results, even for small angular prospective motions.

The algorithm was tested many times for different circumstances. We came to the conclusion that results depends mostly on the quality of tracking. We had the best results while putting in front of robot a chessboard pattern, where point matches are very reliable. In other cases, there were a lot of points along edges in the image. This may lead to tracking drifts, and results may degrade a bit.

V. CONCLUSIONS

We have presented the principles for a vision based automatic calibration procedure for determining the initial unknown angles of pan-tilt kinematic structures. The method

Fig. 6. Calibration procedure for the neck pan. The robot start up with significant offset in the neck pan joint, with respect to its canonical position (top). A motion of 10 degrees is applied to the neck tilt (middle). The proposed method estimates the initial pan angle and compensates it (bottom).

is based on the computation of the homography induced by the rotation of the tilt axis. A set of points is tracked in the images before and after the prospective motions. A robust estimation architecture allows the estimation of the homographies from the tracked points, even in the presence of tracker failures (outliers). By relating the homography entries with the unknown initial angles, it is possible to estimate them reliably from the visual measurements.

In future work we will further characterize the precision

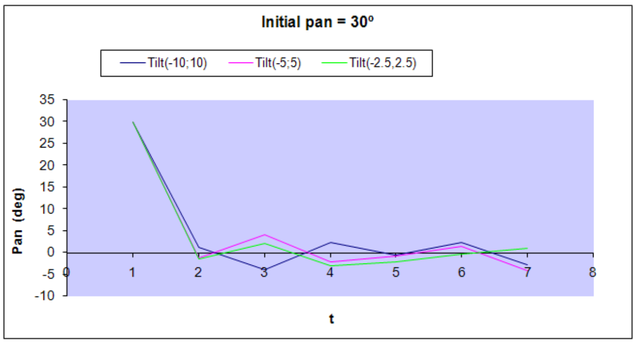

Fig. 7. Evolution of the pan angle for an iterative application of the calibration method. Different colors represent different prospective motion amplitudes.

of the method as a function of the initial conditions and motion amplitudes. We also aim at investigating how errors propagate if the method is to be employed in longer kinematics chains.

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge all members of Vislab that, along the years, have made this work possible due to the constant developments on the iCub head hardware, electronics and software.

REFERENCES

[1] Beira, R., Lopes, M., Praça, M., Santos-Victor, J., Bernardino, A., Metta, G., Becchi, F., Saltarén, R. Design of the Robot-Cub (iCub) Head, *Proc. IEEE International Conference on Robotics and Automation, ICRA*, Orlando, May, 2006.

[2] Metta, G., Fitzpatrick, P., Natale., L. YARP: yet another robot platform *International Journal on Advanced Robotics Systems, Special Issue on Software Development and Integration in Robotics*, March, 2006.

[3] Galassi, M. et al, GNU Scientific Library Reference Manual - Revised Second Edition, ISBN 0954161734

[4] Intel Research. Open Computer Vision Library. http://www.intel.com/research/mrl/research/opencv/

[5] Hartley, R. and Zisserman, A. Multiple View Geometry in Computer Vision. Cambridge University Press 2000.

[6] Fischler, M. A. and Bolles, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Vol 24, pp 381-395, 1981.

[7] Shi, J. and Tomasi, C. Good Features to Track. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593-600, 1994.

[8] Bouguet, J. Pyramidal Implementation of the Lucas Kanade Feature Tracker, OpenCV Documents, Microprocessor Research Labs, Intel Corporation, (1999).