IEEE ICDL-EPIROB 2014, pp.454-460, DOI: 10.1109/DEVLRN.2014.6983023

Markerless Online Stereo Calibration for a Humanoid Robot

Nuno Moutinho*, Ricardo Ferreira*, José Gaspar*, Alexandre Bernardino*, José Santos-Victor*
*Institute for Systems and Robotics
Instituto Superior Técnico, Lisbon, Portugal
Email: {nmoutinho, ricardo, jag, alex, jasv}@isr.ist.utl.pt

***Abstract*—In this paper we present an online stereo calibration system that calibrates the eyes of a humanoid robot at a kinematic chain level using information from the embedded cameras and the motor encoders. The system calibrates the joints' offsets and mounting errors of the cameras using advanced filtering techniques based in the epipolar constraint. By calibrating the stereo system at a kinematic chain level we are able to continuously provide an accurate stereo reconstruction of the world with version and vergence of the eyes. Our method is able to converge even with large initial errors and is independent of camera movements - can operate with still or moving cameras. The filtering approach is essential for smoothing out errors arising from spurious image matches and converges in just a few seconds in case of sudden model changes.**

Experiments using the iCub robotic head are presented and illustrate the performance of the methodology as well as the advantages of using such an approach.

I. Introduction

Humanoid robots are in vogue, with different sizes, shapes and features. This variety makes each one unique, requiring specific procedures in order to correctly perform the tasks they were designed for. However, the biological inspiration used in their designing process somehow standardized the basic type of sensors they should all include: two cameras for stereo vision, inertial measurement units and encoders in their motor joints. We humans are able to correctly integrate and fuse information from similar sensors due to a perfectly calibrated internal model which is continuously adapted through life-long developmental learning mechanisms. The ability to obtain and keep an accurate model of a humanoid robot's internal state, adapting both to sudden changes during operation or repair, and slow changes due to material changing properties, is an essential but challenging task.

In [7] we presented an online calibration system to learn the internal model of a humanoid robot head by fusing information from these three types of sensors. Even though the system was accurate, the kinematic model used to represent the robot eyes was incomplete since it did not contemplate possible mounting errors of the cameras. Figures 1(d) and 1(e) show how these mounting errors influence the quality of the stereo reconstruction and the difference between both cases is quite noticeable. In this work we decided to change our focus to stereo calibration thus requiring a new system suitable to work in dynamic environments, unlike our previous work where the world was assumed to be static.

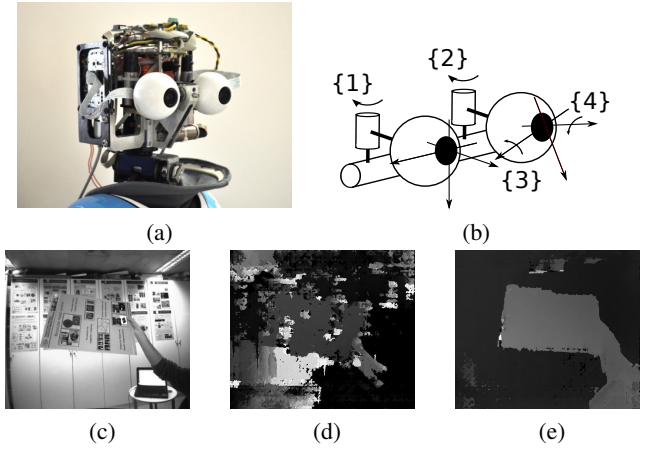


Fig. 1. The iCub head used in this work (a). The complete kinematic model of the robot eyes, including the mounting errors of the left camera (b). Original scene to be reconstructed (c). Stereo reconstruction using the kinematic model of the robot eyes considering perfect assembly (only joints $\{1\}$ and $\{2\}$) (d). Stereo reconstruction after calibration of the complete kinematic model of the robot eyes, considering the mounting errors of the left camera ($\{3\}$ and $\{4\}$) (e). The stereo reconstruction code is taken off-the-shelf from OpenCv.

Stereo reconstruction is very sensitive to small misalignments between the two images: a small misalignment of just one pixel between the two images is enough to get a miss reconstruction. This sensitivity is even more serious when working with active vision where the transformation between the two images changes considerably through time. There are numerous methods capable of accurately calibrate the stereo system for a unique configuration of the cameras, such as the Bouguet toolbox [1]. However, calibration is usually performed in an offline fashion and when this configuration is changed by moving the cameras, the calibrated status is lost and a new calibration is required. In robotics, very few works tackle the continuous update of the system's parameters as it moves in the environment with active vision. In this paper we present a stereo calibration system that performs online calibration at a kinematic model level, meaning the robot can move the eyes and still reconstruct the world using information solely from the kinematic model at each time. This calibration system does not require any external visual markers thus permanently adjusting the robot calibration state

by using natural visual information from the world.

When considering state-of-the-art calibration systems, the Bouguet toolbox [1] is one of the most well known and precise system given the amount of information it uses to reach a calibration state. This toolbox is able to provide intrinsic and extrinsic parameters for both cameras but requires the acquisition of (many) images with a visible calibration grid and human intervention to control and validate the acquisition process. The calibration patterns must fully intersect in all the images otherwise stereo calibration can not be achieved. In [5] a new type of calibration pattern is presented for multi-camera calibration in which only a small part of the calibration pattern needs to be visible (without images intersection) in order to calibrate the system. In spite of their precision and accuracy, these systems are highly time consuming and can only work in an offline manner. The authors of [2] present an accurate offline calibration method for multi-cameras that uses a bundle adjustment technique to optimize some previously obtained calibration parameters without requiring any human intervention, unlike the previous ones.

Online stereo calibration systems allow for small changes during operation of the cameras since the system can adapt to that changes. In [8] a system to estimate the extrinsic parameters between two cameras without any reference objects or human intervention, by integrating visual information with angular velocities given by a gyroscope to reject image features that are useless to the estimation, is presented. The system described in [3] uses a Kalman Filter to continuously estimate the extrinsic parameters of the stereo cameras and the one explained in [6] uses recursive filtering techniques and plane induced homographies between successive frames to optimize the calibration parameters between the cameras, in an online manner. The method explained in [10] implements a new form of partitioned bundle adjustment to calibrate stereo cameras and simultaneously provide an estimate for the pose of the cameras. When working with active vision, it is important to know the exact transformation between the two cameras so stereo reconstruction can be obtained for every configuration of the cameras. The authors of [11] tried a different approach and calibrated the stereo cameras at a kinematic level so the transformation between the two cameras is always known during operation with active vision. However this system requires the use of a calibration pattern and human intervention to achieve calibration state. In comparison with the state-of-the-art, our system calibrates the cameras at a kinematic model level in an online fashion simply by observing the world, without requiring any human intervention during the calibration process. It uses natural visual features instead of external visual markers which gives the flexibility to calibrate the system anywhere. The fact it provides correct stereo transformations that work with active vision is a major advantage when compared with the previously mentioned systems.

The structure of the paper is as follows. Section II does an introduction to the calibration problem. Section III describes the proposed methodology for the stereo calibration. Results obtained with a humanoid robotic head are shown in section IV and conclusions are drawn in section V.

II. Problem Formulation

Let ${}^{R}T_{L}$ represent the transformation between the left and right images. In a perfectly calibrated active stereo system, it is possible to compute the transformation ${}^{R}T_{L}$ at a kinematic level using solely the readings from the motor encoders, which allows quality stereo reconstruction even with cameras motion. The problems arise because very often humanoid robot heads are equipped with relative encoder sensors and exhibit mounting errors in the camera's assembly. On one hand, the relative encoder sensors lack the capability of providing an absolute encoder position with respect to some factory calibrated value, e.g. having the cameras looking forward with parallel projection planes. These encoders are not persistent and fix their zero value at the position in which they are turned on, consequently demanding recalibration each time the system is powered on. On the other hand, mounting errors of the cameras are persistent within the same platform but are not usually considered in a kinematic model of a robot since they change between robots.

In order to generalize the calibration problem we include the mounting errors into the kinematic model of the robot. We assume a model where one of the eyes has only one rotation joint (pan joint, $\{1\}$), e.g. the right camera, thus representing the mounting errors as two virtual joints for the left camera only (virtual tilt and swing joints, $\{3\}$ and $\{4\}$, respectively) placed in front of its last real joint (pan joint, $\{2\}$) as seen in figure 1(b). These virtual joints had no encoder readings thus representing only the physical mounting rotations of one camera in respect to the other. The complete transformation ${}^{R}T_{L}$ from the left camera's reference frame L to the right camera's reference frame R can be represented as a composition of elementary transformations between all the consecutive joints

$$\begin{aligned} &{}^{R}T_{L}\left(\theta_{1}\ldots\theta_{4}\right) \\ &= {}^{R}T_{\{4\}}\left(\theta_{4}\right).{}^{\{4\}}T_{\{3\}}\left(\theta_{3}\right).{}^{\{3\}}T_{\{2\}}\left(\theta_{2}\right).{}^{\{2\}}T_{\{1\}}\left(\theta_{1}\right).{}^{\{1\}}T_{L} \end{aligned} \tag{1}$$

where each transformation is represented as

$${}^{j}T_{i}\left(\theta_{i}\right) = \begin{bmatrix} {}^{j}R_{i}\left(\theta_{i}\right) & {}^{j}t_{i} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{2}$$

and is a function of the rotation angle θ_i associated to the ith joint. The final transformation ${}^{R}T_{L}$ is then a function of all the rotation angles θ_i associated to each individual joint. Due to the referred encoder resetting when powered up, the joint angles θ_i are given by the encoder readings e_i subtracted by a constant δ_i which needs to be estimated. In the case of the virtual joints, the value for e_i is always zero (non existent). The objective of this work is to estimate the offsets δ_i such that the relation

$$\theta_i = e_i - \delta_i \tag{3}$$

holds. We then collected the joint offsets δ_i in a single vector χ, the system state to be estimated at each time step.

Using this concept of virtual joint, we can generalize the calibration problem to any stereo system equipped only with cameras by forcing all the joints to be pure virtual with no measurements thus having all the values of e_i equal to zero.

III. Calibration Methodology

The base of our system is an Implicit Kalman Filter [9], which, given the state transition and the sensor observation equations is able to update the state in order to reduce the measurement error, thus providing the best estimates for the joints offsets. An Implicit Kalman Filter is used since none of the measurement equations can be written in explicit form, defining instead a constraint that the measurements, together with the system state, need to satisfy.

A. State Transition Model

In our system state χ the joint offsets are assumed to be constant over time, thus the state transition equation $\mathcal{F}$ simply propagates the previous values with some state transition noise w^k.

$$\chi^{k|k-1} = \chi^{k-1} + w^k \tag{4}$$

Here $w^k \sim N\left(0, Q^k\right)$ where Q^k represents the covariance matrix of the zero mean state transition noise w^k, assumed Gaussian. The system can be adapted to be more or less responsive to variations in the offsets by changing this covariance matrix.

B. Observation Model

The cameras provide N image features represented by their image coordinates $f_i = [u_i v_i]$ which can be collected in a feature measurement vector Z_F at a certain time instant k:

$$Z_F^k = \left[f_0^k \dots f_{N-1}^k\right] \tag{5}$$

The N image features for the left and right cameras, Z_{FL}^k and Z_{FR}^k respectively, are sorted by match. The features are matched between the two images by applying the SIFT detector, extracting their descriptors and performing a K-nearest neighbor search (in our case $k = 2$).

Considering M kinematic joints, a scan of all the encoders, Z_E^k, consists of M measurements of the relative position of the joints taken at time instant k,

$$Z_E^k = \left[e_0^k \dots e_{M-1}^k\right]. \tag{6}$$

In the case of a virtual joint i, the value for e_i^k is always zero. The encoders are sampled at the same time instants as the cameras so synchronization between sensors is ensured.

The system measurements Z^k are, at each time instant k, given by:

$$Z^k = \left[Z_{FL}^k Z_{FR}^k Z_E^k\right] + v^k \tag{7}$$

where $v^k \sim N\left(0, R^k\right)$ is the observation noise assumed to be a zero mean Gaussian with covariance matrix R^k. These measurements provide geometrical constraints which are used by the filter to improve the estimates.

In order to obtain the correct and complete transformation ${}^R T_L$, from the left camera to the right camera, the absolute value of each joint (real and virtual) is needed. Collecting equation (3) in vector form, the absolute values of the joints Θ are given by

$$\Theta^k = Z_E^k - \chi^k \tag{8}$$

For the purpose of describing our filter's observation function, let's assume we known the correct transformation ${}^R T_L$ between the left and right images. We can take the rotational component of this transformation, ${}^R R_L$, as well as the translational component, ${}^R t_L$, and compute the Essential Matrix E [4] that relates corresponding points in both images:

$$E = \left[{}^R t_L\right]_\times {}^R R_L \tag{9}$$

where $\left[{}^R t_L\right]_\times$ is a skew-symmetric matrix using the vector components of ${}^R t_L$. In perfect conditions, if we take a pair of image features $\tilde{x}_L$ and $\tilde{x}_R$, represented by their homogeneous normalized coordinates, the following relation holds:

$$\tilde{x}_R^T E \tilde{x}_L = 0 \tag{10}$$

This relation, called the Epipolar Constraint [4], can itself be rewritten in a different form in order to use image coordinates instead of normalized coordinates of the features by a change of variables $\tilde{x}_R$ and $\tilde{x}_L$. Thus, we have a new constraint given by:

$$x_R^T F x_L = 0 \tag{11}$$

with

$$F = \left(K^R\right)^{-T} E \left(K^L\right)^{-1} \tag{12}$$

where x_L and x_R correspond to the coordinates of the features, in pixel units, and F is the Fundamental Matrix [4]. In non-perfect conditions, using a pair of noisy matched image points, equation (11) will not be zero but an ϵ instead. Next we will use the distance to the epipolar line, ϵ, as a measurement error. Considering the right part of the product of equation (11):

$$\begin{bmatrix} A & B & C \end{bmatrix}^T = F x_L \tag{13}$$

it gives the three components of the epipolar line, describing a line, where the point correspondence x_R should lie in the right image. To obtain the distance ϵ from x_R to its closest projection in the line, in pixel units, we have to normalize these three components

$$\begin{bmatrix} A^* & B^* & C^* \end{bmatrix}^T = \begin{bmatrix} A & B & C \end{bmatrix}^T / \sqrt{A^2 + B^2} \tag{14}$$

and apply them to the corresponding point x_R

$$x_R^T \begin{bmatrix} A^* & B^* & C^* \end{bmatrix}^T = \epsilon \tag{15}$$

This distance ϵ tells us how far, in pixel units, point x_R is from its corresponding epipolar line in the right image. We use this distance ϵ in our filter's observation function as an implicit constraint that guarantees the correct estimation of the joint offsets. To accomplish this, we first obtain the transformations ${}^{R}T_{L}\left(\Theta^{k}\right)$ and ${}^{L}T_{R}\left(\Theta^{k}\right)$, using equations (1) and (8). We rewrite these two transformations as in (12) and apply equations (13) to (15) to each pair of features i represented in Z^{k}_{FL} and Z^{k}_{FR}. From each pair of features i we obtain two distance measurements, ${}^{L}\epsilon_{i}$ and ${}^{R}\epsilon_{i}$, representing the distances from the left and right features to their corresponding epipolar lines, respectively. We then compute the squared norm of the error

$$g_i = \left({}^{L}\epsilon_{i}\right)^{2} + \left({}^{R}\epsilon_{i}\right)^{2} \tag{16}$$

into a larger vector $\mathcal{G}^{k}$ with N entries, according to the number of matched image features

$$\mathcal{G}^{k}\left(\chi^{k|k-1}, Z^{k}\right) = \begin{bmatrix} g_0 & \cdots & g_{N-1} \end{bmatrix}^{T} \tag{17}$$

representing the Implicit Kalman filter's innovation function, depending both on the system state prediction $\chi^{k|k-1}$ and on the measurements Z^{k}. By using this function as the filter's innovation, the system is able to estimate all the joint offsets that minimize the distance between each point to its corresponding epipolar line.

IV. Results

This section describes the experiments realized to test the behaviour and performance of the online stereo calibration system. The experiments were performed using the iCub robotic head at VisLab, IST-Lisboa [1], as seen in figure 1(a).

The calibration procedure consists of initializing the eyes of the robot in an arbitrary uncalibrated position, setting them to the zero position. Normally, a mechanical calibration procedure at initialization time is performed, either manually by the user or automatically by driving the joint until a limit, so that the eyes are looking forward with approximately parallel optical axes. Then we run the system until a convergence state is reached. In order to validate our system, two specific sets of experiments were realized with the iCub head testing the (IV-A) accuracy and (IV-B) repeatability of the calibration results obtained. We compared our results with those obtained using the Bouguet toolbox, as a benchmark, using a calibration grid with known dimensions (with a length L of 381mm and a height H of 296mm).

The intrinsic parameters of the cameras were obtained using the Bouguet toolbox and the calibration grid previously mentioned. The standard deviations of the observation noises were measured in order to have an accurate characterization of the sensors and define the covariance matrix R^{k} in the observation model. The measured values are shown in table I along with the frequency associated with each sensor.

[1] http://vislab.isr.ist.utl.pt

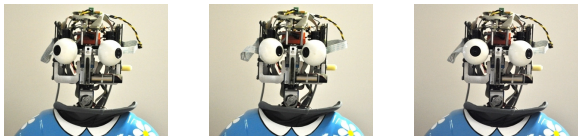
TABLE I
Characterization of the sensors

Sensor	Noise Std. Dev.	Frequency (Hz)
Cameras	5pixel	30
Encoders	0.0005rad	> 30

A. Accuracy analysis of the calibration results

We started the iCub eyes at three different positions, for experiments 1, 2 and 3, as shown in figures 2(a), 2(b) and 2(c) respectively, and set the eyes to their zero position, looking forward. We exaggerated in the initialization errors to show the ability of the system to converge even in extreme cases. Figures 3(a), 3(b) and 3(c) show the convergence of the system to the corresponding offsets of the joints in approximately 350 iterations, 150 iterations and 950 iterations, respectively, with a frame rate of 10 frames per second. During these periods the cameras were statically looking to a dynamic scenario. The filter was initialized with all offsets to zero. Due to unfavourable initializations, these values are higher than in more reasonable cases. In a normal initialization case the system converges in a few iterations (< 30).

To test the system's accuracy after convergence we took 8 images with the calibration grid with different orientations and obtained the calibrated stereo transformation using the Bouguet toolbox. The results obtained with the Bouguet toolbox are meant to work as a ground truth for our system due to its precision.



(a) (b) (c)

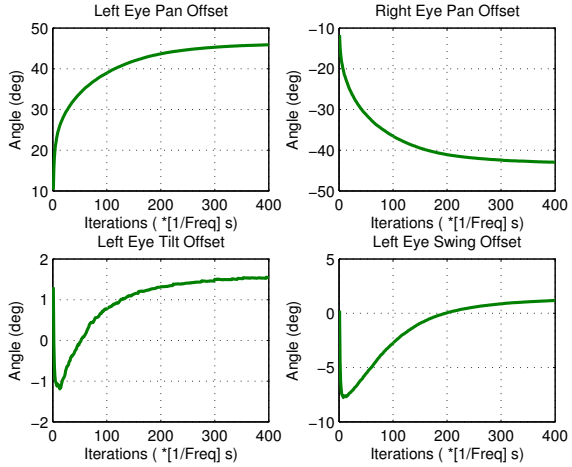
Fig. 2. Accuracy: Convergence of the joints offsets for three different starting positions of the robot eyes. Starting position of the eyes for experiment 1 (a), experiment 2 (b) and experiment 3 (c).

In tables II, III and IV are represented the rotational (rx, ry and rz) and translational (tx, ty and tz) vector parameters, extracted from the transformation ${}^{R}T_{L}$ given by the two systems for each configuration of the eyes.

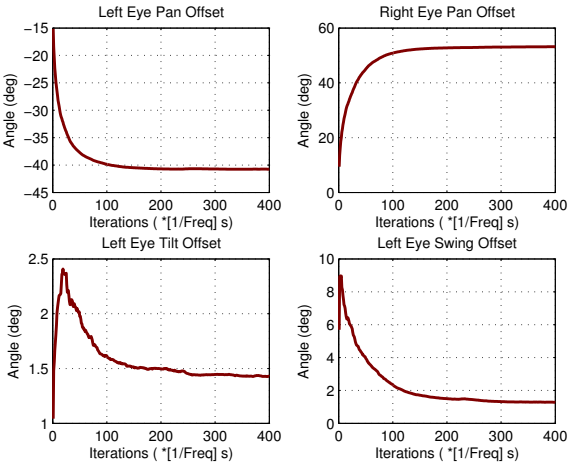
TABLE II
Rotational and translational vectors parameters, representing the transformation from left to right camera for experiment 1

Experiment 1	Our System	Bouguet	error
rx (deg)	-1.5976	-1.0657	0.5319
ry (deg)	0.9786	0.4137	0.5649
rz (deg)	-1.1375	-1.2026	0.0651
tx (mm)	-66.7387	-66.9005	0.1618
ty (mm)	0.0	1.08175	1.08175
tz (mm)	1.9805	-0.36044	2.34094

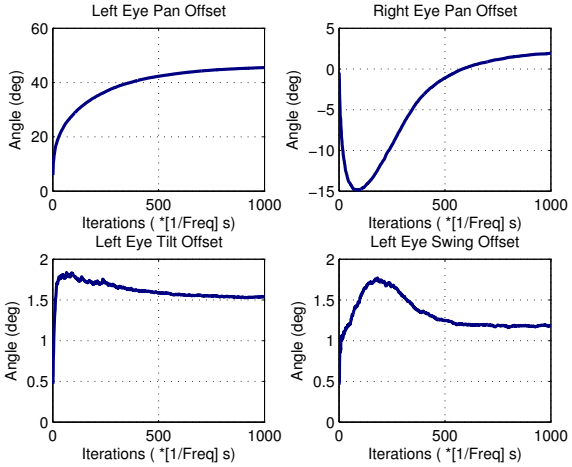
We get a mean error of 0.3322deg in the rotational parameters and a mean error of 1.1704mm in the translational



(a) Convergence of the joints offsets for experiment 1



(b) Convergence of the joints offsets for experiment 2



(c) Convergence of the joints offsets for experiment 3

Fig. 3. Accuracy: Convergence of the joints offsets for three different starting positions of the robot eyes.

TABLE III
ROTATIONAL AND TRANSLATIONAL VECTORS PARAMETERS, REPRESENTING THE TRANSFORMATION FROM LEFT TO RIGHT CAMERA FOR EXPERIMENT 2

Experiment 2	Our System	Bouguet	error
rx (deg)	-1.5818	-1.0267	0.5551
ry (deg)	1.0777	1.107	0.0293
rz (deg)	-1.3037	-1.1998	0.1039
tx (mm)	-66.6952	-66.6854	0.0098
ty (mm)	0.0	1.0803	1.0803
tz (mm)	2.6482	0.5001	2.1481

TABLE IV
ROTATIONAL AND TRANSLATIONAL VECTORS PARAMETERS, REPRESENTING THE TRANSFORMATION FROM LEFT TO RIGHT CAMERA FOR EXPERIMENT 3

Experiment 3	Our System	Bouguet	error
rx (deg)	-1.5449	-1.0434	0.5015
ry (deg)	0.8004	0.2126	0.5878
rz (deg)	-1.1776	-1.2273	0.0497
tx (mm)	-66.7849	-66.8968	0.1119
ty (mm)	0.0	1.1071	1.1071
tz (mm)	1.932	-0.5598	2.4918

parameters.

To analyse the influence of these errors in the stereo reconstruction, we extracted three corners from an image where the calibration grid was at a distance in between 380mm and 450mm, almost parallel to the cameras, and estimated the dimensions of the calibration grid (L and H). The estimated results, for each configuration of the eyes, are shown in tables V and VI.

TABLE V
ESTIMATING THE DIMENSIONS OF H ($296mm$)

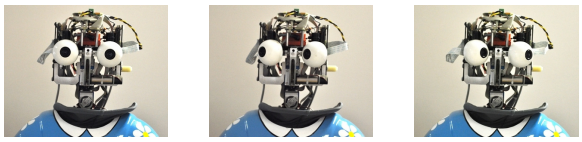
Experiment #	Our system (mm)	Bouguet (mm)
1	275.482	298.429
2	294.854	297.291
3	268.108	291.895

TABLE VI
ESTIMATING THE DIMENSIONS OF L ($381mm$)

Experiment #	Our system (mm)	Bouguet (mm)
1	354.029	377.302
2	375.38	374.379
3	348.317	369.629

Using the Bouguet toolbox as a benchmark, we obtain a mean reconstruction error of 19.138mm, in comparison with the Bouguet's mean error of $4.919mm$.

To test the accuracy of our system with motion of the eyes (version), we started the eyes encoders at their zero position as shown in figure 4(a). We then moved both eyes, by hand, to the right and left, without turning the system off, as seen in figures 4(b) and 4(c). During this operation, 8 images containing the calibration grid were acquired at each position of the eyes, for each camera, so we could compare our results with the ones obtained using the Bouguet toolbox (a total of 48 calibration images were used).



(a) (b) (c)

Fig. 4. Accuracy: Eyes positions during online calibration. (a) Position 1 - Starting position of the eyes. (b) Position 2 - Eyes looking to the right. (c) Position 3 - Eyes looking to the left.

To compare the results, we extracted the three corners of the calibration grid from calibration images, at a distance in between $430mm$ and $550mm$, and again estimated the dimensions of L and H for the three positions of the eyes. The estimated results, for each position of the eyes, are shown in tables VII and VIII.

TABLE VII
ESTIMATING THE DIMENSIONS OF H $(296mm)$

Eyes position #	Our system (mm)	Bouguet (mm)
1	309.647	298.073
2	276.267	308.612
3	274.659	296.154

TABLE VIII
ESTIMATING THE DIMENSIONS OF L $(381mm)$

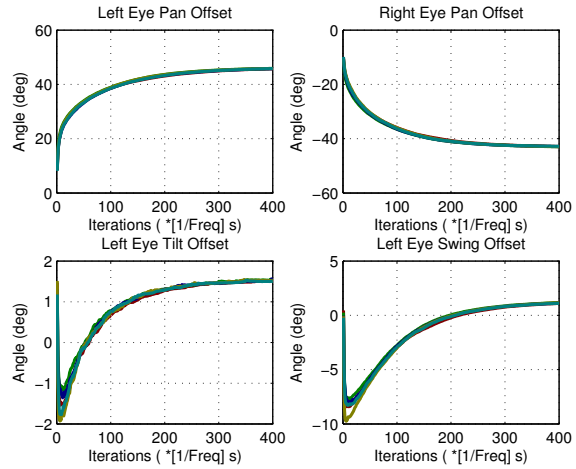
Eyes position #	Our system (mm)	Bouguet (mm)
1	388.669	370.681
2	364.112	392.749
3	359.939	387.592

In this case we obtained a mean reconstruction error of $16.723mm$ for our system, in comparison with the Bouguet's of $7.249mm$. We clearly see the advantage of our system when working with active since the quality of the reconstruction remains almost the same for different positions of the eyes.

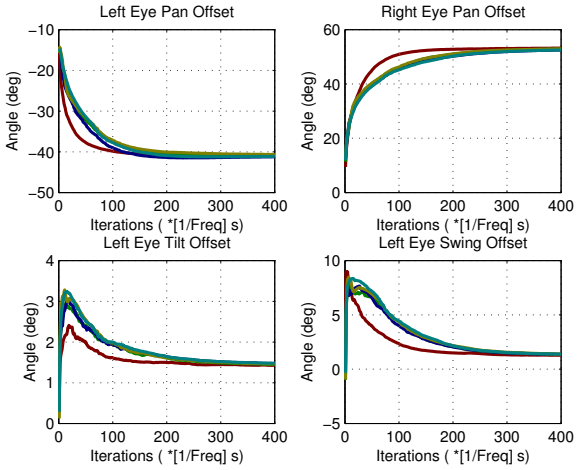
B. Repeatability analysis of the calibration results

The correct behavior of the system is also validated by ensuring its repeatability. To test the repeatability of the calibration procedure, we ran the algorithm 5 times for each configuration of the eyes, shown in figures 2(a), 2(b) and 2(c), without a full reset of the encoders between each trial. Figures 5(a), 5(b) and 5(c) show the convergence of the offsets estimates for the right pan offset (δ_1), the left pan offset (δ_2), the virtual left tilt offset (δ_3) and the virtual left swing offset (δ_4) in each experiment with the mean values taken in the last 50 iterations shown in table IX, X and XI along with the standard deviations.

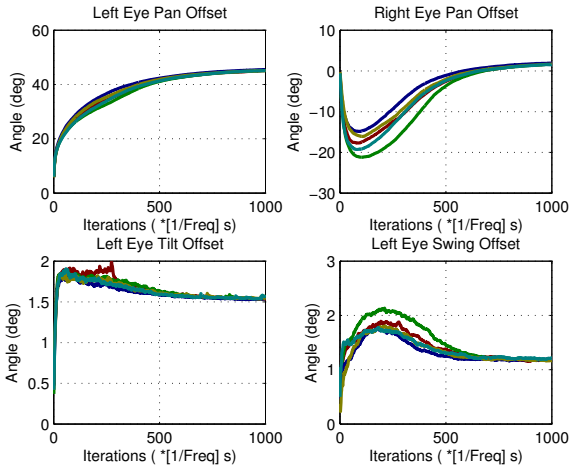
Considering the two last joints represent the mounting errors of the left camera, the estimates of the offsets should be roughly the same, independently of the starting position of the eyes. By analysing the results for the three experiments, we have a standard deviation of just $0.0503deg$ for the virtual left tilt offset (δ_3) and 0.0612 for the virtual left swing offset



(a) Convergence of the joints offsets for experiment 1 (5 trials)



(b) Convergence of the joints offsets for experiment 2 (5 trials)



(c) Convergence of the joints offsets for experiment 3 (5 trials)

Fig. 5. Repeatability: Convergence of the (real and virtual) joints offsets for three different starting positions of the robot eyes (5 trials per experiment).

TABLE IX
REPEATABILITY TEST - EXPERIMENT 1

Trial #	δ_1 (deg)	δ_2 (deg)	δ_3 (deg)	δ_4 (deg)
1	-43.0344	45.9615	1.5336	1.2083
2	-43.0198	46.0341	1.5342	1.228
3	-43.0763	46.0349	1.5419	1.2498
4	-43.0752	46.1249	1.5538	1.2676
5	-42.8647	45.9609	1.5165	1.1617
std.dev.	0.0871	0.0676	0.0136	0.0409

TABLE X
REPEATABILITY TEST - EXPERIMENT 2

Trial #	δ_1 (deg)	δ_2 (deg)	δ_3 (deg)	δ_4 (deg)
1	53.2242	-40.7295	1.4286	1.2693
2	52.9378	-40.8335	1.4434	1.3446
3	52.5676	-41.2097	1.4687	1.3508
4	53.1539	-40.6823	1.4367	1.2843
5	53.3059	-40.5371	1.4241	1.2751
std.dev.	0.2963	0.2534	0.0175	0.0396

TABLE XI
REPEATABILITY TEST - EXPERIMENT 3

Trial #	δ_1 (deg)	δ_2 (deg)	δ_3 (deg)	δ_4 (deg)
1	1.6535	45.1342	1.5367	1.1884
2	1.6026	45.1669	1.5655	1.1863
3	1.9534	45.6184	1.5365	1.1873
4	1.7522	45.2004	1.5364	1.1663
5	2.0143	45.5788	1.5376	1.1811
std.dev.	0.1817	0.2379	0.0129	0.0092

(δ_4), for a total of 15 trials. The results show that the different experiments converge to similar values in their trials thus empirically proving the robustness of the proposed method.

V. CONCLUSIONS

In this paper we have presented a new markerless stereo calibration system that calibrates the eyes of a humanoid robot at a kinematic model level in an online fashion. The system does not require any visual marker in the world thus using natural image points to achieve a calibration state. The calibration is obtained at a kinematic model level meaning it allows the use of active vision without losing quality in the stereo reconstruction due to the system's robustness.

We tested the accuracy and repeatability of our system using the results with the Bouguet toolbox as ground truth. We are able to correctly calibrate the stereo system independently of the starting position of the eyes as seen by the correct estimates of the calibration pattern dimensions with acceptable reconstruction errors, considering the sensitivity of stereo reconstruction. The same calibration state is achieved after restarting the system without resetting the encoders which demonstrates the repeatability, and therefore the robustness, of the system. The online calibration at a kinematic model level is tremendously useful when working with active vision so a correct stereo reconstruction is always possible for every position of the robot's eyes, as seen in performed tests.

We believe that, as the efforts on miniaturisation of robots grow and their complexity increase, automatic and self-sources of uncertainty, will be key to the usability and reliability of future robots.

ACKNOWLEDGEMENTS

This work was supported by the European Commission project POETICON++ (FP7-ICT-288382) and by the FCT project [PEst-OE/EEI/LA0009/2013].

REFERENCES

[1] J.Y. Bouguet. *Camera Calibration Toolbox for Matlab*. 2008.

[2] Y. Furukawa and J. Ponce. "Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment". In: *International Journal of Computer Vision* 84.3 (2009), pp. 257–268.

[3] Peter Hansen et al. "Online Continuous Stereo Extrinsic Parameter Estimation". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 17.2 (2012), pp. 234–238.

[4] Richard I. Hartley. "In Defense of the Eight-Point Algorithm". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997).

[5] Lionel Heng, Bo Li, and Marc Pollefeys. "A Multiple-Camera System Calibration Toolbox Using A Feature Descriptor-Based Calibration Pattern". In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on* (2013).

[6] Moritz Knorr, Wolfgang Niehsen, and Christoph Stiller. "Online Extrinsic Multi-Camera Calibration Using Ground Plane Induced Homographies". In: *IEEE Intelligent Vehicles Symposium* (2013).

[7] Nuno Moutinho et al. "Online calibration of a humanoid robot head from relative encoders, IMU readings and visual data". In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on* (2012).

[8] Pawe Peczyski and Bartosz Ostrowski. "Automatic Calibration of Stereoscopic Cameras in an Electronic Travel Aid for the Blind". In: *Metrology and Measurement Systems* 20 (2013), pp. 229–238.

[9] S. Soatto. "A geometric framework for dynamic vision". PhD thesis. California Institute of Technology, 1996.

[10] Michael Warren, David McKinnon, and Ben Upcroft. "Online Calibration of Stereo Rigs for Long-Term Autonomy". In: *IEEE International Conference on Robotics and Automation (ICRA)* (2013).

[11] Kai Welke et al. "Kinematic Calibration for Saccadic Eye Movements". In: *Technical Report, U. Karlsruhe* (2008).