Extrinsic Calibration of a Camera-Robot System under Non-Holonomic Constraints

Gastón Araguás, Gonzalo Perez Paina, Guillermo Steiner, Luis Canali

Research Centre in Informatics for Engineering (CIII)
National Technological University, Córdoba Regional Faculty (UTN-FRC)

Abstract. A novel approach for the extrinsic calibration of a camera-robot system, i.e. the estimation of the pose of the camera with respect to the robot coordinate system, is presented. The method is based on the relative pose of a planar pattern as seen by the camera, estimated along a predefined set of simple robot motions. This set has been generated so as to exploit the kinematic constraints imposed by the robot architecture and the relative pose between the pattern and the camera coordinate system. The resulting calibration procedure is very simple, making it suitable to be used in a broad range of applications. Experimental evaluations on both synthetic and real data demonstrate the validity of the proposed method.

1 Introduction

In the field of mobile robotics, and especially on those applications involving some degree of autonomy or interaction, cameras have become one of the most used sensors. In order to relate the observed 3D world to its image representation, different parametric models governing the geometry of the image formation process have been proposed, with the projective (pinhole) camera model being the *de-facto* standard. Among all the parameters of a projective model are those known as *extrinsic*: the set of parameters defining the camera pose with respect to a given reference system in 3D space. Unlike other sensor types, where the reference system can be regarded as being located on a known point inside the physical device, the reference system for a camera depends both on the geometry of the sensor as well on the lens array being used.

Standard camera calibration schemes [15] refer the camera system to an arbitrary pose given by a planar pattern taken from a particular view among the image set used for calibration. The common approach to refer the camera to the robot coordinate system is therefore "measuring by hand". This is not well suited for mobile robots aiming to perform autonomous tasks in complex scenarios, where the control strategy of the robot critically depends on the visual information or where different sensors are used cooperatively.

A possible solution to the problem was presented in [11], where the extrinsic calibration of the camera is performed by looking at known points on the robot's body with the help of a mirror. Their approach makes use of the mechanical

model (CAD) of the robot and does not take into account its possible inaccuracies or deviations (mechanical wear, loss of calibration, etc.). Thus, the proposed solution estimate the pose parameters with respect to a system that can be different from the one determining the robot's kinematics. In [12], a method based on the Extended Kalman Filter (EFK) is proposed. The method integrates the robot's odometry and the angle between a light source and the camera principal axis, which is assumed to be parallel to the navigation plane of the robot thus making the method not well suited for practical problems.

In this paper, a simple yet powerful method for the estimation of the camera pose w.r.t. the actual robot reference system is given. The method makes no assumptions about the robot's mechanical layout and allows to relate the camera coordinate system to that of greatest importance for autonomous navigation: the robot's reference frame. The method uses a fiducial marker, similar to those used in augmented reality applications [7], to track the relative pose of the camera along a small set of motions. The resulting trajectories, as seen by the camera, are then used to infer its relative position and orientation with respect to the system that constrained them. As the method doesn't uses the robot's odometry, it is not faced with problems arising from error integration along trajectories.

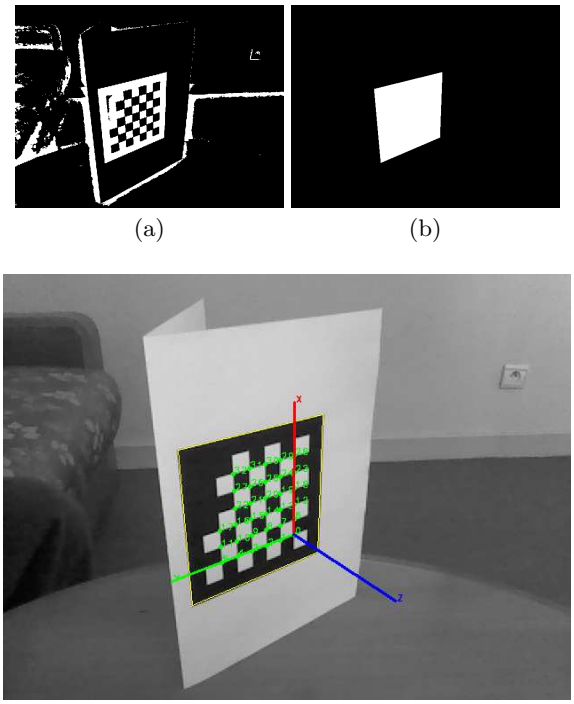
The paper is organized as follows: In Section 2, a brief overview of the fiducial detection and pose estimation system used in the experiments is given. The description is not intended to be complete but to show the main characteristics of the system over which the calibration method is built. In Section 3, the model for a camera mounted statically on a non-holonomic mobile robot is presented. Sections 4.1 and 4.2 formalizes the calibration method for both the translation vector and rotation matrix that uniquely define the pose of the camera w.r.t. the robot reference system. Experimental results on both simulated and real data are presented in Section 5. Finally, conclusions are drawn in Section 6.

2 Fiducial detection and pose estimation

The fiducial detection and pose estimation system consists of a planar pattern with a chessboard-like structure, where the L shaped pattern of interior points (X-shaped corners) limits the number of possible solutions determining its pose, as shown in Fig. 1.

The detection process consists of the following steps:

1. *Adaptive image binarization* by means of locally defined thresholds, as in [6]. This processing step requires the generation of an "integral image" representation of the input image, where each pixel on the former has a value equal to the sum of those lying within the block defined by the current pixel location and the image top-left corner of the last. This representation is kept in memory for further usage.
2. *Quadrilateral detection* by means of simple heuristics (straight segment detection and filtering) on the contours extracted from the binary image obtained during the previous step.



(a) (b)

(c)

Fig. 1. Example of the fiducial detection system, showing a) the binarized image, b) the detected quadrilateral pattern and c) the control points extracted superimposed on the original image, together with the estimated pose.

3. *Control point extraction* by detection of local maxima of the function

$$H_\sigma(x,y) = \max\{0, (I_{xy}(x,y))^2 - I_{xx}(x,y)I_{yy}(x,y)\} \quad (1)$$

with I_{xx}, I_{xy} and I_{yy} the second spatial derivatives of the input image $I(x,y)$ computed at scale σ. Note that (1) is just the negative of the Hessian matrix, giving a high score to regions having saddle-point like structures (X-shaped corners).

The use of saddle-points instead of more usual corner points [10], was motivated by the displacement effect that is observed on the last one when the image defocusing increases [3, 13]. This is particularly important when using lenses with relatively shorts depth-of-fields.

In order to speed-up the detection process, derivatives in (1) are computed using a block approximation as in [4] by means of the integral image generated during the binarization step.

Once the fiducial marker has been detected and its control-points matched against the model, the 3D pose of the pattern is estimated. Here, a closed form

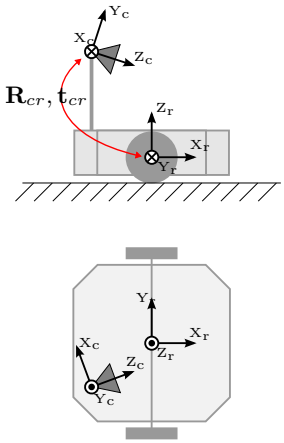


Fig. 2. Schematic diagram of the Robot-Camera model.

solution has been obtained [8] and used as initial guess for a non-linear optimization.

Fig. 1 shows an example of each of the previously described steps.

3 Robot-Camera model

Let's consider a projective camera with known intrinsic parameters [15], mounted on a non-holonomic unicycle mobile robot as shown schematically in Fig. 2. Using homogeneous coordinates, a point $\mathbf{x} \in \mathbb{R}^2$ on the image plane can be modelled as the projection of a 3D point $\mathbf{X} \in \mathbb{R}^3$ by means of the following transformation

$$\tilde{\mathbf{x}} = \mathbf{P}_{cr}\tilde{\mathbf{X}} = (\mathbf{R}_{cr}|\mathbf{t}_{cr})\,\tilde{\mathbf{X}} \tag{2}$$

where $\tilde{\mathbf{x}} = (\mathbf{x}^T, 1)^T$, $\tilde{\mathbf{X}} = (\mathbf{X}^T, 1)^T$ and with $\mathbf{R}_{cr} \in SO(3)$ and $\mathbf{t}_{cr} \in \mathbb{R}^3$ the rotation matrix and translation vectors respectively that determine the 3D pose of the camera coordinate system (CCS) with respect to the robot coordinate-system (RCS). Note that the intrinsic calibration matrix has been factored out in (2), since it does not modifies the present analysis[1].

As it is usual in the literature, the RCS is considered to be located on the midpoint of the wheel's axle and oriented in such that its Y_r-axis and $\{\mathrm{XY}\}_r$-plane are parallel to the wheel's rotation axis and the ground plane respectively. The wheel's diameter d_w and the inter-wheel distance b (wheelbase) are assumed to be known. This implies that the robot's odometry is fully calibrated, e.g. by the methods proposed in [5] or [1].

[1] This can always be done by pre-multiplying the homogeneous image-coordinates by the inverse of the (non-singular) calibration matrix.

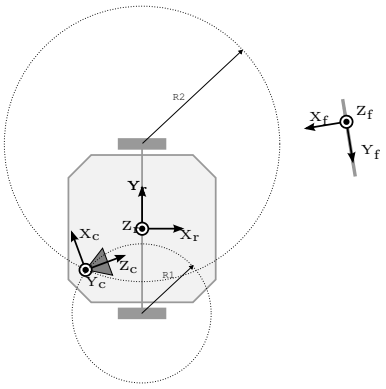


Fig. 3. Calibration of the translation vector, $\mathbf{t}_{cr}$. The dashed lines show the trajectories of the camera.

4 Robot-Camera Extrinsic Calibration

4.1 Calibration of the translation vector, $\mathbf{t}_{cr}$

For the case of a non-holonomic unicycle robot, the translation vector $\mathbf{t}_{cr} = (\mathrm{X_{cr}}, \mathrm{Y_{cr}}, \mathrm{Z_{cr}})^T$ that relates the CCS to the RCS can be estimated by exploiting the constraints imposed by the robot architecture as follows. Let the robot perform a pair of circular motions around each of its wheels while estimating the pose of a fiducial marker located in 3D space, as shown schematically in Fig. 3.

If we denote the inter-wheel distance as b, the $\mathrm{X_{cr}}$ and $\mathrm{Y_{cr}}$ coordinates of the translation vector are solutions of the system

$$\begin{cases} x^2 + \left(y + \frac{b}{2}\right)^2 = R_1^2 \\ x^2 + \left(y - \frac{b}{2}\right)^2 = R_2^2 \end{cases} \tag{3}$$

which are easily shown to be

$$\mathrm{X_{cr}} = \pm\sqrt{R_1^2 - \left(\mathrm{Y_{cr}} + \frac{\mathrm{b}}{2}\right)^2} \tag{4}$$

$$\mathrm{Y_{cr}} = \frac{R_1^2 - R_2^2}{2b} \tag{5}$$

and with the sign of (4) according to which side of the wheel's axis the camera is located on.

It can be seen from the above expressions that the first two components of the translation vector $\mathbf{t}_{cr}$ can be estimated once the radii R_1 and R_2 are known. A simple procedure to estimate those radii arises after observing that, in the FCS (Fiducial Coordinate System), the 3D points describing the position of the

camera along each circular motion lie all on the same spatial plane and describe on it a set of circle-arcs whose radii are those needed to solve (4)–(5).

Let $\{\mathbf{p}_i, i = 1 \cdots N_p\}$ be the set of 3D points determining the position (translation vector) of the camera in the FCS estimated along each of the described motions[2] and let $\mathbf{p}_0$ be its centroid. It can be shown that the *Best-Fit* plane [2], i.e. the plane that minimizes the sum of orthogonal distances from observed points to it, corresponds to that of parameters given by the vector $\mathbf{p}_0$ and the eigenvector corresponding to the smallest eigenvalue of $\mathbf{M}$, the covariance matrix of the point set

$$\mathbf{M} = \sum_{i=1}^{N} (\mathbf{p}_i - \mathbf{p}_0) (\mathbf{p}_i - \mathbf{p}_0)^T \tag{6}$$

Let $\mathbf{n} = (n_x, n_y, n_z)^T$, $\|\mathbf{n}\| = 1$, be the normal vector describing the orientation of this plane. The rotation matrix that brings the Z_f-axis of the FCS parallel to $\mathbf{n}$ can be computed by means of the Rodrigues formula[3] as

$$\mathbf{R}_f = \mathbf{I} + \sin\theta_p [\mathbf{n}_\perp]_\times + (1 - \cos\theta_p) \left(\mathbf{n}_\perp \mathbf{n}_\perp^T - \mathbf{I}\right) \tag{7}$$

where $\mathbf{I}$ denotes the 3×3 identity matrix, $\mathbf{Z}_f$ is the unit vector representing the Z_f-axis, $\theta_p = \cos^{-1}(\mathbf{n} \cdot \mathbf{Z}_f)$ and $\mathbf{n}_\perp = \mathbf{Z}_f \times \mathbf{n}$ a vector orthogonal to the normal $\mathbf{n}$ and the Z_f-axis.

The matrix $\mathbf{R}_f$ can thus be used to rotate the 3D points (forcing all points to have approximately the same z coordinate in the FCS), after which radii R_1 and R_2 can be estimated, e.g. by some of the methods described in [9], from the set of transformed points.

Once the estimation of $\mathrm{X_{cr}}$ and $\mathrm{Y_{cr}}$ has been performed, a complete calibration of the translation vector $\mathbf{t}_{cr}$ can be obtained by computing the pose of the camera relative to a fiducial marker lying on the floor, in which case

$$\mathrm{Z_{cr}} = |t_z| - \frac{d_w}{2} \tag{8}$$

with d_w the diameter of the wheels and t_z the z component of the translation vector estimated in this way.

4.2 Calibration of the rotation matrix, $\mathbf{R}_{cr}$

To find out the rotation matrix $\mathbf{R}_{cr}$ that aligns the CCS and the RCS the follow observation based on the non-holonomic nature of the robot must be done. First, all the observed 3D points determining the fiducial marker position during any robot motion must lie in a plane parallel to the XY_r-plane of the RCS, thus normal vector $\mathbf{n}$ describing the orientation of this plane (found in the previous section from eq. 6) is parallel to $\mathrm{Z_r}$-axis. Second, the direction of the robot

[2] Note that this set of points corresponds to the union of the sets obtained independently along each circular motion.

[3] The notation $[\mathbf{v}]_\times$ refers to the cross-product matrix for the vector $\mathbf{v}$.

moving in a linear motion is parallel to the X_r-axis. This direction vector can be estimated by means of the following simple procedure, shown schematically in Fig. 4: let the robot make a back/forward motion (zero angular velocity) while observing the fiducial marker. Relative to the fiducial, the camera reference point will describe a line in 3D space with a direction-vector parallel to the X_r-axis of the RCS.

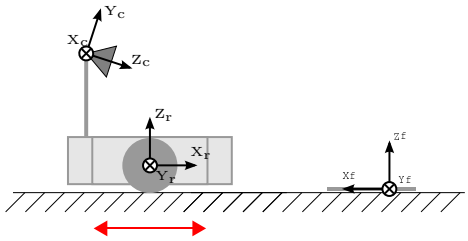


Fig. 4. Calibration of the rotation matrix, $\mathbf{R}_{cr}$.

Let $\{\tilde{\mathbf{p}}_i, i = 1 \cdots N_{\tilde{p}}\}$ be the set of 3D points determining the position (translation vector) of the fiducial marker along the linear motion. It can be shown that the direction of the *Best-Fit* line (in an orthogonal distance sense) is given by the eigenvector corresponding to the largest eigenvalue of the covariance matrix of the point set.

With these two axis directions known, the rotation matrix $\mathbf{R}_{cr}$ can be constructed as

$$\mathbf{R}_{cr} = [\mathbf{q} \quad (\mathbf{n} \times \mathbf{q}) \quad \mathbf{n}] \tag{9}$$

where $\mathbf{q} = (q_x, q_y, q_z)^T$, $\|\mathbf{q}\| = 1$, represents the direction vector of the line, $\mathbf{n}$ is the normal vector to the best fitting plane mentioned before, and $(\mathbf{n} \times \mathbf{q})$ is the cross product between both.

5 Results

We have applied the proposed method to both simulated and real data.

5.1 Simulated data

In order to evaluate the accuracy of the proposed camera robot calibration method a relative error measure is defined, wich reflects the error between the estimated and current values of radii R_1 and R_2 that appears in (4) and (5). It is defined as

$$err_\xi = \frac{|R_\xi - \hat{R}_\xi|}{|R_\xi|} \tag{10}$$

with R_ξ ($\hat{R}_\xi$) denoting the ξ component of the actual (estimated) radius of each circle. It is noted that the z component is estimated independently from the pose of a fiducial marker located at a known distance of the ground, so this error is concerning to the x and y components of the camera position.

All simulations where carried out with the CCS located at $\mathbf{t}_{cr} = \left(\frac{b_w}{2}, 0, 0\right)^T$ and $\mathbf{R}_{cr} = diag(1, 1, 1)$. In order to simulate the pose estimation error (of the camera w.r.t. the fiducial), Gaussian noise of zero mean and a given standard deviation was added. Figure 5 illustrate the mean and standard deviation from the mean for the considered error measures as a function of the noise level, estimated over 100 runs. As it can be seen, the relative error grows quikly up

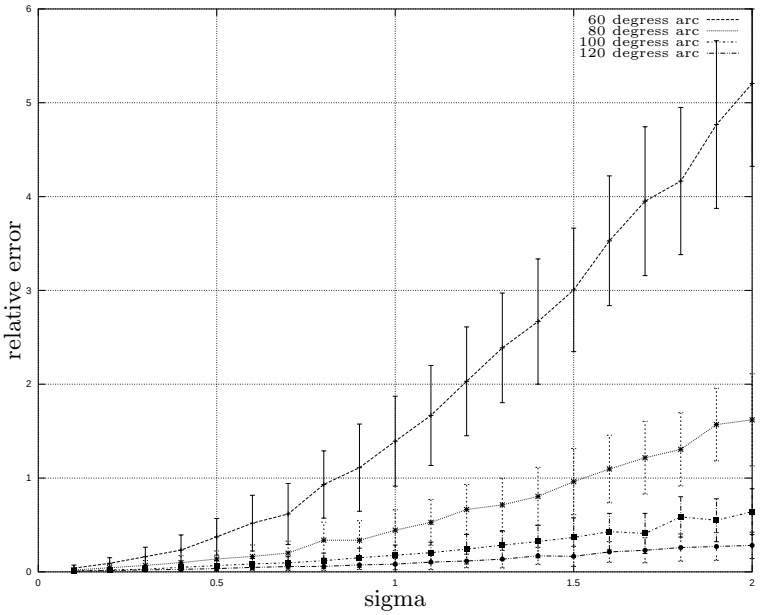


Fig. 5. Relative error of estimated radio vs. noise level

with 60 degrees arc, and becomes more stable over 100 degrees arc. This was to be expected, as in the last case the estimation of the plane normal and circle-arc radii are better conditioned. It is underlined that the radii estimation should be performed in a non-linear fashion in order to avoid the curvature bias of linear least-square methods [14].

It is also noted that as the noise level increases so does the estimation error. In a real scenario, the estimation error is related to the distance from which

the fiducial marker is observed. This implies that the fiducial marker should be located as close as possible to the robot. In our case the noise level in $|\mathbf{t}_{cf}|$ of the estimated pose is less than 1%, that is represented in Figure 5 by $sigma = 1$, which means that to keep the relative error under 1% in the estimation of radii, arcs of 80 degrees must be performed.

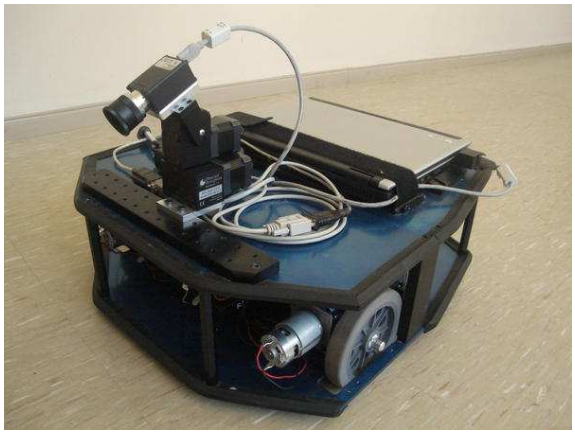


Fig. 6. Mobile robot.

5.2 Real data

The robot-camera system used in our experiments is shown on Figure 6. It consist of a non-holonomic robot platform of parameters $d_w = 138mm$ and $b_w = 455mm$ and a camera located along the X_r-axis of the robot at an approximate distance of $70mm$ from the wheels axle in this direction, $20mm$ along the Y_r-axis and $270mm$ along the Z_r-axis. The camera is oriented in a down-looking manner with a tilt angle of 30 degrees approximately. This angle was measured by setting a pan-and-tilt unit over which the camera was mounted. However, in this way we can only ensure that the chasis of the camera has the so-determined pose, but not necesarily the sensor itself. The intrinsic parameters of the camera were estimated off-line.

The robot describes linear and circular motions around each of its wheels and the pose of the fiducial pattern are estimated and saved for off-line processing. For the reference pose (ground plane) we choose one from the set generated during the linear motion. Table 1 shows the mean (in 15 runs) of the estimated parameters, indicating the standard deviation between parentheses. It can be seen that the estimation of Y_{cr} is the most difficult one. We observed that the accuracy of this value is mostly dependent on the accuracy of the estimated fiducial pose. One way to stabilize this value could be by means of Kalman

filtering. Nevertheless, the estimated values are really close those (carefully) measured by hand.

Table 1. Calibration results.

parameter	estimated mean (std. dev.)	measured by hand
X_{cr}	72.6 (1.86)	70
Y_{cr}	16.5 (8.6)	20
Z_{cr}	277.9 (0.7)	270
tilt angle	29.09 (0.04)	30

6 Conclusions

A new method for the calibration of the pose of a camera mounted on a non-holonomic mobile robot was proposed. By tracking the relative pose of a planar marker along a small set of motions, the parameters determining the position and orientation of the camera where estimated. Although the intrinsic calibration can be easily incorporated during the procedure, the method still relies on the robot design parameters, which have to be calibrated. The calibration of the complete system constitute a perspective of future research.

References

1. T. Abbas, M. Arif, and W. Ahmed. Measurement and correction of systematic odometry errors caused by kinematics imperfections in mobile robots. In *SICE-ICASE, 2006. International Joint Conference*, pages 2073–2078, Oct. 2006.
2. S.J. Ahn, W. Rauh, H.S. Cho, and H.J. Warnecke. Orthogonal distance fitting of implicit curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:620–638, 2002.
3. Luis Alvarez and Freya Morales. Affine morphological multiscale analysis of corners andmultiple junctions. *Int. J. Comput. Vision*, 25(2):95–107, 1997.
4. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.
5. J. Borenstein. Correction of systematic odometry errors in mobile robots. In *IROS '95: Proceedings of the International Conference on Intelligent Robots and Systems-Volume 3*, page 3569, Washington, DC, USA, 1995. IEEE Computer Society.
6. Derek Bradley and Gerhard Roth. Adaptive thresholding using the integral image. *journal of graphics, gpu, and game tools*, 12(2):13–21, 2007.
7. Mark Fiala. Artag, a fiducial marker system using digital techniques. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 590–596, Washington, DC, USA, 2005. IEEE Computer Society.
8. Paul D. Fiore. Efficient linear solution of exterior orientation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):140–148, 2001.

9. Walter Gander, Gene H. Golub, and Rolf Strebel. Least-squares fitting of circles and ellipses. *BIT Numerical Mathematics*, 34:558–578, 1994.
10. C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
11. J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis. Determining the camera to robot-body transformation from planar mirror reflections. In *Proceedings of the IEEE/RSJ International Conference on Robotics and Intelligent Systems (IROS)*, pages 3865–3871, Nice, France, September 22-26 2008.
12. Agostino Martinelli and Davide Scaramuzza. Automatic self-calibration of a vision system during robot motion. In *in Proc. of the IEEE International Conference on Robotics and Automation, ICRA06*, 2006.
13. K. Mikolajczyk. Detection of local features invariant to affine transformations. *Ph.D. Thesis*, 2002.
14. Zhengyou Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15:59–76, 1997.
15. Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330–1334, 1998.