Planar Motion and Hand-eye Calibration using Inter-image Homographies from a Planar Scene

Mårten Wadenbäck and Anders Heyden
Centre for Mathematical Sciences, Lund University, Lund, Sweden

Keywords: Visual Navigation, Planar Motion, SLAM, Homography, Tilted Camera.

Abstract: In this paper we consider a mobile platform performing partial hand-eye calibration and Simultaneous Localisation and Mapping (SLAM) using images of the floor along with the assumptions of planar motion and constant internal camera parameters. The method used is based on a direct parametrisation of the camera motion, combined with an iterative scheme for determining the motion parameters from inter-image homographies. Experiments are carried out on both real and synthetic data. For the real data, the estimates obtained are compared to measurements by an industrial robot, which serve as ground truth. The results demonstrate that our method produces consistent estimates of the camera position and orientation. We also make some remarks about patterns of motion for which the method fails.

1 INTRODUCTION

The development of algorithms for Simultaneous Localisation and Mapping (SLAM) has been a major focus in robotics research the past few decades. Such algorithms aim at enabling a mobile platform to explore and map its surroundings, while at the same time maintaining accurate knowledge of its position. Many types of sensors may be used to this end, and are often combined to supplement each other.

For the mapping part of SLAM, a reconstruction (broadly interpreted) must be created from the scene. For some work on SLAM using visual sensors, see for example (Davison, 2003), (Karlsson et al., 2005) and (Koch et al., 2010). Scene reconstruction from images is a well studied problem in computer vision, and is still a very active research area. Since the introduction of the fundamental matrix in (Faugeras, 1992) and (Hartley, 1992), epipolar geometry has been the foundation of many successful approaches to visual reconstruction.

However, a planar or near-planar scene is well known to be a degenerate or ill-conditioned case for reconstruction based on the fundamental matrix and similar approaches. Since planar scenes and objects are very common in man-made or indoor environments, a navigation system intended to operate in such environments must take care to avoid degeneracy. Planar homographies, on the other hand, are particularly well suited to planar scenes, but are unable to describe general 3D structure. This insight has been utilised for visual navigation in (Liang and Pears, 2002) and (Hajjdiab and Laganière, 2004), among others.

In this paper we shall consider a single camera, with square pixels and zero skew, moving at a constant height above the floor. We will further assume that the internal parameters of the camera are constant, and that the camera orientation is fixed except for a rotation about the normal to the floor plane. Using inter-image homographies not only avoids the degeneracy issue mentioned above, but in addition allows us to use an explicit parametrisation of this particular kind of camera motion.

In the case where the inter-image homographies describe a Euclidean (or, in general, an affine) transformation, the motion parameters are easily recovered using the QR decomposition. This happens when the image plane is parallel to the floor. In the presence of a tilt, however, it is not as straightforward to extract the motion information from the homographies. The main contribution of this paper is a method to compute both the tilt and the motion information from a single homography.

Another reason for estimating the tilt is that only rectified images may be stitched consistently into a mosaic. A visual navigation system based on a sparse feature based map of the floor plane also needs rectified images to construct the map. It is in general not trivial to mount a camera with very high precision, so

Wadenbäck M. and Heyden A..
Planar Motion and Hand-eye Calibration using Inter-image Homographies from a Planar Scene.
DOI: 10.5220/0004293701640168
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 164-168
ISBN: 978-989-8565-48-8

avoiding the need for this would be useful.

Determining the tilt can also be seen as a partial hand-eye calibration. The original formulation of the hand-eye calibration problem was to recover the relative orientation between a robot arm and a camera mounted on the arm. Tsai and Lenz showed that with known 3D feature points, known motion of the robot arm, and known transformations A and B, the unknown relative orientation X can be determined from the equation $AX = XB$ (Tsai and Lenz, 1989). The problem was later reformulated using quaternions to parametrise rotations and 3×4 camera matrices instead of classical transformation matrices (Horaud and Dornaika, 1995).

2 CAMERA PARAMETRISATION

We assume that the camera is mounted rigidly onto a mobile platform, and directed towards the floor. This means that the position and orientation of the camera can be parametrised by a translation vector t and a rotation in the floor plane of an angle φ. The tilt is described by the constant angles ψ and θ. Both the translation and the three angles will be estimated. The camera is assumed to move in the plane $z = 0$, and the ground plane is taken to be at $z = 1$. This is not a restriction, since it only reflects our choice of the world coordinate system and global scale fixation (corresponding to the unknown focal length).

We will consider two consecutive images, A and B, with associated camera matrices

$$\begin{aligned} P_A &= R_{\psi\theta}[I \mid 0], \\ P_B &= R_{\psi\theta}R_{\varphi}[I \mid -t]. \end{aligned} \tag{1}$$

Here $R_{\psi\theta}$ is a rotation of θ around the y-axis followed by a rotation of ψ around the x-axis, and R_{φ} is a rotation of φ around the z-axis (the floor normal).

Using (1), one can easily verify that the homography H from A to B is

$$H = \lambda R_{\psi\theta}R_{\varphi}TR_{\psi\theta}^T, \tag{2}$$

for any non-zero $\lambda \in \mathbb{R}$ and with

$$T = \begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

3 TILT ESTIMATION

The presence of a tilt gives rise to perspective effects. These distort the geometry perceived by the camera, and prevent easy extraction of motion information. If the tilt angles ψ and θ can be determined, one can rectify the image and then use the QR decomposition to retrieve the translation t and the free rotation φ.

To estimate ψ and θ, we derive equations that contain these angles but which do not contain t and φ. These equations will then be solved using an iterative scheme.

3.1 Eliminating φ

Separating the tilt angles ψ and θ from the motion parameters t and φ in (2), we get

$$R_{\psi\theta}^T H R_{\psi\theta} = \lambda R_{\varphi} T. \tag{4}$$

Here, one notes that R_{φ} can be eliminated by multiplying with the transpose from the left on both sides. This results in the relation

$$R_{\psi\theta}^T M R_{\psi\theta} = \lambda^2 T^T T, \tag{5}$$

with (symmetric)

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ m_{13} & m_{23} & m_{33} \end{bmatrix} = H^T H. \tag{6}$$

Since both sides of (5) are symmetric matrices, one obtains six unique equations. Let $\mathcal{L} = R_{\psi\theta}^T M R_{\psi\theta}$ and $\mathcal{R} = \lambda^2 T^T T$ be the left and right hand sides of (5), respectively. Evaluating $\mathcal{R}$, one obtains

$$\mathcal{R} = \lambda^2 \begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ -t_x & -t_y & 1 + t_x^2 + t_y^2 \end{bmatrix}. \tag{7}$$

3.2 Iterative Scheme

As described in Section 2, $R_{\psi\theta} = R_{\psi}R_{\theta}$ is a rotation of θ around the y-axis followed by a rotation of ψ around the x-axis. Direct multiplication of the rotation matrices allows us to evaluate $\mathcal{L}$ (though this margin is too narrow to contain the result), and one finds that $\mathcal{L}$ is a fourth degree expression in $c_{\psi} = \cos\psi$, $s_{\psi} = \sin\psi$, $c_{\theta} = \cos\theta$ and $s_{\theta} = \sin\theta$.

Noting that $\mathcal{R}_{11}$, $\mathcal{R}_{12}$ and $\mathcal{R}_{22}$ are independent of t, the equations for ψ and θ become

$$\begin{cases} \mathcal{L}_{11} - \mathcal{L}_{22} = 0 \\ \mathcal{L}_{12} = 0 \end{cases}. \tag{8}$$

But instead of trying to solve (8) for both ψ and θ at the same time, we will iteratively alternate between solving for one angle, with the other held fixed. This reduces the problem of solving a fourth degree trigonometric equation, so that we instead iterate and solve a second degree equation in each iteration.

Before explaining in detail how these equations are solved, we first outline in Algorithm 1 the iterative scheme which produces an approximation to $R_{\psi\theta}$. Since

$$R_{\psi\theta} = R_\psi R_\theta = \begin{bmatrix} c_\theta & 0 & -s_\theta \\ -s_\psi s_\theta & c_\psi & -s_\psi c_\theta \\ c_\psi s_\theta & s_\psi & c_\psi c_\theta \end{bmatrix} \quad (9)$$

it is trivial to find ψ and θ from this approximation.

Algorithm 1: Iteratively approximate $R_{\psi\theta}$. The steps on line 6 and line 8 are detailed in Sections 3.2.1 and 3.2.2. Since we "embed" into M the current approximation, we may assume that the fixed angle is zero when solving for the free one.

Input: An inter-image homography H
Output: An approximation R of $R_{\psi\theta}$

1: $\widehat{M} \leftarrow H^T H$
2: $\theta_0 \leftarrow 0$
3: $R \leftarrow R_{\theta_0}$
4: **for** $j = 1, \ldots, N$ **do**
5: $\quad \widehat{M} \leftarrow R_{\theta_{j-1}}^T \widehat{M} R_{\theta_{j-1}}$
6: $\quad$ Solve for ψ_j
7: $\quad \widehat{M} \leftarrow R_{\psi_j}^T \widehat{M} R_{\psi_j}$
8: $\quad$ Solve for θ_j
9: **end for**
10: $R \leftarrow R_{\theta_0} R_{\psi_1} R_{\theta_1} R_{\psi_2} R_{\theta_2} \cdots R_{\psi_N} R_{\theta_N}$

3.2.1 Solving for ψ

Since c_ψ and s_ψ cannot both be zero, (8) is equivalent to

$$\begin{cases} \mathcal{L}_{11} - \mathcal{L}_{22} = 0 \\ c_\psi \mathcal{L}_{12} = 0 \\ s_\psi \mathcal{L}_{12} = 0 \end{cases}. \quad (10)$$

By letting $\widehat{M} = R_\theta^T M R_\theta$ this can be written in matrix form as

$$\begin{bmatrix} \widehat{m}_{11} - \widehat{m}_{22} & -2\widehat{m}_{23} & \widehat{m}_{11} - \widehat{m}_{33} \\ \widehat{m}_{12} & \widehat{m}_{13} & 0 \\ 0 & \widehat{m}_{12} & \widehat{m}_{13} \end{bmatrix} \begin{bmatrix} c_\psi^2 \\ c_\psi s_\psi \\ s_\psi^2 \end{bmatrix} = 0. \quad (11)$$

This means that $(c_\psi^2, c_\psi s_\psi, s_\psi^2)$ lies in the null space of the coefficient matrix in (11). For the moment[1], let us assume that we expect a one dimensional null space. Due to measurement noise this will not be the case, so instead we use the singular vector $v = (v_1, v_2, v_3)$ corresponding to the smallest singular value as our null vector.

[1]This is true for most patterns of motion. Section 4.2 demonstrates some specific patterns of motions for which this is fragile.

Provided the singular vector v one obtains ψ as

$$\psi = \frac{1}{2} \arcsin \frac{2v_2}{v_1 + v_3}. \quad (12)$$

3.2.2 Solving for θ

Now θ can be found in much a similar way as ψ. Physical considerations imply that, at least for moderately sized angles, $R_\psi R_\theta$ has approximately the same effect on the camera as $R_\theta R_\psi$. Examination of the matrices confirms this for small angles.

Therefore, if $\widehat{M} = R_\psi^T M R_\psi$, then

$$\begin{cases} \mathcal{L}_{11} - \mathcal{L}_{22} = 0 \\ c_\theta \mathcal{L}_{12} = 0 \\ s_\theta \mathcal{L}_{12} = 0 \end{cases}, \quad (13)$$

can be written in matrix form as

$$\begin{bmatrix} \widehat{m}_{11} - \widehat{m}_{22} & 2\widehat{m}_{13} & \widehat{m}_{33} - \widehat{m}_{22} \\ \widehat{m}_{12} & \widehat{m}_{23} & 0 \\ 0 & \widehat{m}_{12} & \widehat{m}_{23} \end{bmatrix} \begin{bmatrix} c_\theta^2 \\ c_\theta s_\theta \\ s_\theta^2 \end{bmatrix} = 0. \quad (14)$$

We find, in the same way as in Section 3.2.1 that the null vector v can be used to find

$$\theta = \frac{1}{2} \arcsin \frac{2v_2}{v_1 + v_3}. \quad (15)$$

4 EXPERIMENTS

In order to test how well the tilt estimation works in practice, fifty homographies of the form (2) were generated with random values for ψ, θ, φ and t. The true angles and their corresponding estimates can be seen in Figure 1.

4.1 Path Reconstruction

A simple path estimation has also been tried on both synthetic and real data using the QR decomposition to determine translation and planar rotation, after estimating the tilt as described in Section 3. In the simulation, noise of the magnitude corresponding to a few pixels have been added to the points used to estimate the homographies. Results for this experiment are shown in Figure 2 and 3.

We have also carried out experiments with real data. A camera mounted onto an industrial robot has been used to take images, from which homographies were computed. The resulting reconstruction can be seen in Figure 4. For comparison, we have additionally estimated the non constant angle φ using a method based on conjugate rotations, see (Liang

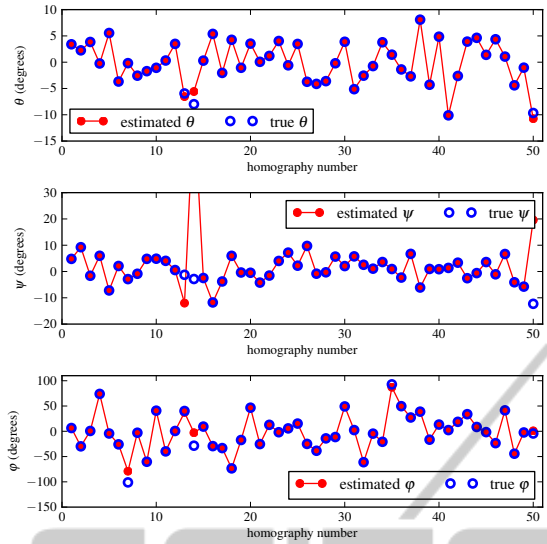

Figure 1: True and estimated values for ψ, θ and φ for fifty randomly generated homographies. As can be seen, the estimation works well in most instances.

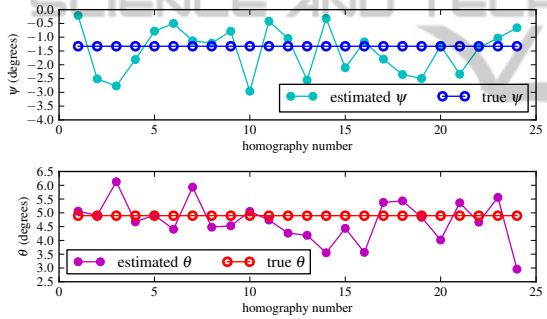

Figure 2: We see that the estimated values of ψ and θ are, on average, close to the true values. Since ψ and θ are constant, temporal filtering could be used to get better estimates over time.

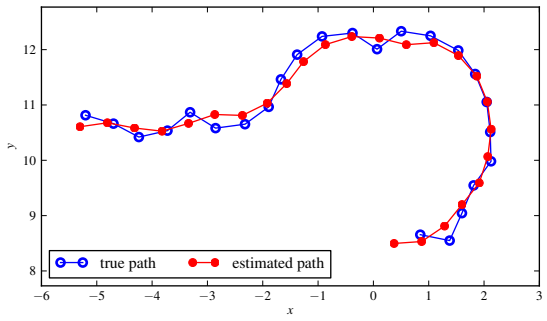

Figure 3: The simulated and the estimated paths. Procrustes analysis has been carried out to align the path curves for easy comparison.

and Pears, 2002) for details. This method computes φ from the eigenvalues of the homography without estimating the tilt. Figure 5 shows this estimate com-

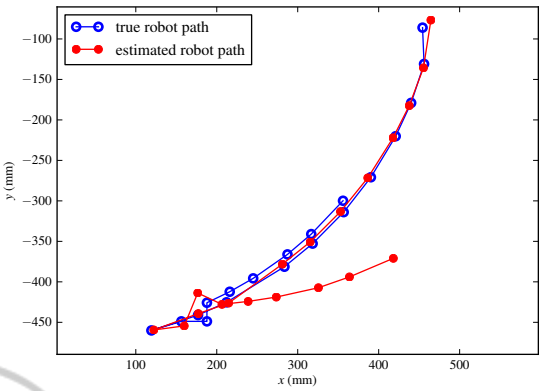

Figure 4: True and estimated paths for the robot experiment. At the lower part of the plot some erroneous estimates are made, which results in the estimated path being deflected away.

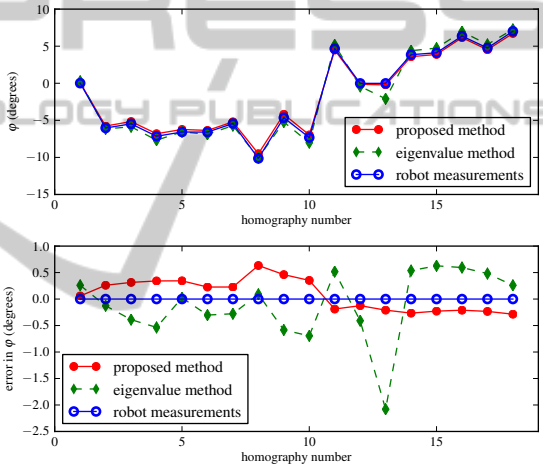

Figure 5: The upper plot shows the difference in orientation between consecutive images, and the lower plot shows the angular error. The plots show that our estimates of φ and the eigenvalue-based estimates of φ are both close to the truth (robot measurements).

Table 1: Mean, median and variance of the magnitude of the angular error. For the eigenvalue-based method, measurement 13 is considered an outlier and has been omitted. Despite this, the proposed method is clearly seen to give more accurate estimates.

	Mean	Median	Variance
Proposed (QR)	0.2759	0.2467	0.0161
Eigenvalue	0.3947	0.4092	0.0403

pared to our estimate and the true value (as measured by the robot). Both methods perform well, however some statistical measures shown in Table 1 suggest that tilt estimation followed by QR decomposition has a slightly favourable performance.

4.2 Ill-Conditioned Motion

Empirical evidence suggests that the instances where the tilt estimation fails are the ones where the translation t is close to either a pure x-translation or a pure y-translation. Randomly generating homographies with this pattern of motion provides further evidence for this. It can further be seen that a pure x-translation gives rise to a poor estimate of ψ, while a pure y-translation results in a poor estimate of θ. Results for this experiment are presented in Figures 6 and 7. Theoretical understanding of this will be necessary if the instability is to be addressed.

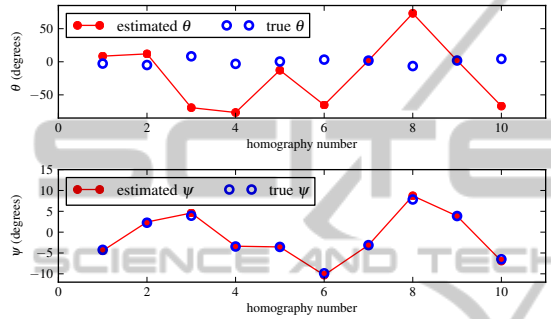

Figure 6: When t is a pure x-translation, ψ seems to be unreliably estimated.

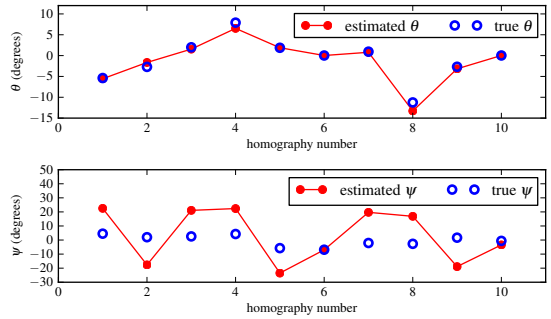

Figure 7: When t is a pure y-translation, θ seems to be unreliably estimated.

5 CONCLUSIONS

Tilt estimation is a prerequisite for constructing consistent floor maps using images from a tilted camera. In this paper we have presented an iterative scheme for determining the tilt from a single homography. Experiments with a simple path reconstruction have been conducted, which show that if the tilt is rectified then the correct Euclidean motion can be found using the QR decomposition. Experiments using synthetic data show that the estimated tilt angles are close to the true tilt angles in most instances, however some especially troublesome motions have been found.

ACKNOWLEDGEMENTS

This work has been funded by the Swedish Foundation for Strategic Research through the SSF project ENGROSS, http://www.engross.lth.se.

REFERENCES

Davison, A. J. (2003). Real-Time Simultaneous Localisation and Mapping with a Single Camera. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, volume 2 of *ICCV '03*, pages 1403–1410, Nice, France. IEEE Computer Society.

Faugeras, O. D. (1992). What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the Second European Conference on Computer Vision*, volume 588 of *ECCV '92*, pages 563–578, Santa Margherita Ligure, Italy. Springer-Verlag.

Hajjdiab, H. and Laganière, R. (2004). Vision-Based Multi-Robot Simultaneous Localization and Mapping. In *CRV '04: Proceedings of the 1st Canadian Conference on Computer and Robot Vision*, pages 155–162, Washington, DC, USA. IEEE Computer Society.

Hartley, R. I. (1992). Estimation of Relative Camera Positions for Uncalibrated Cameras. In *Proceedings of the Second European Conference on Computer Vision*, volume 588, pages 579–587, Santa Margherita Ligure, Italy. Springer-Verlag.

Horaud, R. and Dornaika, F. (1995). Hand-Eye Calibration. *International Journal of Robotics Research*, 14(3):195–210.

Karlsson, N., Bernardo, E. D., Ostrowski, J. P., Goncalves, L., Pirjanian, P., and Munich, M. E. (2005). The vSLAM Algorithm for Robust Localization and Mapping. In *ICRA '05: Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 24–29, Barcelona, Spain. IEEE.

Koch, O., Walter, M. R., Huang, A. S., and Teller, S. J. (2010). Ground Robot Navigation using Uncalibrated Cameras. In *ICRA '10: IEEE International Conference on Robotics and Automation*, pages 2423–2430, Anchorage, Alaska, USA. IEEE.

Liang, B. and Pears, N. (2002). Visual Navigation using Planar Homographies. In *ICRA '02: Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, pages 205–210, Washington, DC, USA.

Tsai, R. and Lenz, R. (1989). A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358.