Relative Pose Estimation in Binocular Vision for a Planar Scene using Inter-Image Homographies

Marcus Valtonen Örnhag and Anders Heyden
Centre for Mathematical Sciences, Lund University, Sweden

Keywords: Relative Pose Estimation, SLAM, Visual Odometry, Binocular Vision, Planar Motion, Homography.

Abstract: In this paper we consider a mobile platform with two cameras directed towards the floor mounted the same distance from the ground, assuming planar motion and constant internal parameters. Earlier work related to this specific problem geometry has been carried out for monocular systems, and the main contribution of this paper is the generalization to a binocular system and the recovery of the relative translation and orientation between the cameras. The method is based on previous work on monocular systems, using sequences of inter-image homographies. Experiments are conducted using synthetic data, and the results demonstrate a robust method for determining the relative parameters.

1 INTRODUCTION

In robotics research, it is of interest to accurately track the position of a mobile robot relative to its surroundings. The emergence of artificial intelligence and autonomous vehicles in recent years demand robust algorithms to handle such problems. During the years of research in the field many kinds of sensors have been used—LIDAR, rotary encoders, inertial sensors and GPS, to mention a few—and often in combination. The type of sensor one chooses to work with restricts what algorithms that can be used, and how the resulting map of the robot and its surroundings will look.

One sensor of particular interest for the robotics and computer vision community is the image sensor and there are many reasons why it is popular. One important factor is that the wide range of algorithms used in computer vision, e.g. visual feature extraction and pose estimation, can be used in this setting; however, from an industrial point of view image sensors are an often considered design choice since they are relatively cheap compared to other sensors. Furthermore, they are often available on consumer products, such as smartphones and tablets, where similar techniques can be used, e.g. in Augmented Reality (AR). With image sensors one is not limited to sparse 3D clouds of feature points, but can model the map using dense and textured 3D models.

Visual SLAM systems have been developed for nearly three decades, with (Harris and Pike, 1988) being one of the first. Since then, several improvements have been made, and with the aid of modern computing power, a variety of methods for real-time SLAM are available. Among the more recent once are MonoSLAM (Davison et al., 2007), LSD-SLAM (Engel et al., 2014) and ORB-SLAM2 (Mur-Artal and Tardós, 2017), where the latter includes support for monocular, stereo and RGB-D cameras.

2 RELATED WORK

In epipolar geometry, the fundamental matrix, introduced by (Faugeras, 1992) and (Hartley, 1992), has been a tool for many algorithms concerning scene reconstruction; however, planar motion is known to be ill-conditioned, see e.g. (Hartley and Zisserman, 2004). The problem geometry considered in this paper is forced to planar motion, which is common in e.g. indoor environments. To overcome this issue algorithms that take advantage of planar homographies have been devised, which by construction are constrained to planar motion and therefore do not suffer from being ill-conditioned. Some early work on planar motion using homographies include that of (Liang and Pears, 2002) and (Hajjdiab and Laganière, 2004). More recent work on ego-motion recovery in a monocular system using inter-image homographies for a planar scene has been covered in (Wadenbäck and Heyden, 2013) for a single homogra-

Örnhag, M. and Heyden, A.
Relative Pose Estimation in Binocular Vision for a Planar Scene using Inter-Image Homographies.
DOI: 10.5220/0006695305680575
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 568-575
ISBN: 978-989-758-276-9

phy and by the same authors for several homographies in (Wadenbäck and Heyden, 2014). In (Wadenbäck et al., 2017) the same methods are used to calibrate the fixed parameters initially, transforming the subsequent problem to a two-dimensional rigid body motion problem.

The stereo rig problem involving two cameras with fixed relative orientation is investigated for autocalibration in (Hartley and Zisserman, 2004). In (Nyman et al., 2010) a method for multi-camera platform calibration using multi-linear constraints is developed; however, this method does not rely on the inter-image homographies, but rather using the camera matrices.

3 THEORY

3.1 Problem Geometry

In this paper we consider a mobile platform with two cameras directed towards the floor mounted the same distance from the ground. By a suitable choice of the world coordinate system the cameras move in the plane $z = 0$ and relative to the ground plane positioned at $z = 1$. Both cameras are assumed to be mounted rigidly onto the platform and no common scene point is assumed to be visible in the cameras simultaneously. Furthermore, the mobile platform's center of rotation is assumed to be located in the first camera center. In this setting the second camera center is connected to the first by a rigid body motion.

The 3D rotations are parametrized using Tait-Bryan angles

$$\boldsymbol{R}(\psi,\theta,\varphi) = \boldsymbol{R}_x(\psi)\boldsymbol{R}_y(\theta)\boldsymbol{R}_z(\varphi), \tag{1}$$

where $\boldsymbol{R}_x$, $\boldsymbol{R}_y$ and $\boldsymbol{R}_z$ denote the rotation around the respective coordinate axes with a given angle. The problem geometry is illustrated in Figure 1.

3.2 Camera Parametrization

As in (Wadenbäck and Heyden, 2013), consider two consecutive images, A and B, for the first camera. The camera matrices are then

$$\begin{aligned} \boldsymbol{P}_A &= \boldsymbol{R}_{\psi\theta}[\boldsymbol{I} \mid \boldsymbol{0}], \\ \boldsymbol{P}_B &= \boldsymbol{R}_{\psi\theta}\boldsymbol{R}_\varphi[\boldsymbol{I} \mid -\boldsymbol{t}], \end{aligned} \tag{2}$$

where $\boldsymbol{R}_{\psi\theta}$ is a rotation θ around the y-axis followed by a rotation of ψ around the x-axis. The movement of the mobile platform is modelled by a rotation φ around the z-axis, corresponding to $\boldsymbol{R}_\varphi$ and translation vector $\boldsymbol{t}$.

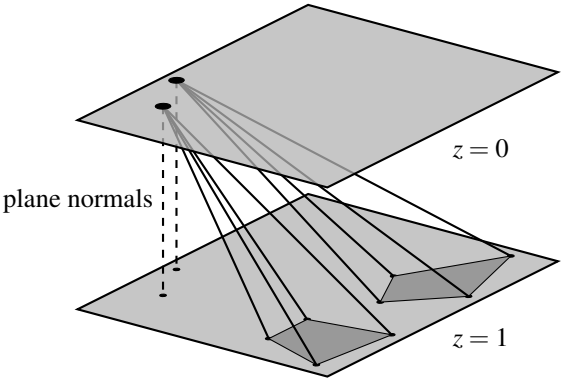

Figure 1: The problem geometry considered in this paper. The cameras are assumed to move in the plane $z = 0$ and the relative orientation between them as well as the tilt towards the floor normal is assumed to be fixed as the mobile platform moves freely.

The camera matrices for the second camera can be parametrized as

$$\begin{aligned} \boldsymbol{P}'_A &= \boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_\eta\boldsymbol{T}_{\boldsymbol{\tau}}[\boldsymbol{I} \mid \boldsymbol{0}], \\ \boldsymbol{P}'_B &= \boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_\eta\boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_\varphi[\boldsymbol{I} \mid -\boldsymbol{t}], \end{aligned} \tag{3}$$

where $\boldsymbol{R}_{\psi'\theta'}$ is the tilt, defined as for the first camera. Furthermore, $\boldsymbol{R}_\eta$ is a fixed rotation by η degrees around the z-axis relative to the first camera, and $\boldsymbol{\tau}$ is the rigid body translation vector between the first and the second camera center. The matrix $\boldsymbol{T}_{\boldsymbol{\tau}}$ corresponds to a translation by $\boldsymbol{\tau}$ defined as $\boldsymbol{T}_{\boldsymbol{\tau}} = \boldsymbol{I} - \boldsymbol{\tau}\boldsymbol{n}^T$, where $\boldsymbol{n} = (0,\ 0,\ 1)^T$ is a floor normal.

3.3 Homographies

Given point correspondences $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, in homogeneous coordinates, the homography $\boldsymbol{H}$ transforms the points such that $\boldsymbol{x}_2 = \lambda\boldsymbol{H}\boldsymbol{x}_1$, where $\lambda \neq 0$ is due to universal scale ambiguity. In (Wadenbäck and Heyden, 2013) the homography for the first camera is derived and is given by

$$\lambda\boldsymbol{H} = \boldsymbol{R}_{\psi\theta}\boldsymbol{R}_\varphi\boldsymbol{T}_{\boldsymbol{t}}\boldsymbol{R}_{\psi\theta}^T\,. \tag{4}$$

Similarly, the homography $\boldsymbol{H}'$ for the second camera is given by

$$\lambda'\boldsymbol{H}' = \boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_\eta\boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_\varphi\boldsymbol{T}_{\boldsymbol{t}}\boldsymbol{T}_{\boldsymbol{\tau}}^{-1}\boldsymbol{R}_\eta^T\boldsymbol{R}_{\psi'\theta'}^T\,. \tag{5}$$

3.4 Parameter Recovery

By separating the fixed angles from φ and the translation $\boldsymbol{t}$ the following relation holds

$$\boldsymbol{R}_\varphi\boldsymbol{T}_{\boldsymbol{t}} = \lambda\boldsymbol{R}_{\psi\theta}^T\boldsymbol{H}\boldsymbol{R}_{\psi\theta} = \lambda'\boldsymbol{T}_{\boldsymbol{\tau}}^{-1}\boldsymbol{R}_\eta^T\boldsymbol{R}_{\psi'\theta'}^T\boldsymbol{H}'\boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_\eta\boldsymbol{T}_{\boldsymbol{\tau}}, \tag{6}$$

It is shown in (Wadenbäck and Heyden, 2013) how to recover the parameters for the monocular case, and

by doing so the parameters ψ, θ, φ and $\boldsymbol{t}$ aswell as ψ' and θ' can be recovered; the latter two from treating the second camera as a monocular system. Furthermore, we shall assume that all homographies $\boldsymbol{H}$ are normalized such that $\det \boldsymbol{H} = 1$.

3.4.1 Recovering the Relative Translation $\boldsymbol{\tau}$

The relative translation and rotation can be separated by putting (3.4) in the form

$$\boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_{\varphi}\boldsymbol{T}_{\boldsymbol{t}}\boldsymbol{T}_{\boldsymbol{\tau}}^{-1} = \lambda'\boldsymbol{R}_{\eta}^{T}\boldsymbol{R}_{\psi'\theta'}^{T}\boldsymbol{H}'\boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_{\eta}, \tag{7}$$

and multiplying with the transpose from the left on both sides yield

$$\boldsymbol{T}_{\boldsymbol{t}-\boldsymbol{\tau}}^{T}\boldsymbol{R}_{\varphi}^{T}\boldsymbol{T}_{\boldsymbol{\tau}}^{T}\boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_{\varphi}\boldsymbol{T}_{\boldsymbol{t}-\boldsymbol{\tau}} = \lambda'\boldsymbol{R}_{\eta}^{T}\boldsymbol{R}_{\psi'\theta'}^{T}\boldsymbol{H}'^{T}\boldsymbol{H}'\boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_{\eta}\,. \tag{8}$$

The left hand side of (3.4.1) can be simplified to

$$\mathscr{L} = \begin{bmatrix} 1 & 0 & \ell_1 \\ 0 & 1 & \ell_2 \\ \ell_1 & \ell_2 & \ell_3 \end{bmatrix}, \tag{9}$$

where

$$\begin{aligned} \ell_1 &= \tau_x - t_x - \tau_y \sin\varphi - \tau_x \cos\varphi, \\ \ell_2 &= \tau_y - t_y + \tau_x \sin\varphi - \tau_y \cos\varphi, \\ \ell_3 &= k_1\tau_x + k_2\tau_y + c\tau_x^2 + c\tau_y^2 + |\boldsymbol{t}|^2 + 1, \end{aligned} \tag{10}$$

and

$$\begin{aligned} k_1 &= 2(t_x\cos\varphi - t_y\sin\varphi - t_x), \\ k_2 &= 2(t_x\sin\varphi + t_y\cos\varphi - t_y), \\ c &= 2(1-\cos\varphi)\,. \end{aligned} \tag{11}$$

The eigenvalues of $\mathscr{L}$ are given by $\lambda_2 = 1$ and λ_1, λ_3 such that $\lambda_1\lambda_3 = \ell_3 - \ell_1^2 - \ell_2^2 = 1$. Furthermore, the right hand side of (3.4.1) has the same eigenvalues as $\boldsymbol{H}'^{T}\boldsymbol{H}'$, as they are similar. Since the sum of the eigenvalues is the trace of the corresponding matrix, the following relation holds

$$\operatorname{tr}\boldsymbol{H}'^{T}\boldsymbol{H}' = 2 + \ell_3, \tag{12}$$

which is independent of η. By letting $h = \operatorname{tr}\boldsymbol{H}'^{T}\boldsymbol{H} - 3 - |\boldsymbol{t}|^2$ the relation becomes

$$k_1\tau_x + k_2\tau_y + c\tau_x^2 + c\tau_y^2 - h = 0\,. \tag{13}$$

The other invariants do not give any new relations for $\boldsymbol{\tau}$ since, $\det\mathscr{L} = 1$ and $\frac{1}{2}\left((\operatorname{tr}\mathscr{L})^2 - \operatorname{tr}\mathscr{L}^2\right) = \operatorname{tr}\mathscr{L}$.

3.4.2 Solving for the Relative Translation $\boldsymbol{\tau}$

With only one pair of homographies one cannot determine $\boldsymbol{\tau}$ explicitly; however, using multiple pairs one equation on the form (13) for each pair of homography is given, which yields a system of equations

$$\begin{aligned} k_1^{(1)}\tau_x + k_2^{(1)}\tau_y + c^{(1)}(\tau_x^2+\tau_y^2) - h^{(1)} &= 0, \\ k_1^{(2)}\tau_x + k_2^{(2)}\tau_y + c^{(2)}(\tau_x^2+\tau_y^2) - h^{(2)} &= 0, \\ &\vdots \\ k_1^{(n)}\tau_x + k_2^{(n)}\tau_y + c^{(n)}(\tau_x^2+\tau_y^2) - h^{(n)} &= 0\,. \end{aligned} \tag{14}$$

The system in (14) is over-determined for $n > 2$, hence minimizing

$$\min_{\boldsymbol{\tau}\in\mathbb{R}^2} \sum_{i=1}^{n} \left| k_1^{(i)}\tau_x + k_2^{(i)}\tau_y + c^{(i)}(\tau_x^2+\tau_y^2) - h^{(i)} \right|^2, \tag{15}$$

gives the desired result. This can be re-formulated as

$$\min_{\boldsymbol{\tau}\in\mathbb{R}^2} \|\boldsymbol{K}\boldsymbol{\tau} + \boldsymbol{c}\boldsymbol{\tau}^T\boldsymbol{\tau} - \boldsymbol{h}\|_2^2, \tag{16}$$

where

$$\boldsymbol{K} = \begin{bmatrix} k_1^{(1)} & k_2^{(1)} \\ k_1^{(2)} & k_2^{(2)} \\ \vdots & \vdots \\ k_1^{(n)} & k_2^{(n)} \end{bmatrix}, \boldsymbol{c} = \begin{bmatrix} c^{(1)} \\ c^{(2)} \\ \vdots \\ c^{(n)} \end{bmatrix} \text{ and } \boldsymbol{h} = \begin{bmatrix} h^{(1)} \\ h^{(2)} \\ \vdots \\ h^{(n)} \end{bmatrix} \tag{17}$$

By introducing a new variable $r = |\boldsymbol{\tau}|^2$, an equivalent problem is obtained

$$\min_{\boldsymbol{\tau}\in\mathbb{R}^2,\, r=|\boldsymbol{\tau}|^2} \|\boldsymbol{K}\boldsymbol{\tau} + \boldsymbol{c}r - \boldsymbol{h}\|_2^2 = \min_{\boldsymbol{x}\in\mathbb{R}^3,\, r=|\boldsymbol{\tau}|^2} \|\boldsymbol{M}\boldsymbol{x} - \boldsymbol{h}\|_2^2, \tag{18}$$

where $\boldsymbol{x} = (\tau_x,\ \tau_y,\ r)^T$ and $\boldsymbol{M} = [\boldsymbol{K} \mid \boldsymbol{c}]$, where the objective function can be written as

$$\|\boldsymbol{M}\boldsymbol{x} - \boldsymbol{h}\|_2^2 = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{x} + \boldsymbol{h}^T\boldsymbol{h}, \tag{19}$$

where $\boldsymbol{Q} = \boldsymbol{M}^T\boldsymbol{M}$ and $\boldsymbol{d}^T = -2\boldsymbol{h}^T\boldsymbol{M}$. In conclusion, one may consider minimizing $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{x}$, subject to $x_1^2 + x_2^2 - x_3 = 0$. Note that, the constraint can be written as $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} = 0$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}. \tag{20}$$

The Lagrangian is given by

$$\mathscr{L}(\boldsymbol{x};\lambda) = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{x} + \lambda(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}), \tag{21}$$

and solving $\nabla_{\boldsymbol{x}}\mathscr{L}(\boldsymbol{x};\lambda) = \boldsymbol{0}$ results in

$$\boldsymbol{x} = -\frac{1}{2}(\boldsymbol{Q} + \lambda\boldsymbol{A})^{-1}(\boldsymbol{d} + \lambda\boldsymbol{b})\,. \tag{22}$$

The constraint $\nabla_{\lambda}\mathscr{L}(\boldsymbol{x};\lambda) = 0$ yields a rational equation in λ, which can be turned into finding the roots of a fifth degree polynomial. This in turn can be translated into an eigenvalue problem, and solved robustly. Using this approach solving (16) takes $\sim$100 μs which is suitable for real-time applications. Furthermore, due to the precision of modern eigenvalue solvers, the error is usually negligable.

3.4.3 Solving for the Relative Orientation η

Given $\boldsymbol{\tau}$ from the previous step consider (3.4) and multiply the first homography with the matrix corresponding to the translation $\boldsymbol{\tau}$. This yields

$$\boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_{\psi\theta}^{T}\boldsymbol{H}\boldsymbol{R}_{\psi\theta}\boldsymbol{T}_{\boldsymbol{\tau}}^{-1} \sim \boldsymbol{R}_{\eta}^{T}\boldsymbol{R}_{\psi'\theta'}^{T}\boldsymbol{H}'\boldsymbol{R}_{\psi'\theta'}\boldsymbol{R}_{\eta}, \tag{23}$$

where η is the only unknown parameter. Define $\boldsymbol{W} = \boldsymbol{T}_{\boldsymbol{\tau}}\boldsymbol{R}_{\psi\theta}^{T}\boldsymbol{H}\boldsymbol{R}_{\psi\theta}\boldsymbol{T}_{\boldsymbol{\tau}}^{-1}$ and $\boldsymbol{W}' = \boldsymbol{R}_{\psi'\theta'}^{T}\boldsymbol{H}'\boldsymbol{R}_{\psi'\theta'}$ and note that $\boldsymbol{W}$ and $\boldsymbol{W}'$ share the same eigenvalues since they are similar and the corresponding eigenvectors are rotated η degrees.

Let us recall that the null space of a matrix is spanned by the right-singular vectors corresponding to zero—or due to noise, vanishing—singular values. Using the same approach as in (Wadenbäck and Heyden, 2014) we conclude that $\operatorname{nulldim}\boldsymbol{W}'^{T}\boldsymbol{W}' = 1$. Consequently, the eigenvectors spanning the null spaces, $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{W}^{T}\boldsymbol{W})$ and $\boldsymbol{x}' \in \mathcal{N}(\boldsymbol{W}'^{T}\boldsymbol{W}')$, can be obtained using SVD—this ensures that we work with real eigenvectors.

We will use the following theorem to recover η robustly, using all available pairs of homogaphies.

Theorem 1. *Let $\boldsymbol{Y}, \boldsymbol{Y}' \in \mathbb{R}^{3\times N}$ and non-zero. Furthermore, let $\boldsymbol{R}_{\eta} = \boldsymbol{R}_{z}(\eta)$ be a rotation matrix, corresponding to a rotation of angle η around the third axis. Then*

$$\min_{\substack{\eta\in(-\pi,\pi]\\ \lambda\neq 0}} \|\boldsymbol{Y}' - \lambda\boldsymbol{R}_{\eta}\boldsymbol{Y}\|_{F}, \tag{24}$$

is solved when

$$\eta_{opt} = \alpha + \begin{cases} 0, & \text{if } \boldsymbol{y}_3^{T}\boldsymbol{y}_3' > 0,\\ \pi, & \text{otherwise}, \end{cases} \tag{25}$$

where α may be expressed using the programming friendly atan2 *function,*

$$\alpha = \operatorname{atan2}\left(\boldsymbol{y}_1^{T}\boldsymbol{y}_2' - \boldsymbol{y}_2^{T}\boldsymbol{y}_1',\ \boldsymbol{y}_1^{T}\boldsymbol{y}_1' + \boldsymbol{y}_2^{T}\boldsymbol{y}_2',\right) . \tag{26}$$

Here $\boldsymbol{y}_i$ denotes the column vector of dimension N corresponding to the i:th row of $\boldsymbol{Y}$. The vectors $\boldsymbol{y}_i'$ are defined analogously. The angles are considered as equivalence classes, where $\eta \equiv \eta + 2\pi k$, $k \in \mathbb{Z}$, with the class representative being in the interval $(-\pi, \pi]$.

Proof. Using the relation between the Frobenius norm and the trace, the square of the objective function can be simplified

$$\begin{aligned}\|\boldsymbol{Y}' - \lambda\boldsymbol{R}_{\eta}\boldsymbol{Y}\|_{F}^{2} &= \operatorname{tr}\left[(\boldsymbol{Y}' - \lambda\boldsymbol{R}_{\eta}\boldsymbol{Y})(\boldsymbol{Y}' - \lambda\boldsymbol{R}_{\eta}\boldsymbol{Y})^{T}\right]\\ &= \operatorname{tr}\boldsymbol{Y}'\boldsymbol{Y}'^{T} - \lambda\operatorname{tr}\boldsymbol{Y}'\boldsymbol{Y}^{T}\boldsymbol{R}_{\eta}^{T}\\ &\quad - \lambda\operatorname{tr}\boldsymbol{R}_{\eta}\boldsymbol{Y}\boldsymbol{Y}'^{T} + \lambda^{2}\operatorname{tr}\boldsymbol{R}_{\eta}\boldsymbol{Y}\boldsymbol{Y}^{T}\boldsymbol{R}_{\eta}^{T} .\end{aligned} \tag{27}$$

Since the trace is invariant under cyclic permutations it follows that the last term is independent of η. Furthermore,

$$\operatorname{tr}\boldsymbol{Y}'\boldsymbol{Y}^{T}\boldsymbol{R}_{\eta}^{T} = \operatorname{tr}\left(\left(\boldsymbol{Y}'\boldsymbol{Y}^{T}\boldsymbol{R}_{\eta}^{T}\right)^{T}\right) = \operatorname{tr}\boldsymbol{R}_{\eta}\boldsymbol{Y}\boldsymbol{Y}'^{T} . \tag{28}$$

Combining these observations (24) is equivalent to solving

$$\min_{\substack{\eta\in[-\pi,\pi)\\ \lambda\neq 0}} \lambda^{2}\|\boldsymbol{Y}\|_{F}^{2} - 2\lambda\operatorname{tr}\boldsymbol{R}_{\eta}\boldsymbol{Y}\boldsymbol{Y}'^{T} . \tag{29}$$

The reader can easily verify that the optimum is reached when η is on the form (24). □

In conclusion, the angle η may be obtained using Theorem 1 where the i:th column of $\boldsymbol{Y}$ corresponds to the eigenvector spanning the null space of $\boldsymbol{W}_i^{T}\boldsymbol{W}_i$—the matrix $\boldsymbol{Y}'$ is defined analogously.

4 EXPERIMENTS

4.1 Synthetic Data

In order to validate the theory and evaluate the algorithm synthetic data was generated in form of sequences of images mimicking those taken by a mobile platform as described in Section 3.1. A high-resolution image of a planar scene, in this case a textured floor, was chosen to yield many key-points. Furthermore, different paths simulating the mobile platform was defined. In order to simulate the tilt the original image was transformed around a given point along the pre-defined path and then cropped, such that the center point in the cropped image coincide with this point. The parameters used in the transformation serve as ground truth, and the resolution used in each image is 400×400 pixels. The field of view of the simulated camera is normalized to 90 degrees, which affects the impact of the distortion of the images caused by the cameras being tilted.

4.2 Homography Estimation

The homography estimation was done by extracting SIFT keypoints (Lowe, 2004) from every frame, keeping the most prominent once as candidates for key-point matching. The remaining key-points are then matched between subsequent images only, using a brute-force matcher based on the K Nearest Neighbor algorithm. From the matched key-points a random subset is chosen iteratively in the RANSAC framework and from these a homography is estimated. The homography with the highest amount of

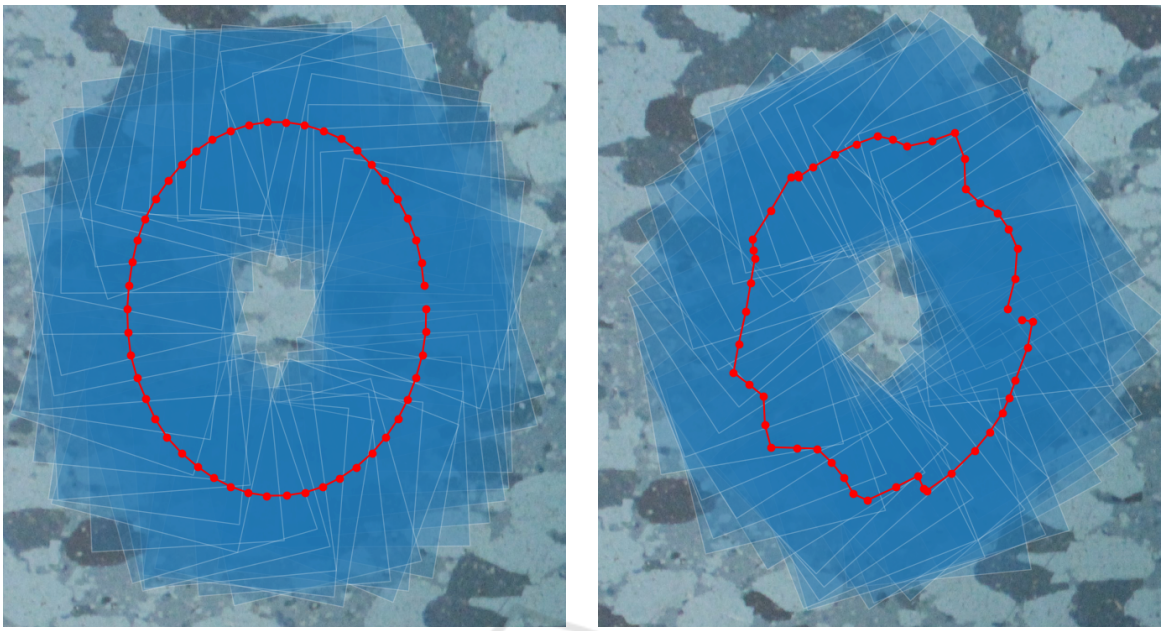

Figure 2: Path of the simulated mobile platform for the first camera (left) and the second camera (right). The red dots represent the absolute position of the camera and the blue squares are the extracted images. The impact of the tilt is illustrated by the frames not being square, but rather slanted. Note that the second camera path is not elliptic as the translational components are affected by the rotation of the mobile platform.

inliers is chosen, where the maximum allowed reprojection error for a point pair to be considered as an inlier is five pixels.

4.3 Parameter Recovery

4.3.1 Monocular Case

The parameters were recovered using the method proposed in (Wadenbäck and Heyden, 2014) for both trajectories, treated as two independent monocular systems, using five homographies to determine the tilt, rotation and translation in each step.

4.3.2 Recovering the Relative Translation

The optimal relative translation vector was obtained by solving (16) using the optimization scheme proposed in Section 3.4.2.

4.3.3 Recovering the Relative Orientation

Using the closed form solution presented in Theorem 1 the relative rotation around the z-axis was estimated for the five pairs of homographies used in the previous step. The computations involves finding the vectors spanning the null spaces in order to compute the matrices used in the closed form expression (24) for η which is computationally inexpensive.

4.4 Test Cases

4.4.1 Elliptic Path

This case simulates the mobile platform moving in an elliptic path while rotating between the images. The test case was choosen as it includes general motions which generates many different combinations of values for the nonfixed parameters. The sequence of images for both cameras used in this case is shown in Figure 2. The parameters used in this experiment are $\psi = 3.3^\circ$, $\theta = -1.2^\circ$, $\psi' = 5.1^\circ$, $\theta' = -4.6^\circ$, $\boldsymbol{\tau} = (100,\ 80)$ and $\eta = 30^\circ$.

The results from analyzing the two paths independently are shown in Figure 3 and the estimation of the relative pose is shown in Figure 4.

4.4.2 Rotation Around the First Camera Center

Estimating the tilt in a monocular system with only rotations and no translation is generally hard. A possible benefit of a binocular system is that the rigid body motion between the cameras results in a translational component in the second camera. The generated paths are shown in Figure 5. All fixed angles are set to 0° and the relative translation $\boldsymbol{\tau} = (800,\ 0)$. Furthermore, the mobile platform is moving with a constant rotation.

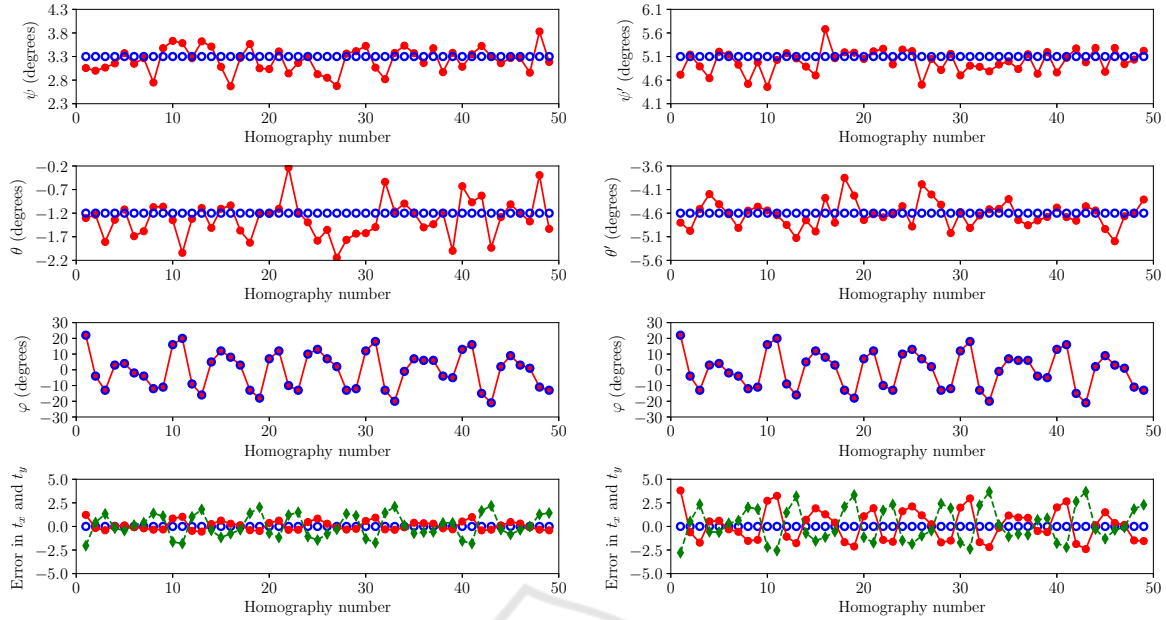

Figure 3: Estimated parameters for the first camera (left) and the second camera (right). In all plots the blue circles represent the ground truth. The red dots are the estimated parameters for the angles (first three subplots) and in the last subplot the red dots and the green diamond are the error in t_x and t_y respectively. The error in translation is measured in pixels. The estimates have been calculated using five homographies at each frame.

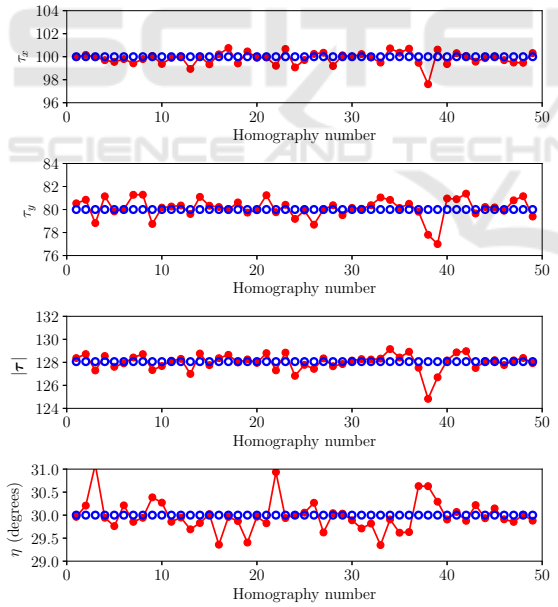

Figure 4: Estimated value of $\boldsymbol{\tau} = (\tau_x, \tau_y)$ and η using five pairs of homographies at each frame. The magnitude of the translational component $|\boldsymbol{\tau}|$ is also shown. The red dots are the estimated parameter and the blue circles represent the ground truth.

Since (3.4.1) degenerates, as $k_1^{(i)} = k_2^{(i)} = 0$ for all homographies, one cannot expect to recover the components of $\boldsymbol{\tau}$ but it is possible to recover $|\boldsymbol{\tau}|$, as shown in Figure 6.

4.4.3 Mean Error vs. Number of Homographies

The relation between the accuracy and the amount of homographies used to estimate the relative pose is shown in Figure 7. The same setup as in Section 4.4.1 was used but the amount of homographies varied. From the figures one can see that it is not a significant improvement in the parameter estimation of the relative pose after approximately twenty homographies. In practice this means that the calibration could be done initially, and then be used to track the position of the mobile platform, without re-computation of the fixed parameters.

5 CONCLUSION

This paper has extended the work of (Wadenbäck and Heyden, 2014) to binocular vision. A method has been deviced to robustly estimate the relative translation and orientation of the two cameras using several pairs of homographies, by reusing the computations from the cameras treated as two monocular systems. The translational component is recovered by solving a non-convex problem, which can be turned into an eigenvalue problem. The proposed optimization scheme is robust and suitable for real-time applications. Furthermore, when solving for the relative

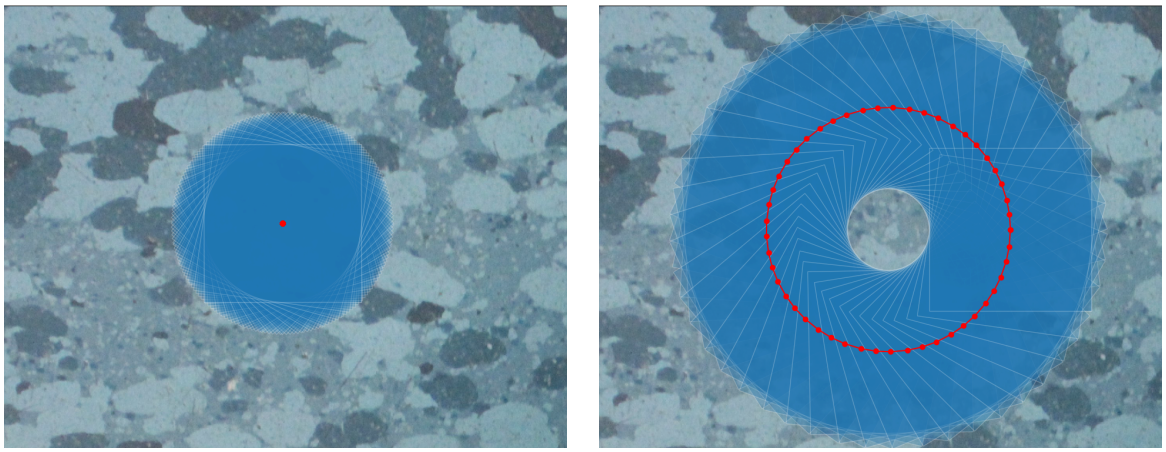

Figure 5: Path of the simulated mobile platform for the first camera (left) in the second test case, and for the second camera (right). The red dots represent the absolute position of the camera and the blue squares are the extracted images. Due to the rigid body motion connecting the cameras the second camera rotates around the first camera center causing a non-zero translational component.

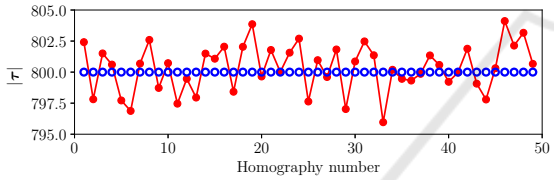

Figure 6: Estimated value of $|\boldsymbol{\tau}|$ using five pairs of homographies at each frame. When considering only rotations for the first camera, the components of $\boldsymbol{\tau}$ cannot be obtained by the proposed method. The red dots are the estimated parameter for the magnitude and the blue circles represent the ground truth.

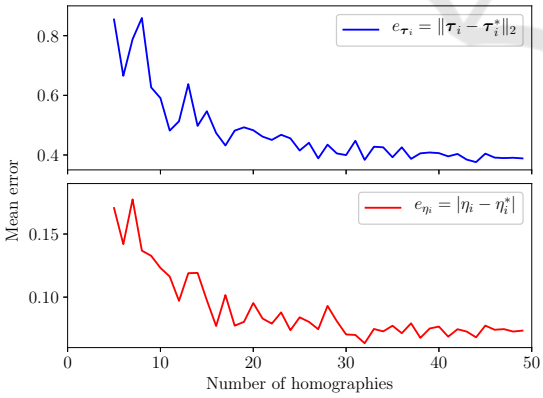

Figure 7: Mean error for the relative translation and rotation as a function of the number of pairs of homographies used in the optimization step. The error for the translational component is measured in the Euclidean norm. The mean error is computed from 49 pairs of homographies estimated from the sequence described in Section 4.4.1.

rotation, the closed form solution presented in Theorem 1 is computationally inexpensive. Experimental results from synthetic data have demonstrated that the method has an acceptable accuracy for most problems, and highlighted problems where the method fails to recover both of the translational components; it is also shown that in this case the magnitude can be recovered accurately.

ACKNOWLEDGEMENTS

This work has been funded by the Swedish Research Council through grant no. 2015-05639 'Visual SLAM based on Planar Homographies'.

REFERENCES

Davison, A. J., Reid, I. D., Molton, N. D., and Stasse, O. (2007). Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067.

Engel, J., Schöps, T., and Cremers, D. (2014). *LSD-SLAM: Large-Scale Direct Monocular SLAM*, pages 834–849. Springer International Publishing, Cham.

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*, ECCV '92, pages 563–578, London, UK. Springer-Verlag.

Hajjdiab, H. and Laganière, R. (2004). Vision-based multi-robot simultaneous localization and mapping. In *Computer and Robot Vision, 2004. Proceedings. First Canadian Conference on*, pages 155–162. IEEE.

Harris, C. and Pike, J. (1988). 3d positional integration from image sequences. *Image and Vision Computing*, 6(2):87 – 90.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition.

Hartley, R. I. (1992). *Estimation of relative camera positions for uncalibrated cameras*, pages 579–587. Springer Berlin Heidelberg, Berlin, Heidelberg.

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.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Mur-Artal, R. and Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262.

Nyman, P., Heyden, A., and Åström, K. (2010). Multi-camera platform calibration using multi-linear constraints. *2010 20th International Conference on Pattern Recognition, Pattern Recognition (ICPR), 2010 20th International Conference on*, page 53.

Wadenbäck, M. and Heyden, A. (2013). Planar motion and hand-eye calibration using inter-image homographies from a planar scene. In *8th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2013), Proceedings of*, pages 164–168.

Wadenbäck, M. and Heyden, A. (2014). Ego-motion recovery and robust tilt estimation for planar motion using several homographies. In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP 2014)*, pages 635–639.

Wadenbäck, M., Karlsson, M., Heyden, A., Robertsson, A., and Johansson, R. (2017). Visual odometry from two point correspondences and initial automatic tilt calibration. In *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 340–346.