Ego-Motion Recovery and Robust Tilt Estimation for Planar Motion Using Several Homographies

Wadenbäck, Mårten; Heyden, Anders

Published in:
9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2014), Proceedings of

DOI:
10.5220/0004744706350639

Published: 2014-01-01

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.
Ego-Motion Recovery and Robust Tilt Estimation for Planar Motion
Using Several Homographies

Mårten Wadenbäck and Anders Heyden
Centre for Mathematical Sciences
Lund University, Sweden
{marten, heyden}@maths.lth.se

Keywords: SLAM: Homography: Robotic Navigation: Planar Motion: Tilt Estimation

Abstract: In this paper we suggest an improvement to a recent algorithm for estimating the pose and ego-motion of a camera which is constrained to planar motion at a constant height above the floor, with a constant tilt. Such motion is common in robotics applications where a camera is mounted onto a mobile platform and directed towards the floor. Due to the planar nature of the scene, images taken with such a camera will be related by a planar homography, which may be used to extract the ego-motion and camera pose. Earlier algorithms for this particular kind of motion were not concerned with determining the tilt of the camera, focusing instead on recovering only the motion. Estimating the tilt is a necessary step in order to create a rectified map for a SLAM system. Our contribution extends the aforementioned recent method, and we demonstrate that our enhanced algorithm gives more accurate estimates of the motion parameters.

1 INTRODUCTION

One of the long-standing aims in robotics research is the development of algorithms for autonomous navigation. A popular class of such algorithms are the ones concerned with so called Simultaneous Localisation and Mapping (SLAM), in which a mobile platform, equipped with an array of suitable sensors (laser scanners, cameras, odometers, sonar, . . .), explores and maps the surrounding environment while keeping track of its own location with respect to the map. The map created in the process should mark notable objects and landmarks in a way which allows for reliable re-identification. The type of map that can be created is highly dependent on the kinds of sensors employed and on the environment being mapped, and can range from sparsely placed points to dense and detailed textured 3D models.

Using cameras to build the map is becoming increasingly attractive, as they are cheap compared to many of the other sensors, and since the traditional obstacle of high computational cost becomes less inhibiting with time as computational power increases. Another advantage of using cameras is that it allows for utilisation of the increasingly sophisticated methods and great experience that the computer vision community has produced during the past few decades. Indeed, scene reconstruction from images is a classical and continually studied problem in computer vision, and various methods have been proposed for both general cases and specialised applications.

Many of the successful general reconstruction techniques are based on epipolar geometry, and in particular the fundamental matrix, which was introduced independently in (Faugeras, 1992) and (Hartley, 1992). Such methods make the implicit assumption that the data are not positioned in one of the so called critical configurations, and in many practical cases such degeneracies are indeed very unlikely to occur. However, one of the less unlikely critical configurations occurs when the data points are coplanar — indeed, the application to navigation that we describe in this paper requires the data points to lie in a plane. Since planar structures are very common in man-made environments, this is an area in which specialised algorithms which can avoid degeneracy can have great advantages.

While invariant local features, for instance SIFT (Lowe, 2004) and other similar features, are standard in Structure from Motion (SfM), their use in camera based SLAM has been less prevalent. One of the main reasons for this is probably, as observed in (Davison et al., 2007), that though such features allow for accurate and robust re-identification, their computational cost has traditionally been obstructive for real time applications. Although this is essentially still a valid
2 RELATED WORK

A robot mapping application not only requires an incremental reconstruction, as data becomes available sequentially, but in contrast to Structure from Motion approaches such as the popular Bundler system described in (Snavely et al., 2008), the order in which views are added is more or less predetermined. Though the views are added to the reconstruction in a fixed order, some SLAM approaches allow the robot path itself to be planned so that the images can be taken from locations which make the reconstruction better (Haner and Heyden, 2011), but we will in this paper consider the path to already be decided. Some very early work which respects the restriction on the order of views is (Harris and Pike, 1988), in which a Kalman filter was used to estimate camera position based on inter-image point correspondences throughout a short image sequence. Probabilistic viewpoints based on extended Kalman filters (EKF) remain popular in later systems such as the vSLAM system (Karlsson et al., 2005) and the MonoSLAM system (Davison et al., 2007).

The systems mentioned above allow general 3D camera motion, but this is not always necessary or even desired. A camera that has been mounted onto a mobile platform will typically perform two-dimensional motion since it remains at a fixed height above the ground, and with this knowledge one can eliminate some of the uncertainty which 3D motion allows. Our work continues in the spirit of (Liang and Pears, 2002) and (Hajjdiab and Laganière, 2004) and others, in that we intend to navigate using images of the floor. Since the scene is planar, the images will be related by planar homographies.

Liang and Pears find the rotation angle $\varphi$ by noting that the eigenvalues of the inter-image homography are (up to scale) 1 and $e^{\pm i \varphi}$, and they derive an expression for the translation from the eigenvectors. One drawback of this method is that it does not determine the tilt. Determining the tilt allows a rectified map to be created, and is therefore highly desirable.

A more recent method described in (Wadenbäck and Heyden, 2013) starts with estimating the tilt $R_{\varphi \theta}$, and then performs a QR decomposition of $R_{\varphi \theta}^T H R_{\varphi \theta}$ to determine $\varphi$ and the translation $(t_x, t_y)$.

We show in this paper how to extend their estimation algorithm to use more than one homography for estimating the tilt. This improves robustness to noise and erroneous measurements.

3 PROBLEM GEOMETRY

We shall consider the navigation of a mobile platform equipped with a single camera that has been mounted rigidly onto the platform and directed towards the floor. This setup means that the camera will move at constant height in a plane parallel to the floor, and have a constant angle to the plane normal (tilt). Figure 1 shows a typical image from one of our datasets, taken under the conditions described here. Figure 2 shows an illustration of the geometrical situation. We will further assume zero skew and square pixels, and that the camera parameters remain constant during the motion (no zooming or refocusing). It will be convenient to work with a global coordinate system in which the camera moves in the plane $z = 0$ and the ground plane is represented by the plane $z = 1$.

As already noted, two images will be related by a planar homography $H$. We model the camera motion by a translation $t = (t_x, t_y)$ and a rotation $R_{\varphi}$ an angle $\varphi$ about the normal of the floor plane (the $z$-axis). Using homogeneous coordinates in the plane, the motion of the camera is represented by the transformation $R_{\varphi} T$. 

Figure 1: A typical image taken by a camera under the conditions described in this paper is shown in Figure 1(a). A rectified version, as if seen straight from above, can be seen in Figure 1(b). In order to rectify such images, it is necessary to be able to estimate the camera tilt.

Figure 2: The camera moves freely in the plane $z = 0$, and can rotate about the normal of the plane, but the angle to the plane normal (tilt) is held constant.
Planar Motion and Hand-Eye Calibration Using Inter-Image Homographies from a Planar Scene

with
t \begin{bmatrix}
1 & 0 & -tx \\
0 & 1 & -ty \\
0 & 0 & 1
\end{bmatrix}.

If the camera is tilted, the camera coordinate system
and the world coordinate system are related by a ro-
tation \( R_{\psi \theta} \). This means that the inter-image
homographies will be of the form
\[ H = \lambda R_{\psi \theta} R_T T^T. \]

where \( \lambda \neq 0 \) is an unknown scale parameter.

Estimating the homographies from the images can
be done using point correspondences and a robust
method such as RANSAC. This is not the focus of
our work, and we will henceforth assume that well-
estimated homographies are available, without con-
cerning ourselves with how they were obtained.

4 PARAMETER RECOVERY

Suppose we have a number of homographies of the
form in (2), that is,
\[ H_j = \lambda_j R_{\psi \theta} R_T T^T, \quad j = 1, \ldots, N, \]

and want to recover the motion parameters. As ob-
erved in (Wadenbäck and Heyden, 2013), the products
\[ M_j = \begin{bmatrix}
m_{11} & m_{12} & m_{13} \\
m_{12} & m_{22} & m_{23} \\
m_{13} & m_{23} & m_{33}
\end{bmatrix} = H_j^T H_j \]

are all independent of \( \varphi \).

An iterative scheme is also presented which altern-
ates between solving for \( \psi \) and \( \theta \), keeping the other
one fixed. Their paper demonstrates that this can be
accomplished by finding the null space of the matrix
\[ \Psi_j = \begin{bmatrix}
m_{11} - \tilde{m}_{11} & -2\tilde{m}_{13} & \tilde{m}_{12} - \tilde{m}_{13} \\
0 & \tilde{m}_{13} & \tilde{m}_{12} \\
0 & \tilde{m}_{12} & \tilde{m}_{13}
\end{bmatrix} \]

in the \( \psi \) case (where \( \tilde{M} = R_{\psi}^T M R_{\psi} \)), and of the matrix
\[ \Theta_j = \begin{bmatrix}
m_{11} - \tilde{m}_{11} & -2\tilde{m}_{13} & \tilde{m}_{12} - \tilde{m}_{13} \\
0 & \tilde{m}_{13} & \tilde{m}_{12} \\
0 & \tilde{m}_{12} & \tilde{m}_{13}
\end{bmatrix} \]

in the \( \theta \) case (with \( \tilde{M} = R_{\theta}^T M R_{\theta} \)). It can clearly be
seen that these matrices have at least rank two, except
in the case where the bottom two rows are identically
zero, so a one dimensional null space is expected. Due
to measurement errors the null space will in practice
be trivial, and a one dimensional approximation is
computed as the singular vector \( \nu = (v_1, v_2, v_3) \)
corresponding to the smallest singular value. In the \( \psi \)
case, any vector \( \nu \) in the null space should be a scalar
multiple of \( (c_\psi, c_\psi s_\theta, s_\psi) \), which gives
\[ \psi = \frac{1}{2} \arcsin \frac{2v_2}{v_1 + v_3}. \]

while in the same way, the the solution in the \( \theta \) case is
a scalar multiple of \( (\frac{c_\theta}{\lambda}, c_\theta s_\psi, s_\theta) \), and
\[ \theta = \frac{1}{2} \arcsin \frac{2v_2}{v_1 + v_3}. \]

This paper presents the insight that if the tilt \( R_{\psi \theta} \)
remains constant, then the matrices \( \Psi_j \) all should have
the same null space. Instead of considering each \( \Psi_j \)
separately, we can therefore solve
\[ \Psi v = \begin{bmatrix}
\Psi_1 \\
\vdots \\
\Psi_N
\end{bmatrix} \begin{bmatrix}
c_\psi \\
c_\psi s_\psi \\
\end{bmatrix} = 0. \]

In the same way, we may combine the equations for \( \theta \)
to
\[ \Theta v = \begin{bmatrix}
\Theta_1 \\
\vdots \\
\Theta_N
\end{bmatrix} \begin{bmatrix}
c_\theta \\
c_\theta s_\psi \\
\end{bmatrix} = 0. \]

The angles are computed from the solution \( v \) in the
same way as above using (7) and (8).

5 EXPERIMENTS

For the purpose of comparing the unmodified al-
gorithm outlined in (Wadenbäck and Heyden, 2013)
with our enhanced version, we have randomly gen-
erated a large number of homographies of the form
in (3). Gaussian noise with standard deviation of 0.5°
was added to each of the angles, intended to simulate
measurement noise. Figure 3 shows the estimation
results obtained using only one homography at a time,
and Figure 4 shows the results using our proposed
method with five homographies used at each step. The
same number of iterations were used for the two meth-
ods. Note that the scale on the axes is the same in
both figures, for the benefit of easier comparison. It
is readily seen that the proposed method drastically
decreases the number of cases where the algorithm
fails to converge.

It should be pointed out that while the results from
the unmodified method can be much improved using
filtering techniques, the same is true for our enhanced
method.

The unmodified algorithm was reported to have
difficulties when the translation was close to a pure
\( x \)-translation or a pure \( y \)-translation. In the case of
an \( x \)-translation, \( \theta \) would be poorly estimated, and

637
Planar Motion and Hand-Eye Calibration Using Inter-Image Homographies from a Planar Scene

6 CONCLUSION

In this paper we have extended the estimation method in (Wadenbäck and Heyden, 2013) to use more than one homography to estimate the tilt. This enhancement produces a robust and more accurate estimate, which demonstrably allows the other motion parameters to be recovered with higher precision. The problems with ill-conditioned motion patterns that were reported in the original algorithm have also been remedied by using more than one homography at a time.

Figure 3: Tilt and motion parameters estimated from one homography at a time using the unmodified method. The starred parameters are the estimates.

Figure 4: Tilt and motion parameters estimated from five homographies at a time using our proposed method. The starred parameters are the estimates.

Figure 5: The x- and y-components of the translation that was used to generate the homographies, normalized by the length of the translation in that step. Some of the translations used are apparently close to pure x-translations or pure y-translations, which were reported to be problematic for the original algorithm.

ACKNOWLEDGEMENTS

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


