A homography formulation to the 3pt plus a common direction relative pose problem

Olivier Saurer, Pascal Vasseur, Cedric Demonceaux, Friedrich Fraundorfer

A homography formulation to the 3pt plus a common direction relative pose problem

Olivier Saurer[1], Pascal Vasseur[2], Cedric Demonceaux[3], and Friedrich Fraundorfer[4]

[1]ETH Zürich, Switzerland
[2]LITIS - Université de Rouen, France
[3]Le2i, UMR CNRS 6306 Université de Bourgogne, France
[4]Remote Sensing Technology, Technische Universität München, Germany

Abstract. In this paper we present an alternative formulation for the minimal solution to the 3pt plus a common direction relative pose problem. Instead of the commonly used epipolar constraint we use the homography constraint to derive a novel formulation for the 3pt problem. This formulation allows the computation of the normal vector of the plane defined by the three input points without any additional computation in addition to the standard motion parameters of the camera. We show the working of the method on synthetic and real data sets and compare it to the standard 3pt method and the 5pt method for relative pose estimation. In addition we analyze the degenerate conditions for the proposed method.

1   Introduction

Reducing the number of required points to be matched or to be tracked in order to estimate the egomotion of a visual system can be very interesting in terms of computation time efficiency and of robustness improvement. In the case of an uncalibrated camera, eight points are at least necessary for estimating the fundamental matrix [1] while only five are sufficient in the calibrated case for the essential matrix [2]. It appears clearly that this point number reduction is only possible if some hypotheses or supplementary data are available. For example, some recent works proposed to use only one point to perform a structure from motion method [3] by supposing a non-holonomic motion or to estimate the metric velocity of a single camera [4] in combining with accelerometer and attitude measurements.

In our method, we propose to use a known common direction between two images. Thus, this knowledge of a common direction reduces the number of rotations from three to one. Obtaining a common direction can be performed either by the use of an IMU (Inertial Measurement Unit) or by some information extracted from the images such as vanishing points or horizon. IMUs are nowadays used in many devices such as smart phones or robotic platforms. The coupling with a camera is then very easy and can then be used for different computer vision tasks [5–9]. However, even if a complete rotation is available from an IMU,

it appears nevertheless that the heading is generally less accurate than the two other angles [10]. Let us note, in the case of a pure vision approach (without external sensor), this common direction can be obtained by one vanishing point in man made environment [11] or by the detection of the horizon line [12] in the case of natural scene.

Recently, it has been shown that only three points are needed to compute the pose between two cameras if we know a common direction [10, 13, 14]. This method derives the epipolar constraint to compute the essential matrix with only 3 degrees of freedom. In this paper we propose to use the homography constraint between two views. We show, contrary to the general case where 4 points are needed, that we can compute the pose with only 3 matching points. Let us note, as we need only 3 points in the scene, our method works even if these 3D points do not belong physically to the same plane. Thus the derived formulation using the homography constraint can be used in the same settings as the standard 3pt method using the epipolar constraint.

Moreover, if we assume the existence of a dominant plane in the scene, our method can be included in a RANSAC process in order to estimate the normal of the plane and the pose of the cameras in presence of noise. Compared to the classical algorithm of homography estimation which needs 4 points, we use only 3 points and thus allows to decrease the number of iterations in the RANSAC. Indeed, as presented in Figure 1, for a probability of success of 0.99 and a rate of outliers equal to 0.5, the number of trials is divided by two for a robust implementation with RANSAC compared to the 4 points algorithm. Using the co-planarity constraint in order to simplify the SFM problem is not new [15][16]. Moreover, if we know roll and pitch of the camera, Saurer et al. [17] prove that we need only 2 points to estimate the full pose of the camera if we also know the normal of the plane. In this paper, we show that only 3 points are needed if roll and pitch angles are known and if the normal is unknown.

2   Related Works

When the camera is not calibrated, at least 8 points are needed to recover the motion [1]. It's now well known that, if this camera is calibrated, only 5 feature points in the scene can be sufficient. Reducing this number of points can be very interesting in order to reduce the computation time and to increase the robustness. To do this, we have to add some hypotheses about the motion of the camera, on the extracted feature points or use supplementary sensors.

For example, if all the 3D points belong on a plane, we need a minimum of 4 points to estimate the motion of the camera between two-views [1]. On the other hand, if the camera is embedded on a mobile robot which moves on a planar surface we need 2 points to recover the motion [18] and if moreover the mobile robot has non-holonomic constraints only one point is necessary [3]. In the same way, if the camera moves in a plane perpendicular to the gravity, Troiani et al. [19] have also shown that one point is enough.
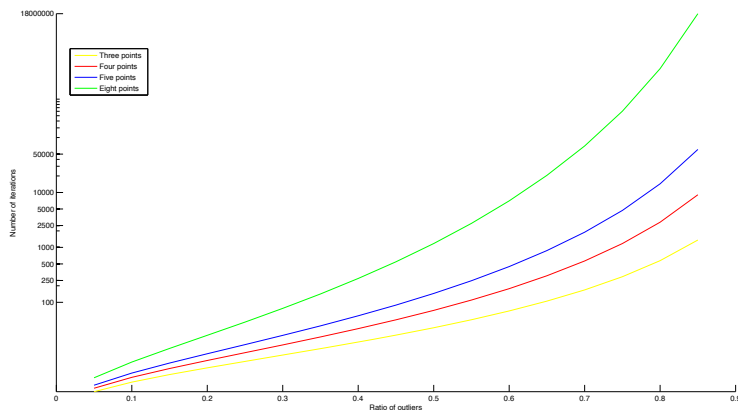
Fig. 1. Comparison of the RANSAC iteration numbers for a 99% probability of success

The number of points needed to estimate the ego motion can also be reduced if we have some information about the relative rotation between two poses. This information can be given by vanishing points extraction in the images [20] or in taking into account extra information given by an other sensor. Thus, Li et al. [21] show if we use an IMU mounted to the sensor only 4 points are sufficient to estimate the relative motion even if we don't know the extrinsic calibration between the IMU and the camera.

Similarly, some different algorithms have been recently proposed in order to estimate the relative pose between two cameras by knowing a common direction. We can show that if we know the roll and pitch angles of the camera at each time, we need only three points to recover the yaw angle and the translation of the camera motion [10, 13, 14]. In these approaches, only the formulation of the problem is different and consequently the way to solve it. All these works start with a simplified essential matrix in order to derive a polynomial equation system. For example, in [10], their parametrization leads to 12 solutions by using the Macaulay matrix method. The good solution has then to be found among this set of solutions. The approach presented in [13] permits to obtain a 4th-order polynomial equation and consequently leads to a more efficient solution. In contrast to the method of [13] the proposed algorithm directly solves for rotation and translation parameters. In [13] the essential matrix is estimated first and to solve for rotation and translation parameters the essential matrix has to be decomposed by an additional step typically involving SVD. In [14], the authors propose a closed-form solution to the 4th-order polynomial equation that allows a faster computation.

If we want to reduce again the number of points, stronger hypotheses have to be added. When the full rotation between the two views are known, we have only

2 degrees of freedom to estimate corresponding to the translation up-to-scale. In this strong hypothesis, 2 points can be use to solve the problem [22]. In this case, the authors compute the translation vector by epipolar geometry with a rotation equal to identity. Thus, these approaches allow to reduce the number of points but the knowledge of the complete rotation between two views makes these methods really sensitive to IMU accuracy. More recently, Martinelli [23] proposes a closed-form solution for structure from motion knowing the gravity axis of the camera in a multiple views scheme. He shows that we need at least 3 features points belong on a same plane and 3 consecutive views to estimate the motion of the camera. In the same way, the plane constraint has been used for reducing the complexity of the bundle adjustment (BA) in a visual simultaneous localization and mapping (SLAM) embedded on a micro-aerial vehicle (MAV) [24].

In [17] a strong assumption about the environment is made to reduce the number of necessary points for motion estimation. Within the Manhattan world assumption any scene plane is either parallel or vertical to a gravity aligned camera plane. Thus only 2 points are necessary to estimate relative camera pose using a homography formulation. This paper also describes the possibility of propagating relative scale by utilizing the recovered distance to the scene plane used for motion estimation as an alternative to the standard way of propagating scale through 3D reconstruction of feature points. As our proposed method also recovers the distance to scene planes this method for scale propagation can also be used when using our proposed method for motion estimation, in particular when a dominant scene plane is present (e.g. ground plane).

3   Relative pose with the knowledge of a common direction

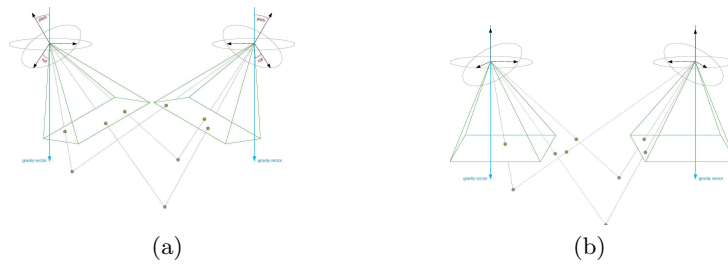(a)   (b)

Fig. 2. (a) 3D scene with two camera frames in general position. (b) Cameras after de-rotation such that both camera frames are aligned with the vertical direction (gravity direction).

Knowing a common direction in images will simplify the estimation of camera pose and camera motion, which are fundamental methods in 3D computer vision. It is then possible to align every camera coordinate system using the known common direction, e.g. aligning them to the vertical direction such that the z-axis of the camera is parallel to the vertical direction and the x-y-plane of the camera is orthogonal to the vertical direction (illustrated in Fig. 2) or any other arbitrary alignment. This alignment can just be done as a coordinate transform for motion estimation algorithms, but also be implemented as image warping such that feature extraction method benefit from it. Relative motion between two such aligned cameras reduces to a 3-DOF motion, which consists of 1 remaining rotation and a 2-DOF translation vector. A general relative pose between two images is represented using the following epipolar constraint

$$p_j^T E p_i = 0, \tag{1}$$

where E is the essential matrix representing a 5-DOF relative pose. Aligning transformations R_i, R_j for the point measurements p_i, p_j can be computed from the known common direction and lead to the aligned measurements q_i, q_j.

$$q_i = R_i p_i \tag{2}$$
$$q_j = R_j p_j \tag{3}$$
$$\tag{4}$$

This leads to the simplified epipolar constraint

$$q_j^T \hat{E} q_i = 0, \tag{5}$$

where the essential matrix $\hat{E}$ represents the simplified 3-DOF relative pose between the aligned cameras. The general relative pose R, t can then be computed by reversing the alignment with $R = R_j^T \hat{R} R_i$ and $t = R_j^T \hat{t}$.

For the case of points on a plane the relative pose can also be written by using the homography constraint:

$$q_j = H_{ij} q_i \tag{6}$$

The homography is composed of

$$\mathbf{H} = \mathbf{R} - \frac{1}{d}\mathbf{t}\mathbf{n}^T, \tag{7}$$

where $\mathbf{R} = \mathbf{R_z}\mathbf{R_y}\mathbf{R_x}$ is a rotation matrix representing the relative camera rotations around the x, y, and z-axis, $\mathbf{t} = [t_x, t_y, t_z]^T$ represents the relative motion, $\mathbf{n} = [n_x, n_y, n_z]^T$ is the plane normal and d is the distance from the first camera center to the plane.

For the 3-DOF relative pose between two aligned cameras the homograpy simplifies as well and writes as

$$H_{ij} = R_z - \frac{1}{d}\mathbf{t}\mathbf{n}^T, \tag{8}$$

where R_z is the remaining rotation around a single axis between the aligned cameras.

3.1 Previous approaches - 3pt relative pose using the epipolar constraint

All the previous approaches that solve the 3pt+1 problem [10, 13, 14] have in common that they utilize the epipolar constraint to set up the equations for relative pose estimation.

$$q_j^T E_{ij} q_i = 0 \tag{9}$$

where E_{ij} is the essential matrix composed of $E_{ij} = [t_{ij}]_x R_{ij}$.

However, for this case, when only 3 point correspondences are necessary, one can consider to use the homography constraint alternatively. 3 points by definition are always co-planar and form a plane. For any such 3 points the homography constraint will hold. It is therefore possible to use the homography constraint instead of the epipolar constraint to solve the 3pt+1 problem. In the next section we show the derivation of the solution to this novel idea.

3.2 3pt relative pose using the homography constraint

The general homography relation for points belonging to a 3D plane and projected in two different views is defined as follows :

$$q_j = H_{ij} q_i \tag{10}$$

with $q_i = \begin{bmatrix} x_i\ y_i\ w_i \end{bmatrix}^T$ and $q_j = \begin{bmatrix} x_j\ y_j\ w_j \end{bmatrix}^T$ the projective coordinates of the points between the views i and j. H_{ij} is given by :

$$H_{ij} = R_{ij} - \frac{1}{d} \mathbf{t} \mathbf{n}^T \tag{11}$$

where R_{ij} and $\mathbf{t}$ are respectively the rotation and the translation between views i and j and where d is the distance between the camera i and the 3D plane described by the normal $\mathbf{n}$.

In our case, we assume that the camera intrinsic parameters are known and that the points q_i and q_j are normalized. We also consider that the attitude of the cameras for both views are known and that these attitude measurements have been used to align the camera coordinate system with the vertical (gravity) direction. In this way, only the yaw angle θ between the views remains unknown.

$$\mathbf{H} = \mathbf{R_z} - \frac{1}{d} [t_x, t_y, t_z]^T [n_x, n_y, n_z] \tag{12}$$

Without loss of generality Eq. 12 can be written as

$$\mathbf{H} = \mathbf{R_z} - \hat{d} [\hat{t_x}, \hat{t_y}, 1]^T [n_x, n_y, n_z]. \tag{13}$$

In Eq. 13 the scale of the parameter t_z has been included in $\hat{d}$ and thus the z-component of the translation can be set to 1. This reformulation of the

parameters proves useful for the used Groebner basis technique for solving the equation system. For the case of very small z-motions this choice could lead to numerical instability but in such cases any other motion direction could be set to 1 instead. However, in practice we did not notice such an instability.

The homography matrix then consists of the following entries:

$$\mathbf{H} = \begin{bmatrix} \cos(\theta) - \hat{d}n_x\hat{t_x} & -\sin(\theta) - \hat{d}n_y\hat{t_x} & -\hat{d}n_z\hat{t_x} \\ \sin(\theta) - \hat{d}n_x\hat{t_y} & \cos(\theta) - \hat{d}n_y\hat{t_y} & -\hat{d}n_z\hat{t_y} \\ -\hat{d}n_x & -\hat{d}n_y & 1 - \hat{d}n_z \end{bmatrix} \tag{14}$$

The unknowns that we are seeking for are the motion parameters $\cos(\theta)$, $\sin(\theta)$, $\hat{d}$, $\hat{t_x}$, $\hat{t_y}$ as well as the plane normals n_x, n_y, n_z of the plane defined by the 3 point correspondences. This is a significant difference to the standard 3pt which does not solve for the plane normals. The proposed 3pt also solves for the plane normals for free.

By utilizing the homography constraints of point correspondences, an equation system to solve for these unknowns can be set up.

$$q_j \times H_{ij}q_i = 0 \tag{15}$$

In this $\times$ denotes the cross product and by expanding the relation we obtain

$$\begin{bmatrix} w_iy_j - \cos(\theta)w_jy_i - \sin(\theta)w_jx_i - \hat{d}n_zw_iy_j - \hat{d}n_xx_iy_j - \hat{d}n_yy_iy_j + \hat{d}n_z\hat{t_y}w_iw_j + \hat{d}n_x\hat{t_y}w_jx_i + \hat{d}n_y\hat{t_y}w_jy_i \\ \cos(\theta)w_jx_i - w_ix_j - \sin(\theta)w_jy_i + \hat{d}n_zw_ix_j + \hat{d}n_xx_ix_j + \hat{d}n_yx_jy_i - \hat{d}n_z\hat{t_x}w_iw_j - \hat{d}n_x\hat{t_x}w_jx_i - \hat{d}n_y\hat{t_x}w_jy_i \end{bmatrix}$$
$$= \mathbf{0} \tag{16}$$

The third equation being a linear combination of the two others is being omitted.

Each point correspondence gives 2 linearly independent equations and there are two additional quadratic constraints in the unknowns that can be utilized.

$$\cos^2\theta + \sin^2\theta - 1 = 0 \tag{17}$$

$$n_x^2 + n_y^2 + n_z^2 - 1 = 0 \tag{18}$$

$$\tag{19}$$

The total number of unknowns is 8 and the two quadratic constraints together with the equations from 3 point correspondences will give a total of 8 polynomial equations in the unknowns. Finding a solution to such a polynomial equation system can be difficult. Most monomials are mixed terms of the unknowns. One way of solving such an equation system in closed form is by using the Groebner basis technique [25]. By computing the Groebner basis a univariate polynomial in a single variable can be found which allows to find the value of this variable by root solving. The remaining variables can then be computed by back-substitution. To solve our problem we utilize the automatic Groebner basis solver by Kukelova et al. [26]. This solver automatically computes Matlab-code to solve for the unknowns of the given polynomial equation system.

The analysis of the Groeber basis solutions shows, that the final univariate polynomial has degree 8, which means that there are up to 8 real solutions to our problem.

3.3   Degenerate conditions

In this section we discuss the degenerate conditions for the proposed 3pt homography method. The degenerate conditions for the standard 3pt method have been discussed in detail in [13, 14, 10]. In these papers it is pointed out that a collinear configuration of 3D points is in general not a degenerate condition for the 3pt method, while it is one for the 5pt method. Degenerate conditions for the standard 3pt algorithm however are collinear points that are parallel to the translation direction and points that are coplanar to the translation vector. We investigated if these scenarios also pose degenerate conditions for the proposed 3pt homography method by conducting experiments with synthetic data. Degenerate cases could be identified by a rank loss of the action matrix within the Groebner basis solver. These tests showed that the proposed method shares the degenerate conditions of the standard 3pt method but in addition also has a degenerate condition for the case of collinear points. This is understandable as the 3pt homography method also solves for the plane normal which then has an undefined degree of freedom around the axis of the collinear points. The results of the comparison are summarized in Table 1.

Table 1. Comparison of the degenerate conditions for the standard 3pt method and the proposed 3pt homography method.

	3pt	3pt-homography
collinear points	no	yes
collinear points parallel to translation direction	yes	yes
points coplanar to translation vector	yes	yes

4   Experiments

The proposed method is evaluated on both synthetic and real datasets.

4.1   Synthetic evaluation

The synthetic evaluation is conducted under the following setup. Focal length of the camera is 1000 pixel with a field of view of 45 degrees. The first camera is set to the origin of the coordinate frame and kept fixed. The base-line between the first and second camera is set to 0.2 i.e., 20% of the average scene depth. The scene consists of 200 randomly sampled points. The algorithm is evaluated under varying image noise and increasing IMU noise (roll and pitch) on two configurations, sideways and forward motion of the camera. Each configuration is evaluated on 100 randomly sampled cameras. For the evaluation of the synthetic data we use the following error measure:

- Angle difference in $\mathbf{R}$: $\xi_R = cos^{-1}((Tr(\mathbf{R}\dot{\mathbf{R}}^\top) - 1)/2)$
- Direction difference in $\mathbf{t}$: $\xi_t = cos^{-1}((\mathbf{t}^\top\dot{\mathbf{t}})/(\|\mathbf{t}\|\|\dot{\mathbf{t}}\|))$

Where $\mathbf{R}$, $\mathbf{t}$ denote the ground-truth transformation and $\dot{\mathbf{R}}$, $\dot{\mathbf{t}}$ are the corresponding estimated transformations.

Fig. 3 and Fig. 4 compare the 3-point homography based algorithm to the general 5-point [2] and the 3-point algorithms [13]. The evaluation shows that the proposed method outperforms the 5pt algorithm, in terms of accuracy. Under perfect IMU measurements the algorithm is robust to image noise and performs significantly better than the 5pt algorithm and equally good as the 3pt algorithm. With increasing IMU noise the performance of the 3pt and 3pt-homography (proposed) algorithm are still comparable to the 5pt algorithm.
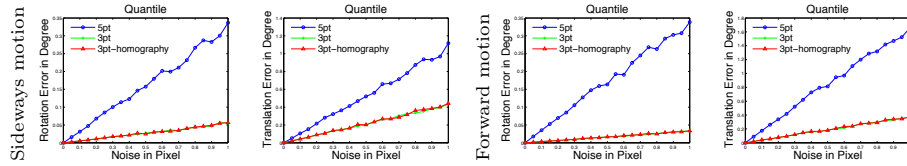

Fig. 3. Evaluation of the 3 point algorithm under varying image noise for two different motion settings (sideways and forward motion)
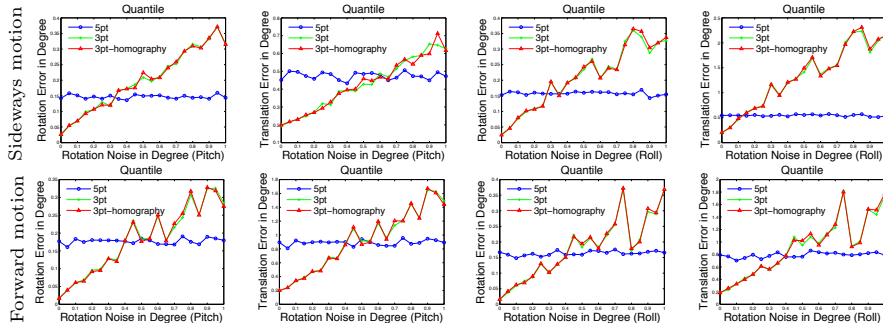

Fig. 4. Evaluation of the 3-point algorithm under different IMU noise and constant image noise of 0.5 pixel standard deviation. First row: sideways motion of the camera with varying pitch angle (left) and varying roll angle (right). Second row: forward motion of the camera with varying pitch angle (left) and varying roll angle (right).

4.2   Experiments on real data with Vicon ground-truth

In order to have a practical evaluation of our algorithm, several real datasets have been collected with reliable ground-truth data. The ground-truth data has been obtained by conducting experiments in a room equipped with a Vicon motion capture system made of 22 cameras. Synchronization between visual and ground-truth data has been obtained by a pre-processing step. We used the VICON data as inertial measures and scale factor in the different experiments. The sequences have been acquired with a perspective camera mounted on a mobile robot and on an handheld system in order to have planar and general trajectories. The lengths of these trajectories are between 20 and 50 meters and the number of images are between 200 and 350 per sequence. We propose a comparison with the five-point algorithm and the general three-point algorithm (implemented after [13] in order to prove the efficiency of the proposed method. Both methods use the same matched feature point sets as input and apply RANSAC algorithm in order to select the inliers. For the case of the 5pt algorithm and the standard 3pt algorithm a least squares estimation is performed on the inlier sets for the estimation of the motion parameters, while for the proposed 3pt method this was not possible.

Figure 5 shows the results for the GT1 sequence. For this sequence the camera has been mounted on a mobile robot in forward direction the z-axis of the camera parallel to the ground plane. The robot was mainly moving forward which poses a difficult situation for a visual odometry system. We can note that all the algorithms show trajectories with a strong drift in z direction (height) even if the global shapes of the trajectories are quite similar. The mean angular error of the translation and the mean error of the rotation of consecutive frames during the sequence are respectively equal to 0.2399 and 0.0052 for the general three-point, 0.2455 and 0.0048 for the three-point homography and 0.4302 and 0.0112 for the five-point methods. The details of these errors during the complete sequence are given in Figures 5 (c) and (d). The three-point homography method presents the best results and also the minimum drift at the end of the motion. Such drift as visible in the plots is typical for visual odometry systems without structure triangulation and local optimization. It is however noticeable that the purely vision based 5pt algorithm has much more difficulties with this sequence than the 3pt methods that use an additional common direction.

Figure 6 shows the results for the GT2 sequence. In this sequence the camera has been handheld, pointing toward the ground in a 45 degree angle and was moved sideways in a circular trajectory. This comprised a less difficult setting than for GT1. In this case, the five-point algorithm presents globally better results but the three-point homography algorithm is better than the general three-point approach. As previously, the details of the rotation and translation errors between consecutive images are given in Figures 6 (c) and (d). We can note that the rotation error of the three-point homography is smaller than the two other approaches.

5   Conclusion

In this paper we presented a new formulation for the 3pt plus common direction relative pose problem using a homography formulation. We show that it is possible to utilize the homography constraint for an alternative formulation for the 3pt problem. With the same number of input point correspondences the homography formulation also solves for the plane normal in addition to the motion parameters. In experiments with synthetic data and real image sequences we show that the alternative formulation produces similar results to the standard 3pt and also demonstrate that additional information in kind of a common direction allows to get better visual odometry results in certain configurations.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
2. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 756–777
3. Scaramuzza, D.: 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. International Journal of Computer Vision **95** (2011) 74–85
4. Kneip, L., Martinelli, A., Weiss, S., Scaramuzza, D., Siegwart, R.: Closed-form solution for absolute scale velocity determination combining inertial measurements and a single feature correspondence. In: IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011. (2011) 4546–4553
5. Viéville, T., Clergue, E., Facao, P.D.S.: Computation of ego-motion and structure from visual and inertial sensors using the vertical cue. In: Computer Vision, 1993. Proceedings., Fourth International Conference on, IEEE (1993) 591–598
6. Corke, P.: An inertial and visual sensing system for a small autonomous helicopter. Journal of Robotic Systems **21** (2004) 43–51
7. Lobo, J., Dias, J.: Vision and inertial sensor cooperation using gravity as a vertical reference. IEEE Trans. Pattern Anal. Mach. Intell. **25** (2003) 1597–1608
8. Weiss, S., Siegwart, R.: Real-time metric state estimation for modular vision-inertial systems. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE (2011) 4531–4537
9. Domke, J., Aloimonos, Y.: Integration of visual and inertial information for ego-motion: a stochastic approach. In: Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, IEEE (2006) 2053–2059
10. Kalantari, M., Hashemi, A., Jung, F., Guédon, J.P.: A new solution to the relative orientation problem using only 3 points and the vertical direction. Journal of Mathematical Imaging and Vision **39** (2011) 259–268
11. Antone, M.E., Teller, S.J.: Automatic recovery of relative camera rotations for urban scenes. In: CVPR, IEEE Computer Society (2000) 2282–2289
12. Oreifej, O., da Vitoria Lobo, N., Shah, M.: Horizon constraint for unambiguous uav navigation in planar scenes. In: IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011. (2011) 1159–1165

13. Fraundorfer, F., Tanskanen, P., Pollefeys, M.: A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In Daniilidis, K., Maragos, P., Paragios, N., eds.: ECCV (4). Volume 6314 of Lecture Notes in Computer Science., Springer (2010) 269–282
14. Naroditsky, O., Zhou, X.S., Gallier, J.H., Roumeliotis, S.I., Daniilidis, K.: Two efficient solutions for visual odometry using directional correspondence. IEEE Trans. Pattern Anal. Mach. Intell. **34** (2012) 818–824
15. Chum, O., Werner, T., Matas, J.: Two-view geometry estimation unaffected by a dominant plane. In: CVPR (1), IEEE Computer Society (2005) 772–779
16. Szeliski, R., Torr, P.H.S.: Geometrically constrained structure from motion: Points on planes. In Koch, R., Gool, L.J.V., eds.: SMILE. Volume 1506 of Lecture Notes in Computer Science., Springer (1998) 171–186
17. Olivier Saurer, F.F., Pollefeys, M.: Homography based visual odometry with known vertical direction and weak manhattan world assumption. In: IEEE/IROS Workshop on Visual Control of Mobile Robots (ViCoMoR 2012). (2012)
18. Ortin, D., Montiel, J.: Indoor robot motion based on monocular images. Robotica **19** (2001) 331–342
19. Troiani, C., Martinelli, A., Laugier, C., Scaramuzza, D.: 1-point-based monocular motion estimation for computationally-limited micro aerial vehicles. In: Mobile Robots (ECMR), 2013 European Conference on, IEEE (2013) 13–18
20. Bazin, J.C., Demonceaux, C., Vasseur, P., Kweon, I.: Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. The International Journal of Robotics Research **31** (2012) 63–81
21. Li, B., Heng, L., Lee, G.H., Pollefeys, M.: A 4-point algorithm for relative pose estimation of a calibrated camera with a known relative rotation angle. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS 2013, Tokyo, Japan. (2013)
22. Troiani, C., Martinelli, A., Laugier, C., Scaramuzza, D.: 2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles. In: IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, 2014. (2014)
23. Martinelli, A.: Closed-form solution of visual-inertial structure from motion. International Journal of Computer Vision **106** (2014) 138–152
24. Lee, G.H., Fraundorfer, F., Pollefeys, M.: Mav visual slam with plane constraint. In: ICRA, IEEE (2011) 3139–3144
25. Cox, D.A., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, 3/e (Undergraduate Texts in Mathematics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
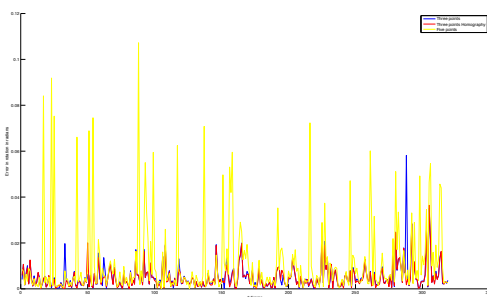26. Kukelova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. (In: ECCV (3)) 302–315
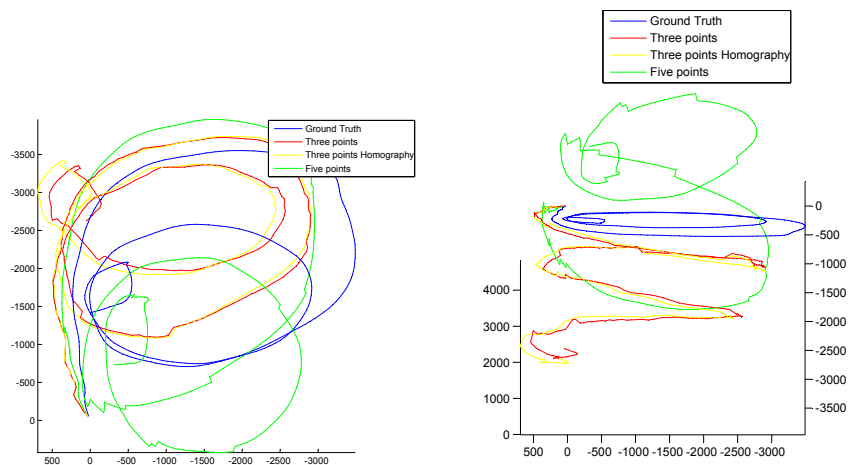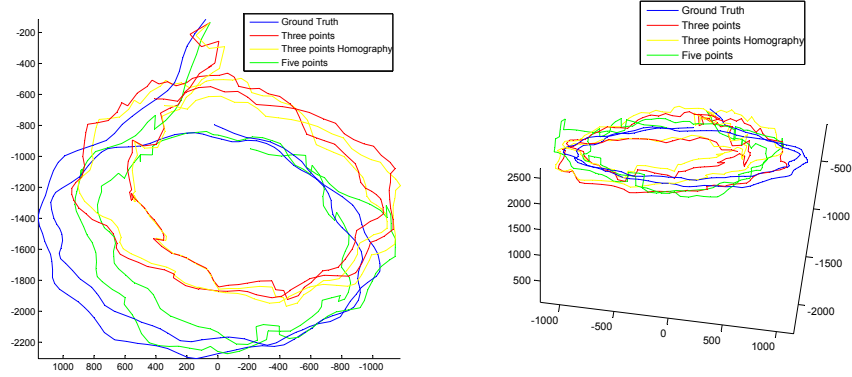
(a)

(b)
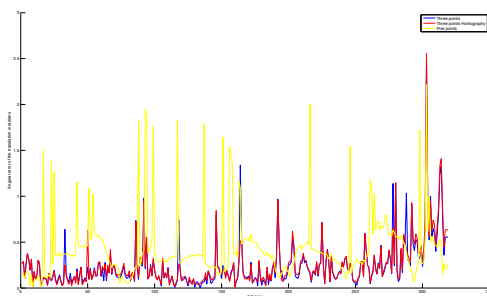
(c)
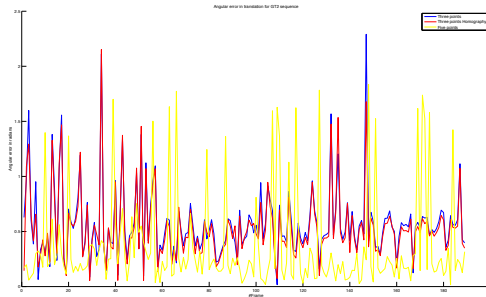
(d)

Fig. 5. Results for sequence GT1 - (a) Top view of the trajectories estimated with general three-point (red curve), homography three-point (yellow curve), five-point (green curve) algorithms compared to the Vicon ground-truth (blue curve) (axis in mm) - (b) Side view - (c) Angular error in translation between consecutive frames during the sequence - (d) Rotation error between consecutive frames during the sequence
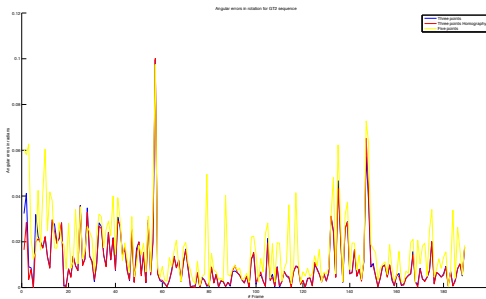
(a)

(b)

(c)

(d)

Fig. 6. Results for sequence GT2 - (a) Top view of trajectories estimated with general three-point (red curve), homography three-point (yellow curve), five-point (green curve) algorithms compared with the Vicon ground-truth (blue curve) (axis in mm) - (b) Side view - (c) Angular error in translation between consecutive frames during the sequence - (d) Rotation error between consecutive frames during the sequence