An Efficient Solution to the Homography-Based Relative Pose Problem With a Common Reference Direction

Yaqing Ding, Jian Yang, Jean Ponce, Hui Kong

An Efficient Solution to the Homography-Based Relative Pose Problem With a Common Reference Direction

Yaqing Ding[1], Jian Yang[1], Jean Ponce[2,3], and Hui Kong[1,4]

[1]Nanjing University of Science and Technology, Nanjing, China [2]INRIA, Paris, France

{dingyaqing,csjyang,konghui}@njust.edu.cn jean.ponce@inria.fr

[3]Département d'informatique de l'ENS, ENS, CNRS, PSL University, Paris, France [4] IAAI Nanjing, Horizon Robotics

Abstract

In this paper, we propose a novel approach to two-view minimal-case relative pose problems based on homography with a common reference direction. We explore the rank-1 constraint on the difference between the Euclidean homography matrix and the corresponding rotation, and propose an efficient two-step solution for solving both the calibrated and partially calibrated (unknown focal length) problems. We derive new 3.5-point, 3.5-point, 4-point solvers for two cameras such that the two focal lengths are unknown but equal, one of them is unknown, and both are unknown and possibly different, respectively. We present detailed analyses and comparisons with existing 6- and 7-point solvers, including results with smart phone images.

1. Introduction

RANSAC [7] has been established as one of the most successful techniques for model estimation from data with outliers. Using a minimal point-correspondence is of extreme importance for handling large amounts of outliers in the image correspondences [8]. However, reducing the number of point correspondence requires reasonable assumptions or extra data, *e.g.*, coplanar points, planar motion, pure rotation or translation, etc. When a common reference direction (*e.g.*,gravity) is known for both cameras, there is only one unknown rotation parameter [41], and the relative pose of the cameras has only four degrees of freedom. In practice, modern cellphones or camera-IMU (inertial measurement unit) systems have accelerometers which can provide the gravity vector. Similar situations arise when we detect a vanishing point. On the other hand, in indoor and outdoor environments, planar parts of the scene are often dominant, such as floor, walls, doors, street or other general structures. In this paper, we *assume* that the scene contains at least one plane, and propose new homography-based minimal solvers to calibrated and partially calibrated relative pose estimation with a common reference direction. The main contributions of this paper are:

- A rank-1 constraint on the difference between the *Euclidean homography matrix* and the corresponding rotation is introduced to formulate the problem efficiently.
- Based on this constraint, we propose an efficient two-step solution for eliminating part of the unknowns from the original equations.
- This allows us to solve the homography-based relative pose problems with a common direction, including the calibrated case (3-point), two cameras with equal focal lengths (3.5-point), two cameras where one is calibrated and the other's focal length is unknown (3.5-point), and two cameras with different focal lengths (4-point).

2. Related work

When the two cameras are semi-calibrated, in the sense that the only unknown intrinsic parameter are the focal lengths, at least seven point correspondences are needed to recover the motion and focal lengths of two cameras [16, 2]. If the two cameras have the same focal length, at least six point correspondences are needed [38, 27, 14, 21, 22]. One the other hand, if 4 points are coplanar, only one extra point is enough [40]. When one camera is fully calibrated and the other one's focal length is unknown, six point correspondences are needed as well [3, 21, 22]. Given a fully calibrated camera, it is well known that the relative pose problem can be solved using five point correspondences [33, 28, 21]. If the points are coplanar, homography-based solver need four point correspondences [15]. Recently, it has been shown that with a common direction, the two-view relative pose problem can be solved with only three point correspondences using epipolar constraints [9, 32, 39] or homographies [37, 36]. In [11], the author propose a homography-based relative pose estimation assuming a known vertical direction and a dominant ground plane.

The most closely related work to ours for the calibrated case is by Saurer *et al*. [37, 36]. In [36] the authors generate a 443×451 G-J (Gauss-Jordon) elimination template using [19]. The two-step solution proposed in Section 4 gives a template with size 24×34. Since the computational complexity of G-J elimination is $o(n^3)$, the speedup is significant. However, for the calibrated relative pose with known direction, essential-matrix based solvers are more efficient, *e.g*., the solver proposed by Sweeney *et al*. [39] only needs to solve a quadratic eigenvalue problem.

For the partially calibrated two-view relative pose problem with equal focal lengths, the most closely related work to ours is Malis *et al*. [29]. They compute the homography between two views, then try all the possible focal lengths and choose the one which minimizes the cost function. In fact, their method needs multiple views to improve accuracy. Other related works are the 6-point solvers [38, 27, 14, 21, 22], which are all based on rank-2 and trace constraints, and the 5-point (4 coplanar points plus 1 point out of the plane) solver [40]; for the one unknown focal length problem, the most closely related work is the single-side 6-point solver [3, 21, 22]; for the varying focal lengths problem, it is the 7-point solver using Bougnoux formula [2]. Compared to the SOTA (state-of-the-art) solvers, the partially calibrated solver proposed in Sec. 4 has two advantages. 1) It needs 3.5 or 4 instead of 6 or 7 points, which is important for RANSAC, 2) it is more accurate.

3. Problem formulation

Let π denote the observed plane and $n = [n_x, n_y, n_z]^T$ be its unit normal vector in the first camera frame. The distance from the optical center of the first camera to the plane is denoted by d. Suppose this plane is observed from two views with projection matrices $P_1 = K_1[\ I \mid 0\ ]$ and $P_2 = K_2[R \mid t\ ]$, where K_1, K_2 are the camera intrinsic matrices and $\{R, t\}$ are the rotation matrix and translation vector, respectively. The homography for points belonging to the plane between consecutive frames is defined as follows:

$$\lambda m_2 = G m_1, \tag{1}$$

with

$$G \sim K_2 H K_1^{-1}, \tag{2}$$

where λ is a scale factor, $\sim$ indicates equality up to a scale factor, and $m_1 = [u_1, v_1, 1]^T$ and $m_2 = [u_2, v_2, 1]^T$ are the homogeneous coordinates of the point in the first and second images. H, the so-called *Euclidean homography matrix* is defined by

$$H = (R - \frac{1}{d} t n^T). \tag{3}$$

In this paper, we assume that the views have a common reference direction. We can use a vanishing point or the gravity direction obtained by an IMU on mobile phones and robots for this direction. Without loss of generality, we can align the y-axes of the two cameras to the common reference direction (see Figure 1). Using this direction, we can compute the rotation matrices R_1, R_2 of the two cameras for the alignment. Applying the rotations to the normalized image points we obtain

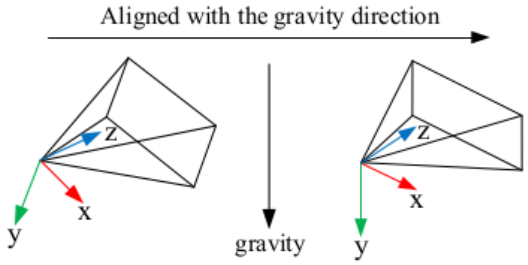


Figure 1: One can calculate the roll, pitch angles of the camera coordinate with respect to the world coordinate (gravity) using the IMU data, and align the y-axes with the gravity.

$$\lambda R_2^T K_2^{-1} m_2 = H_y R_1^T K_1^{-1} m_1, \tag{4}$$

where

$$H_y = (R_y - \frac{1}{d'} t' n'^T), \tag{5}$$

with

$$R_y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{6}$$

$$R = R_2 R_y R_1^T,\ \frac{t}{d} = R_2 \frac{t'}{d'},\ n = R_1 n'. \tag{7}$$

Our aim is to estimate the relative pose using homography.

4. Our approach: Rank-1 constraint

Given two matching image points $m_1 = [u_1, v_1, 1]^T$ and $m_2 = [u_2, v_2, 1]^T$, the homography constraint (1) can also be written as

$$\begin{bmatrix} 0 & 0 & 0 & -u_1 & -v_1 & -1 & v_2 u_1 & v_2 v_1 & v_2 \\ u_1 & v_1 & 1 & 0 & 0 & 0 & -u_2 u_1 & -u_2 v_1 & -u_2 \end{bmatrix} g = 0, \tag{8}$$

$$g = [g_1\ g_2\ g_3\ g_4\ g_5\ g_6\ g_7\ g_8\ g_9]^T. \tag{9}$$

By stacking the constraint rows for n point correspondences, (8) leads to a system of equations of the form

$$Ag = 0, \tag{10}$$

where A is a $2n \times 9$ matrix. Then g and the 2D homography matrix G can be formulated as the null space of A. Based on (1), (2), (4) we obtain the relationship between the *Euclidean homography* H_y and 2D homography G

$$H_y \sim R_2^T K_2^{-1} G K_1 R_1, \tag{11}$$

where $\sim$ indicates equality up to a scale factor. Unlike the essential matrix which obeys rank-2 and trace constraints, there are no explicit constraints for a *Euclidean homography matrix*. In this case, we define a new matrix Q

$$Q = R_2^T K_2^{-1} G K_1 R_1 - k R_y, \tag{12}$$

which is the difference between the homography matrix H_y and the corresponding rotation R_y. Note that there is a scalar k in front of the rotation matrix, because (11) is up to scale. Using (5), we obtain

$$Q = -kt'n'^T, \tag{13}$$

and Q has rank 1 (when $t' = 0$, rank$(Q) = 0$, which means a pure rotation, which will be discussed in Section 5). In particular, each of the 2×2 submatrices of Q must have zero determinant. For succinct representation, let $c = k\cos\theta, s = k\sin\theta$, and we obtain

$$Q = \begin{bmatrix} h_1 - c & h_2 & h_3 - s \\ h_4 & h_5 - k & h_6 \\ h_7 + s & h_8 & h_9 - c \end{bmatrix}, \tag{14}$$

with

$$k^2 = c^2 + s^2. \tag{15}$$

This expression can reduce the degree of the monomials. Since each 2×2 submatrices of Q must vanish, (14) leads to the following system of nine equations

$$h_1(h_5 - k) - c(h_5 - k) - h_2 h_4 = 0, \tag{16}$$
$$h_9(h_5 - k) - c(h_5 - k) - h_6 h_8 = 0, \tag{17}$$
$$h_3(h_5 - k) - s(h_5 - k) - h_2 h_6 = 0, \tag{18}$$
$$h_7(h_5 - k) + s(h_5 - k) - h_4 h_8 = 0, \tag{19}$$
$$sh_4 - ch_6 + h_1 h_6 - h_3 h_4 = 0, \tag{20}$$
$$sh_6 + ch_4 + h_6 h_7 - h_4 h_9 = 0, \tag{21}$$
$$sh_2 + ch_8 + h_2 h_7 - h_1 h_8 = 0, \tag{22}$$
$$sh_8 - ch_2 + h_2 h_9 - h_3 h_8 = 0, \tag{23}$$
$$h_1 h_9 - h_3 h_7 - ch_1 - ch_9 - sh_3 + sh_7 + k^2 = 0. \tag{24}$$

We need to solve the system of 10 equations (15)-(24).

Efficient two-step solution

However, a well known property of a *Euclidean homography matrix* is that it can be decomposed into rotation, translation and normal of the plane $\{R, t, n\}$ using the SVD. Once the *Euclidean homography matrix* is determined, the rotation matrix can be extracted too. It means that we do not need to calculate the unknowns of the homography matrix and the rotation matrix simultaneously. In equation (16)-(24), we can divide them into two subsets: the homography components $\{h, k\}$ and the rotation components $\{c, s\}$. In this case, we can try to eliminate the rotation components first instead of solving the 10 polynomial equations directly.

By exploiting relationship among these nine equations: (16)-(17), (18)+(19), $(20)^2$+$(21)^2$, $(22)^2$+$(23)^2$ and substituting (16) (18) into (24), we get a total of 5 polynomial equations without the rotation parameters $\{c, s\}$

$$\begin{aligned}
(h_1 - h_9)(h_5 - k) - h_2 h_4 + h_6 h_8 &= 0, \\
(h_3 + h_7)(h_5 - k) - h_2 h_6 - h_4 h_8 &= 0, \\
k^2(h_2^2 + h_8^2) - (h_1 h_8 - h_2 h_7)^2 - (h_3 h_8 - h_2 h_9)^2 &= 0, \\
k^2(h_4^2 + h_6^2) - (h_1 h_6 - h_3 h_4)^2 - (h_4 h_9 - h_6 h_7)^2 &= 0, \\
(h_5 - k)(h_7^2 + h_9^2 - k^2) + (h_3 - h_7) h_4 h_8 - (h_1 + h_9) h_6 h_8 &= 0.
\end{aligned} \tag{25}$$

The system of polynomial equations (25) can also be generated using the computer algebra system Macaulay2 [10].

4.1. Calibrated case

With 3 point correspondences, the general solution of g in (10) is a 3-dimensional null space that can be written as

$$g = \alpha g_a + \beta g_b + g_c, \tag{26}$$

Substituting (26) into the new system of polynomial equations (25), we get 5 equations in three unknowns $\{\alpha, \beta, k\}$. The degrees of these 5 equations are $\{2, 2, 4, 4, 3\}$, respectively. Now we solve the system of equations (25) using action matrix method [5]. As described in [4], we found that, in order to create the action matrix, we need to generate additional equations by multiplying the initial equations with monomials. The degrees of the additional equations are up to four, which means that we need to multiply the first and second equations in (25) with $\{k, \alpha, \beta, k^2, \alpha^2, \beta^2, k\alpha, k\beta, \alpha\beta\}$ and the last one in (25) with $\{k, \alpha, \beta\}$. There are in total 26 polynomials with 35 monomials. After removing 2 unnecessary equations and 1 monomial which is not used in the action matrix, we obtain a 24×34 template for the G-J elimination. Then we construct the 10×10 action matrix for solving $\{k, \alpha, \beta\}$. The template size of the solver is much smaller than that of [36], which is 443×451. Since the computational complexity of G-J elimination is $o(n^3)$, the speedup is significant. Once $\{\alpha, \beta, k\}$ are calculated, $\{c, s\}$ can be linearly obtained from two of the nine equations (16)-(24). Finally the translation and normal vector can be extracted from matrix Q. We do not need to use the SVD to compute $\{R, t, n\}$. Once H_y is calculated, the rotation is uniquely determined, with two possible opposite translations.

Note that although the proposed solver is much more efficient than the SOTA homography-based solver [36] (600x speedup in practice), there are more efficient solvers based on the essential matrix if one wants to solve for relative pose with known direction, *e.g.*, the one proposed by Sweeney *et al*. [39]. The solver [39] only needs to solve a quadratic eigenvalue problem and the only difference is that the homography-based solvers simultaneously calculate the

normal of the plane. Since both formulations are solving the same problem with 3 point correspondences, they should return very similar results. In this case, for the calibrated relative pose estimation, we recommend [39]. We just show an efficient solution to the formulation which is used in [36].

4.2. Partially calibrated case

Assuming the only unknown camera parameter is the focal length. For most CCD or CMOS cameras, it is reasonable to assume the cameras have square-shaped pixels, and the principal point coincides with the image center [14]. In addition, smart phones, tablets are special because the relationship between the axes of the camera and the IMU are usually approximate to $0°$, $\pm 90°$, $180°$ [20, 12]. Therefore, for uncalibrated smart phones and tablets we can directly use the images and gravity direction.

Equal and unknown focal length: f+H+f 3.5-point

The first situation is when the two cameras have the same focal length, and this is a 7-DOF problem. In this case, we need at least 3.5 point correspondences (we still need to sample 4 point correspondences, but only need one equation from the last correspondence) and the general solution of g in (10) is a 2-dimensional null space that can be expressed by

$$g = \eta g_a + g_b. \tag{27}$$

H_y can be formulated with $\{f, \eta\}$ using (11), and then substituting H_y into (25) results in a system of 5 equations with 3 unknowns: the focal length f, the scalar η of the null space g_a and the scale factor k of the rotation matrix. In order to solve the system efficiently, we use an automatic generator for the polynomial equations, *e.g.*, [24, 25, 23]. We obtain a template of size 152×176, and the system has 24 solutions. In practice, 4 of these solutions have been found to be always imaginary. The remaining pose and normal can be extracted as in the calibrated case.

One unknown focal length: H+f 3.5-point

We also address the problem where the first camera is calibrated and the only unknown is the focal length of the second camera. In this case, the general solution for g can be expressed by (27) as well. Then H_y can be formulated with $\{f, \eta\}$ using (11). Substituting H_y into the equation system (25) we may obtain a G-J elimination template of size 64×76, and the system has up to 12 solutions.

Different and unknown focal lengths: $\mathbf{f_1}$+H+$\mathbf{f_2}$ 4-point

For two cameras with different focal lengths, we have an 8-DOF problem and we need at least 4 points. With 4 point correspondences, g is uniquely determined and H_y can be formulated with f_1, f_2 using (11). Substituting H_y into (25) results in a system of 5 equations with 3 unknowns: the focal lengths f_1, f_2 and the scalar k of the rotation matrix. The remaining steps are the same as the equal and unknown focal length case. Solving the system of equations (25) results in a template of size 33×41, and the system has 8 solutions. See the supplementary material for more details.

We need to mention that the 4-point (f_1Hf_2) solver has multiple physically possible solutions. The 3.5-point solvers have one parameter in the 2-dimensional null space, and a unique solution can be found using RANSAC. However, with 4 points the 2D homography is uniquely determined, and the f_1Hf_2 solver has up to 8 solutions. Under positive focal length and depth constraint, there may still be multiple (usually two) physical possible solutions. This is similar to the standard 4-point homography algorithm which has two physically possible poses [6, 31]. In order to choose the good solution, it is necessary to have additional information such as an estimate of the normal to the target plane [30], or using the points out of this plane.

Improving numerical stability

In order to improve the numerical stability, the points should be normalized [17]. We use the varying focal lengths problem as an example. First we translate the origin to the image center, then scale the points so that the average distance from the origin is equal to $\sqrt{2}$. In this case, we may obtain two transformation matrices: $T_1 = \mathrm{diag}(\sigma_1, \sigma_1, 1)$, $T_2 = \mathrm{diag}(\sigma_2, \sigma_2, 1)$. Applying the transformation to the image points, and the corresponding camera matrices become $K'^{-1}_1 = \mathrm{diag}(1, 1, \sigma_1 f_1), K'^{-1}_2 = \mathrm{diag}(1, 1, \sigma_2 f_2)$. Once the equations are solved, the focal length only need to be divided by σ_1, σ_2. This normalization is necessary. We show more details in the supplementary material.

5. Degenerate configurations

In our case, there are three kinds of degeneracies. The first is caused by the data, *e.g.*, three or more points collinear, which leads to a rank loss of the matrix A in (10). It can be easily eliminated by RANSAC [7] or its variants [26, 35]. The second one is induced by the formulation. In (13), we have mentioned that a pure rotation will result in $\mathrm{rank}(Q) = 0$. In this case, the zero determinant of the 2×2 submatrices of Q is not a full constraint, since every element of Q should be zero. However, based on experiments we find that the proposed solvers can deal with the pure rotation case. The last one is that arbitrary planar motions when the optical axes lie in the plane are critical motions for the standard 6-point and 7-point solvers [18]. The proposed *fHf* solver can deal with this case (non-zero rotation angle). The 3.5-point Hf solver can deal with pure translation (except for $R_1 = R_2 = I, t_y = 0$). For more details see the supplementary material.

6. Experiments

In this section, we test the solvers on synthetic data with known ground truth. The data with different noise levels are used to evaluate the numerical stability and precision. For

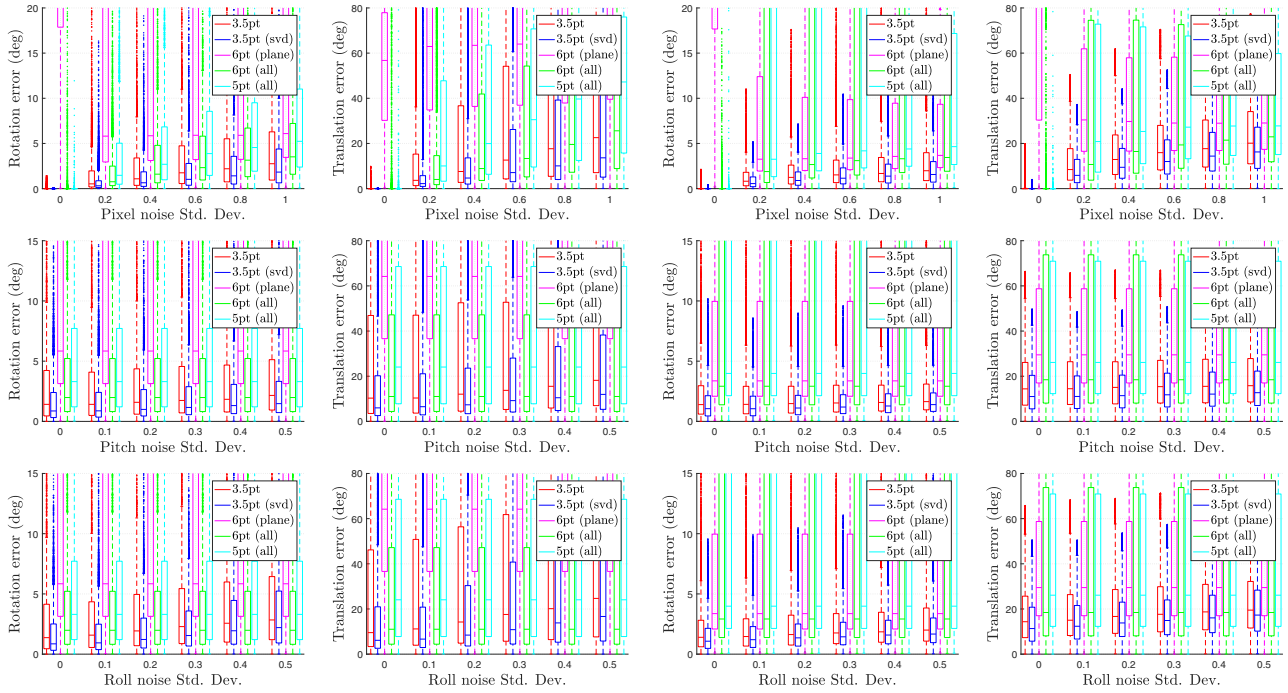


Figure 2: Rotation and translation errors for equal and unknown focal length problem, respectively. (a) Sideways motion. (b) Forward motion. From top to bottom: increased image noise, increased Pitch noise and 0.5 pixel standard deviation image noise, increased Roll noise with constant 0.5 pixel standard deviation image noise.

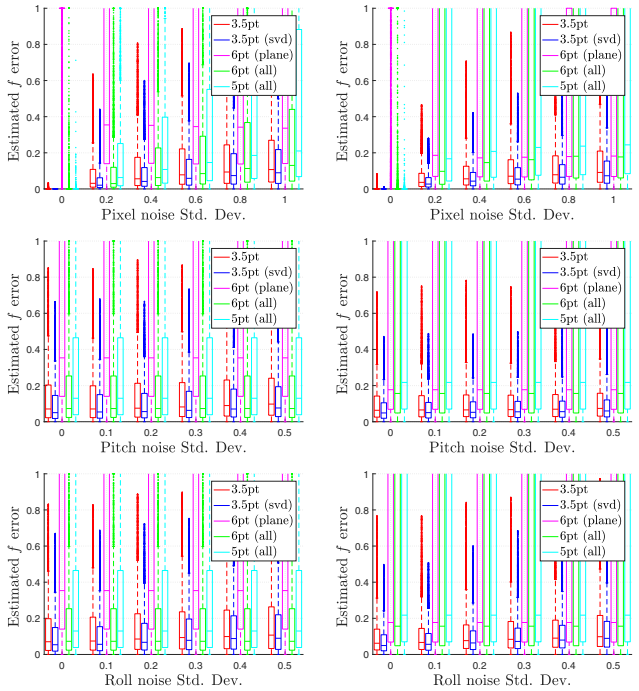


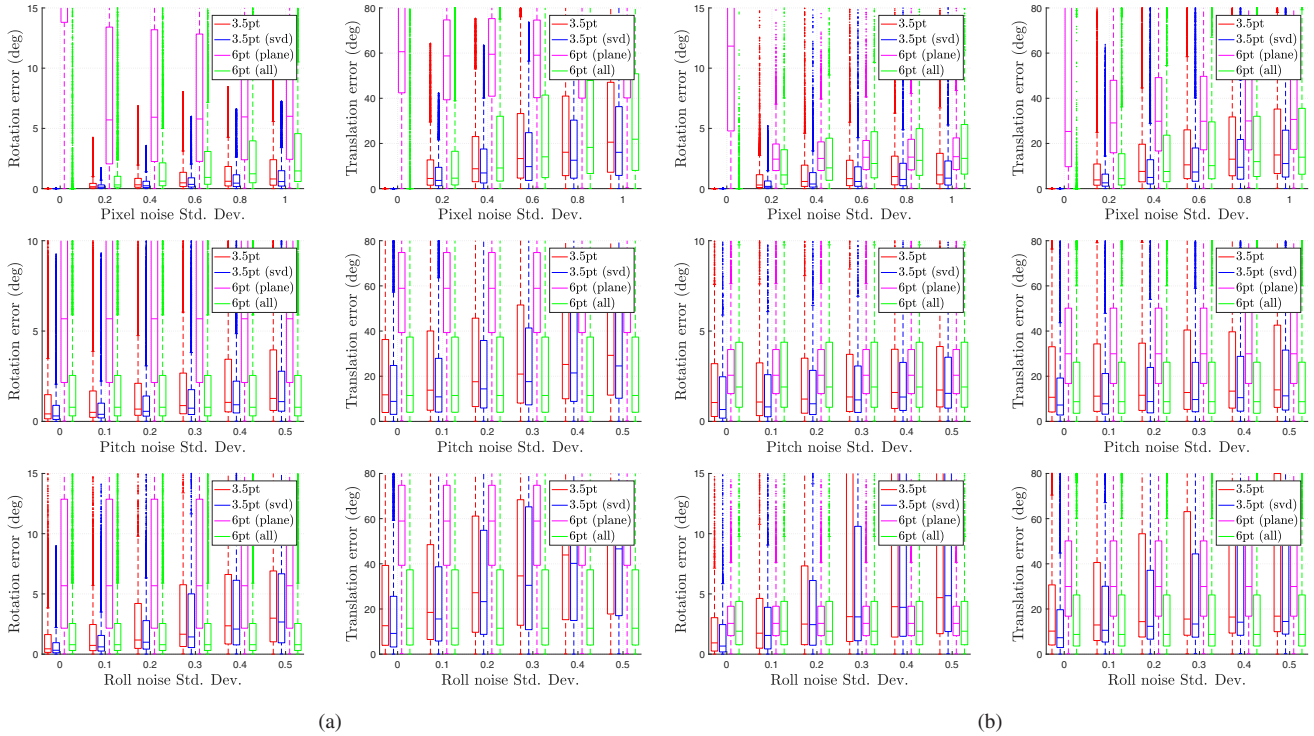
Figure 3: Focal length errors for equal and unknown focal length problem. Left column: Sideways motion. Right column: Forward motion.

the equal focal length problem, since [27, 38, 14, 22] use the same formulation and constraints (rank-2 and trace constraints of the essential matrix), so we only compare with one of them, the SOTA solver [22]. In addition, we also compare with the 4-plus-1 solver (4 coplanar points plus 1 point out of the plane) [40]. For the one unknown focal length problem, [3] and [22] also use same formulation and constraints, so we only compare with SOTA [22]. For the varying focal lengths problem, we compare with the 7-point solver using Bougnoux formula [2]. Since we still need to sample 4 points for the 3.5-point solvers, we show the results of both the minimal case and the 4 points case which uses SVD to compute the 2-dimensional null space.

The synthetic data are generated in the following setup. We sample 200 3D points distributed on a random but feasible plane. The focal length of the camera is randomly generated: $f_g \in [300, 3000]$ pixels with a field of view of 90 degrees, and the resolution of the image is set to $2f \times 2f$. The rotation error ξ_R, translation error ξ_t and focal length error ξ_f are defined as:

- $\xi_R = \arccos((\mathrm{tr}(R_g R_e^T) - 1)/2)$,
- $\xi_t = \arccos((t_g^T t_e)/(\|t_g\|\|t_e\|))$,
- $\xi_f = \frac{|f_e - f_g|}{f_g}$,

where R_g, t_g, f_g represent the ground-truth rotation, translation and focal length, and R_e, t_e, f_e are the estimated rotation, translation and focal length, respectively. Note that,



(a) (b)

Figure 4: Rotation and translation errors for one unknown focal length problem, respectively. (a) Sideways motion. (b) Forward motion. From top to bottom: increased image noise, increased pitch noise and 0.5 pixel standard deviation image noise, increased roll noise with constant 0.5 pixel standard deviation image noise.

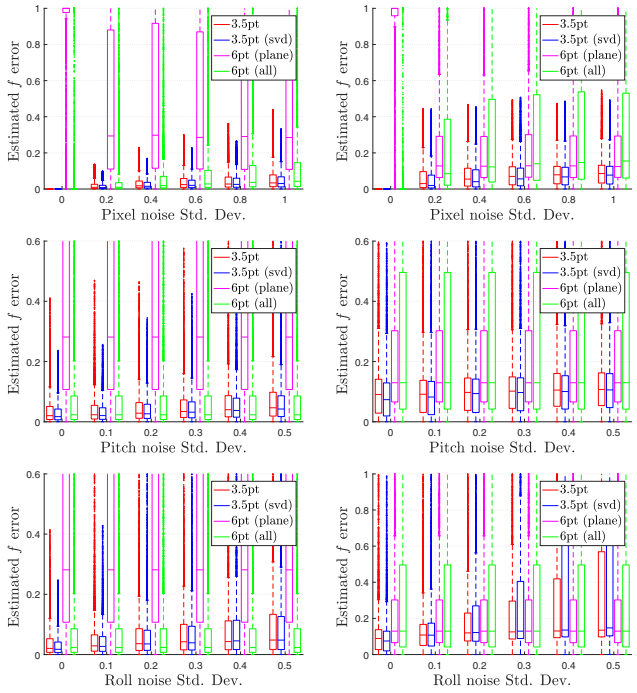


Figure 5: Focal length errors for one unknown focal length problem. Left column: Sideways motion. Right column: Forward motion.

for our solvers, we measure the *full* rotation and translation errors based on the noisy (roll, pitch) angles (using (7)).

We generated 2,000 random scenes with 3D points on different planes and different transformations between two views. For each solver, we test 5 times on one scene and obtain 10,000 results for each solver. Since the solvers have multiple solutions, by evaluating the geometric error of the fit of each solution with respect to a set of points, we can pick the best one. We evaluate each solver under image noise (point location) with different standard deviation and increased (roll, pitch) noise. The (roll, pitch) noise can be considered as the direction error. We focus on sideways (parallel to the scene) and forward (along the z-axis) motions. The base line between two cameras is set to be 10 percent of the average scene distance. Additionally, the cameras are rotated around every axis. This is similar to [34, 9].

Figure 2 and Figure 3 show the estimated rotation, translation and focal length errors under sideways and forward motion for the equal and unknown focal length problem. The figures are plotted using boxplot function in Matlab. The bottom and top edges of the box indicate the lower and upper quartile, and the central mark indicates the median. The points plotted individually are considered as outliers. We compare the proposed 3.5-point and 3.5-point (SVD) solver with the 6-point solver [22] and 5-point solver [40]. We also evaluate the 6- and 5-point solver with general

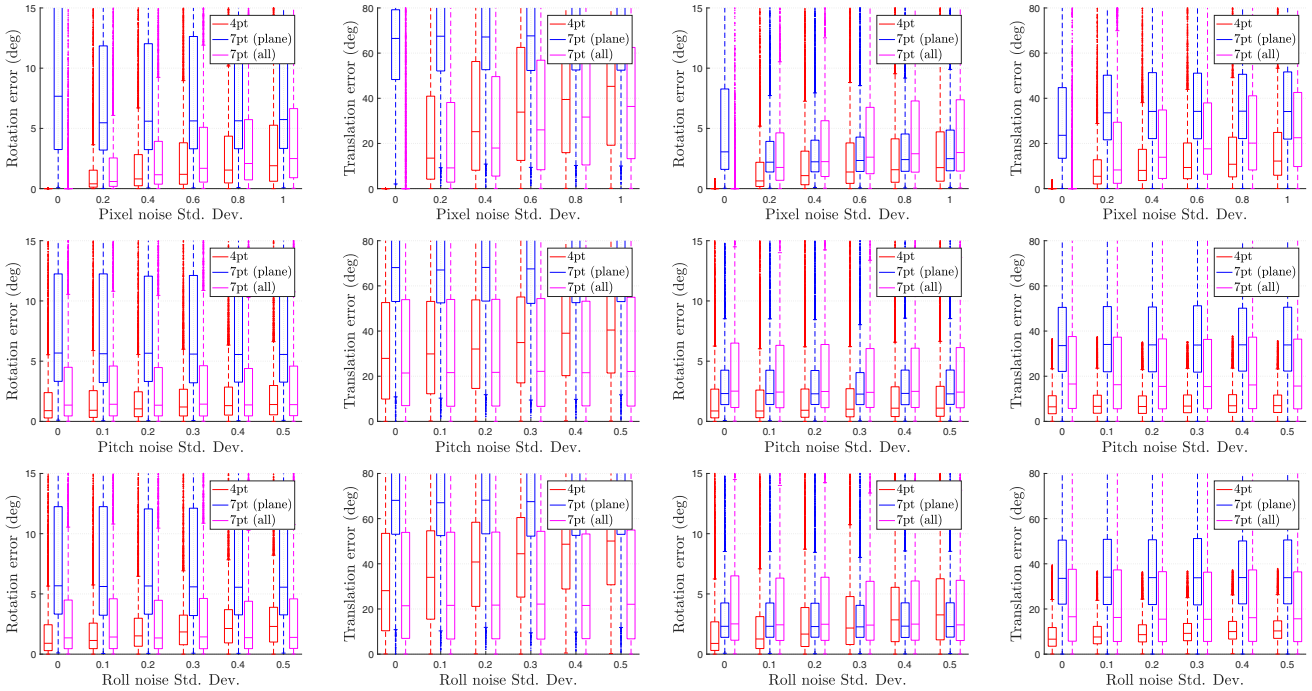


Figure 6: Rotation and translation errors for different and unknown focal length problem, respectively. (a) Sideways motion. (b) Forward motion. From top to bottom: increased image noise, increased pitch noise and 0.5 pixel standard deviation image noise, increased roll noise with constant 0.5 pixel standard deviation image noise.

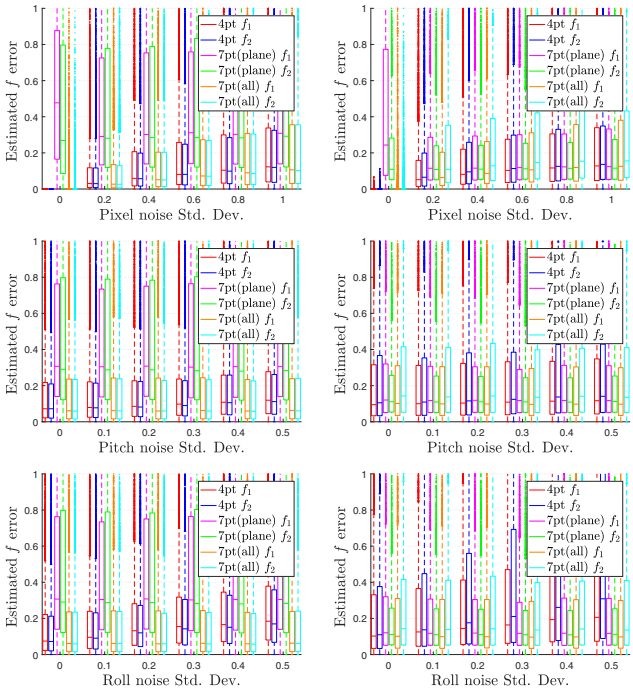


Figure 7: Focal length errors for different and unknown focal length problem. Left column: Sideways motion. Right column: Forward motion.

scenes (with 200 extra non-planar points generated randomly). Figure 4 and Figure 5 show the estimated rotation, translation and focal length errors under sideways and forward motion for the one unknown focal length problem. Figure 6 and Figure 7 show the estimated rotation, translation and focal length errors under sideways and forward motion for the different focal lengths problem.

In general, under perfect direction information the proposed solvers outperform the standard solvers. Sideways motion seems to be more challenge for the 6- and 7-point solvers. Feature points lie in a plane is a degenerate configuration for the standard methods. Our solvers are focus on this case. Under small (roll, pitch) noise, the proposed solvers are still comparable to the standard algorithms with general scenes (50% of the points lie in a plane). It is practical since good accelerometers today have noise levels of around 0.02 degrees in the computed angles [9]. The only drawback of our solvers may be that the translation part is slightly sensitive to direction noise. The experiments also show that using 4 points to compute the 2-dimensional null space is much better than the minimal 3.5-point case.

7. Real data

To illustrate the usefulness of the proposed solvers, we captured the images and the corresponding gravity vector using an uncalibrated iphone 6S which has a precision of

Seq.	3.5-point	Malis [29]	6-point [22]
1	11.3	43.4	**7.8**
2	**4.9**	38.7	62.1
3	**9.7**	37.2	48.5
4	**6.2**	57.6	68.3
5	**5.1**	44.7	63.6
6	**5.2**	53.7	53.4
7	**5.9**	46.9	60.6
8	**3.2**	16.2	89.3

(a) Equal focal length

3.5-point	6-point [22]
2.1	**0.2**
2.3	50.6
6.6	43.7
1.1	13.0
2.8	69.2
1.6	19.2
1.5	14.5
1.3	74.5

(b) One unknown focal length

4-point	7-point [2]
(3.7, 3.4)	(30.0, 32.8)
(6.5, 4.7)	(55.3, 55.7)
(3.3, 6.3)	(68.3, 65.9)
(3.7, 3.8)	(76.9, 72.0)
(3.3, 5.5)	(73.4, 77.3)
(3.1, 3.2)	(60.6, 61.1)
(2.6, 3.2)	(64.2, 63.8)
(0.6, 1.0)	(70.0, 68.5)

(c) Varying focal lengths

Table 1: Median relative error (%) in the estimated focal lengths on the 8 sequences.

1°. 8 short sequences with more than 150 images were captured and each contains one dominant plane. Example images of the sequences are shown in Figure 8. We extracted SURF [1] feature points and descriptors of all images. The ground truth focal length of our mobile phone is 4.2*mm* (3442 pixels at the resolution of 4032×3024). We randomly sample 2 images C_2^κ times for each sequence (κ is the number of images in this sequence). For the single-side and varying focal lengths problems, we resize the resolution of the second view to half the original size. We use RANSAC with a number of iterations fixed to 1000. As shown in the synthetic experiments, the 5-point solver performs very similar to the 6-point solver. So we compare our solvers with the standard 6-point solver, the single-side 6-point solver, the homography-based sampling method and the 7-point solver. Since the comparisons of rotation and translation have already been shown in the extensive synthetic experiments, we only show the focal length comparison because rotation and translation is usually good once we have a good focal length estimate. Table 1 shows the median error (%) in the estimated focal lengths. As we can see, the proposed solvers perform very well when the scene contains one dominant plane even without calibration of relative rotation of IMU and camera inside the smart phone. The standard solvers fail with most cases when most feature points lie in a plane. Recall that the standard solvers are general methods, and there are also a lot of factors which may influence the performance, e.g., the quality of the plane, motions. The proposed solvers need direction information, which is largely motivated by the availability of smart phones, tablets.

Figure 8: Example images of the 8 sequences.

Fast implementation

Since the uncalibrated solvers need to sample 4 points, for the 3.5pt solvers we can first use the standard 4-point homography algorithm to reject outliers. It is very fast since we only need to calculate the null space of an 8×9 matrix. In addition, the standard 4-point homography algorithm has unique solution. Second, using the 3.5pt solvers to select the solution with most inliers using RANSAC based on the inlier set. The 4 point correspondences which return the most inliers for the standard 4-point homography do not necessarily mean they are the best choice for the 3.5pt solvers. Because the 3.5pt solvers need 2-dimensional null space. So we still need to use RANSAC with the inlier set. Once the inlier set is known, the number of iterations would not be large. In contrast, the 6- and 7-point solvers need to estimate the hypothesis for every RANSAC loop and verify all the possible solutions with more iterations. The experiments were run on a laptop computer with an Intel Core i5-8300H 2.3GHz CPU using Matlab, and only the polynomial solvers were used mex files (based on Eigen linear algebra library [13]) for speed up. The runtime of one hypothesis estimation for the uncalibrated solvers are 1.4*ms*, 0.157*ms*, 0.056*ms*, respectively.

8. Conclusion

In this paper, we propose new minimal solvers for homography-based relative pose estimation with a common direction, among which, the proposed partially calibrated solvers can be used in the structure-from-motion pipelines for uncalibrated smart phones, tablets. They mainly correspond to three real-life applications. (i) scenes captured by a smart device with unknown but fixed focal length, (ii) scenes captured by multiple devices but one of them is calibrated, (iii) scenes captured with a zoom camera or multiple uncalibrated smart devices. We believe that the proposed solvers can be used as alternative solutions to increase the speed and robustness of structure-from-motion systems for smart phones, tablets.

Acknowledgments. This work was supported in part by the Inria/NYU collaboration and the Louis Vuitton/ENS chair on artificial intelligence.

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[2] Sylvain Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pages 790–796. IEEE, 1998.

[3] Martin Bujnak, Zuzana Kukelova, and Tomas Pajdla. 3d reconstruction from image collections with a single known focal length. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1803–1810. IEEE, 2009.

[4] Martin Byröd, Klas Josephson, and Kalle Åström. Fast and stable polynomial equation solving and its application to computer vision. *International Journal of Computer Vision*, 84(3):237–256, 2009.

[5] David A Cox, John Little, and Donal O'shea. *Using algebraic geometry*, volume 185. Springer Science & Business Media, 2006.

[6] Olivier D Faugeras and Francis Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(03):485–508, 1988.

[7] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[8] Friedrich Fraundorfer and Davide Scaramuzza. Visual odometry: Part ii: Matching, robustness, optimization, and applications. *IEEE Robotics & Automation Magazine*, 19(2):78–90, 2012.

[9] Friedrich Fraundorfer, Petri Tanskanen, and Marc Pollefeys. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In *European Conference on Computer Vision*, pages 269–282. Springer, 2010.

[10] Daniel R Grayson and Michael E Stillman. Macaulay 2, a software system for research in algebraic geometry, 2002.

[11] Banglei Guan, Pascal Vasseur, Cédric Demonceaux, and Friedrich Fraundorfer. Visual odometry using a homography formulation with decoupled rotation and translation estimation using minimal solutions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2320–2327. IEEE, 2018.

[12] Banglei Guan, Qifeng Yu, and Friedrich Fraundorfer. Minimal solutions for the rotational alignment of imu-camera systems using homography constraints. *Computer vision and image understanding*, 170:79–91, 2018.

[13] Gaël Guennebaud, Benoît Jacob, Philip Avery, Abraham Bachrach, Sebastien Barthelemy, et al. Eigen v3, 2010.

[14] Richard Hartley and Hongdong Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 34(12):2303–2314, 2012.

[15] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[16] Richard I Hartley. Estimation of relative camera positions for uncalibrated cameras. In *European conference on computer vision*, pages 579–587. Springer, 1992.

[17] Richard I Hartley. In defence of the 8-point algorithm. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 1064–1070. IEEE, 1995.

[18] Fredrik Kahl and Bill Triggs. Critical motions in euclidean structure from motion. In *International Conference on Computer Vision & Pattern Recognition (CVPR'99)*, pages 366–372. IEEE Computer Society, 1999.

[19] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Automatic generator of minimal problem solvers. In *European Conference on Computer Vision*, pages 302–315. Springer, 2008.

[20] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Closed-form solutions to minimal absolute pose problems with known vertical direction. In *Asian Conference on Computer Vision*, pages 216–229. Springer, 2010.

[21] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Polynomial eigenvalue solutions to minimal problems in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1381–1393, 2012.

[22] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomas Pajdla. A clever elimination strategy for efficient minimal solvers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, page 4, 2017.

[23] Viktor Larsson and Kalle Åström. Uncovering symmetries in polynomial systems. In *European Conference on Computer Vision*, pages 252–267. Springer, 2016.

[24] Viktor Larsson, Kalle Åström, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *CVPR*, volume 2, page 4, 2017.

[25] Viktor Larsson, Magnus Oskarsson, Kalle Astrom, Alge Wallis, Zuzana Kukelova, and Tomas Pajdla. Beyond grobner bases: Basis selection for minimal solvers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3945–3954, 2018.

[26] Karel Lebeda, Jırı Matas, and Ondrej Chum. Fixing the locally optimized ransac–full experimental evaluation. In *British machine vision conference*, pages 1–11. Citeseer, 2012.

[27] Hongdong Li. A simple solution to the six-point two-view focal-length problem. In *European Conference on Computer Vision*, pages 200–213. Springer, 2006.

[28] Hongdong Li and Richard Hartley. Five-point motion estimation made easy. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 630–633. IEEE, 2006.

[29] Ezio Malis and Roberto Cipolla. Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1268–1272, 2002.

[30] Ezio Malis, Tarek Hamel, Robert Mahony, and Pascal Morin. Dynamic estimation of homography transformations on the special linear group for visual servo control. In *2009 IEEE*

international conference on robotics and automation, pages 1498–1503. IEEE, 2009.

[31] Ezio Malis and Manuel Vargas. *Deeper understanding of the homography decomposition for vision-based control*. PhD thesis, INRIA, 2007.

[32] Oleg Naroditsky, Xun S Zhou, Jean Gallier, Stergios I Roumeliotis, and Kostas Daniilidis. Two efficient solutions for visual odometry using directional correspondence. *IEEE transactions on pattern analysis and machine intelligence*, 34(4):818–824, 2012.

[33] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):756–770, 2004.

[34] David Nistér and Frederik Schaffalitzky. Four points in two or three calibrated views: Theory and practice. *International Journal of Computer Vision*, 67(2):211–231, 2006.

[35] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: a universal framework for random sample consensus. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):2022–2038, 2013.

[36] Olivier Saurer, Pascal Vasseur, Rémi Boutteau, Cédric Demonceaux, Marc Pollefeys, and Friedrich Fraundorfer. Homography based egomotion estimation with a common direction. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):327–341, 2017.

[37] Olivier Saurer, Pascal Vasseur, Cedric Demonceaux, and Friedrich Fraundorfer. A homography formulation to the 3pt plus a common direction relative pose problem. In *Asian Conference on Computer Vision*, pages 288–301. Springer, 2014.

[38] Henrik Stewénius, David Nistér, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 789–794. IEEE, 2005.

[39] Chris Sweeney, John Flynn, and Matthew Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. *3DV*, 2(4):5, 2014.

[40] Akihiko Torii, Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. The six point algorithm revisited. In *Asian Conference on Computer Vision*, pages 184–193. Springer, 2010.

[41] Thierry Viéville, Emmanuelle Clergue, and PE Dos Santos Facao. Computation of ego-motion and structure from visual and inertial sensors using the vertical cue. In *Computer Vision, 1993. Proceedings., Fourth International Conference on*, pages 591–598. IEEE, 1993.