Five-point Fundamental Matrix Estimation for Uncalibrated Cameras

Daniel Barath[12]
[1] Machine Perception Research Laboratory, MTA SZTAKI, Budapest, Hungary
[2] Centre for Machine Perception, Czech Technical University, Prague, Czech Republic

Abstract

We aim at estimating the fundamental matrix in two views from five correspondences of rotation invariant features obtained by e.g. the SIFT detector. The proposed minimal solver[1] first estimates a homography from three correspondences assuming that they are co-planar and exploiting their rotational components. Then the fundamental matrix is obtained from the homography and two additional point pairs in general position. The proposed approach, combined with robust estimators like Graph-Cut RANSAC, is superior to other state-of-the-art algorithms both in terms of accuracy and number of iterations required. This is validated on synthesized data and 561 *real image pairs. Moreover, the tests show that requiring three points on a plane is not too restrictive in urban environment and locally optimized robust estimators lead to accurate estimates even if the points are not entirely co-planar. As a potential application, we show that using the proposed method makes two-view multi-motion estimation more accurate.*

1. Introduction

This paper investigates the problem of estimating the relative motion of two *non-calibrated cameras* from rotational invariant features. In particular, we are interested in the minimal case, i.e. to estimate fundamental matrix $\mathbf{F} \in \mathbb{R}^{3\times3}$ exploiting *five* point correspondences together with rotational components obtained by, e.g. SIFT detector [16]. The method requires three points to be co-planar and two additional ones in arbitrary position (see Fig. 1).

The classical way of estimating $\mathbf{F}$ for non-calibrated cameras is to apply the eight- or seven-point algorithms [10]. They are both widely-used in the literature and fundamental tools of computer vision applications. The eight-point algorithm estimates the direct linear transformation induced by the epipolar constraint. The seven-point algorithm enforces the rank-two constraint by solving the cubic polynomial equation which it implies. From theoretical point of view, getting *more information exclusively from point correspondences is not possible*. However, of course, there are approaches to reduce the number of unknowns. For example, knowing the intrinsic parameters of the cameras (i.e. the principal point, focal length, pixel ratios) enables to enforce the trace constraint. The problem becomes solvable using six point pairs [14, 13, 23, 25] if all intrinsics parameters but a common focal length are known, or five correspondences [19, 15, 5, 13, 9] are enough for fully calibrated cameras. One can also restrict the camera movement, e.g. the one point method proposed by Davide Scaramuzza [22] assumes the cameras to move on a plane and the so-called non-holonomic constraint to hold.

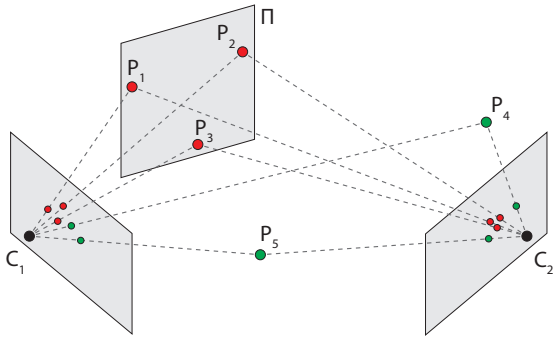


Figure 1: The proposed minimal solver estimates a fundamental matrix between views C_1 and C_2. It first estimates a homography from three correspondences of co-planar points ($\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$) lying on plane π. The fundamental matrix is then obtained from the homography and two additional points ($\mathbf{P}_4$ and $\mathbf{P}_5$) in general position.

By looking the other way, it is very rare nowadays to get solely the point coordinates from the applied feature detector. As an example, the widely-used SIFT detector provides a rotation and scale besides the coordinates. This additional information is rarely exploited in state-of-the-art geometric model estimators and just thrown away at the very beginning. *This information is available* in most of the cases. In this paper, we aim at involving these additional *affine parameters*, e.g. rotation of the feature, into the process to

[1] Available at http://web.eee.sztaki.hu/ dbarath/

reduce the size of the minimal sample required for fundamental matrix estimation.

Exploiting full affine correspondences (point correspondence, rotation, scales along both image axes and shear) for fundamental or essential matrix estimation, of course, is not a new idea. Perdoch et al. [20] proposed techniques for approximating the relative camera motion using two and three correspondences. Bentolila and Francos [6] proposed a method to estimate the exact, i.e. with no approximation, $\mathbf{F}$ from three correspondences. Raposo et al. [21] proposed a solution for direct essential matrix estimation using two correspondences. Using only a part of an affine correspondence, e.g. exclusively the rotation component, is a well-known technique for example in wide-baseline feature matching [17]. However, to the best of our knowledge, the only work involving them into geometric model estimation is that of Barath et al. [1]. In [1], $\mathbf{F}$ is assumed to be known a priori and a technique is proposed for estimating a homography using two SIFT correspondences exploiting their scale and rotation components. Even so, an assumption is made, considering that the scales along axes u and v equal to that of the SIFT features – which is generally not true in practice. Thus, the method yields only an approximation.

The contributions of the paper are: (i) we propose a technique for estimating homography $\mathbf{H}$ using three rotation invariant feature correspondences. To recover $\mathbf{H}$, in addition to the point coordinates, the rotations of the features are exploited. (ii) The recovered homography is then used to calculate fundamental matrix $\mathbf{F}$ using two additional correspondences. (iii) It is reported on both synthesized and real worlds tests, that combining the proposed method with a robust estimator, e.g. LO-RANSAC [7], leads to results superior to the state-of-the-art in term of accuracy and the number of iterations required. Moreover, we demonstrate that using the proposed method in two-view multi-motion fitting is beneficial and leads to more accurate clusterings.

2. Theoretical Background

Affine Correspondences. In this paper, we consider an affine correspondence (AC) as a triplet: $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{A})$, where $\mathbf{p}_1 = [u_1 \quad v_1 \quad 1]^\mathrm{T}$ and $\mathbf{p}_2 = [u_2 \quad v_2 \quad 1]^\mathrm{T}$ are a corresponding homogeneous point pair in the two images (the projections of the 3D points in Fig. 1), and $\mathbf{A}$ is a 2×2 linear transformation which we call *local affine transformation*. To define $\mathbf{A}$, we use the definition provided in [18] as it is given as the first-order Taylor-approximation of the 3D $\rightarrow$ 2D projection functions. Note that, for perspective cameras, $\mathbf{A}$ is the first-order approximation of the related 3×3 *homography* matrix $\mathbf{H}$ as follows:

$$\begin{array}{rclcrcl} a_1 & = & \frac{\partial u_2}{\partial u_1} = \frac{h_1 - h_7 u_2}{s}, & & a_2 & = & \frac{\partial u_2}{\partial v_1} = \frac{h_2 - h_8 u_2}{s}, \\ a_3 & = & \frac{\partial v_2}{\partial u_1} = \frac{h_4 - h_7 v_2}{s}, & & a_4 & = & \frac{\partial v_2}{\partial v_1} = \frac{h_5 - h_8 v_2}{s}, \end{array} \quad (1)$$

where u_i and v_i are the coordinates in the ith image ($i \in \{1, 2\}$), h_j is the jth element of $\mathbf{H}$ in row-major order ($j \in [1, 9]$) and $s = u_1 h_7 + v_1 h_8 + h_9$ is the projective depth.

Fundamental matrix $\mathbf{F}$ is a 3×3 transformation matrix ensuring the so-called epipolar constraint $\mathbf{p}_2^\mathrm{T} \mathbf{F} \mathbf{p}_1 = 0$ for rigid scenes. Since its scale is arbitrary and $\det(\mathbf{F}) = 0$, $\mathbf{F}$ has seven degrees-of-freedom (DoF). Its elements are denoted by f_i ($i \in [1, 9]$) in a row-major order. These properties will help us to recover the fundamental matrix from five rotation invariant feature correspondences.

3. Homography from Three Correspondences

In this section, it is shown how a homography can be estimated from three rotation invariant feature correspondences. First, we show the relationship of homographies and affine correspondences. Then this is decomposed into affine components establishing the way to exploit them independently. Selecting the appropriate equations from the obtained system, we finally use the given rotations to get the homography parameters.

3.1. Homographies and Affine Correspondences

To form a linear equation system using $\mathbf{A}$, Eqs. 1 are multiplied by the common denominator (s – projective depth), then rearranged as follows:

$$\begin{array}{rcl} h_1 - (u_2 + a_1 u_1) h_7 - a_1 v_1 h_8 - a_1 & = & 0 \\ h_2 - (u_2 + a_2 v_1) h_8 - a_2 u_1 h_8 - a_2 & = & 0 \\ h_4 - (v_2 + a_3 u_1) h_7 - a_3 v_1 h_8 - a_3 & = & 0 \\ h_5 - (v_2 + a_4 v_1) h_8 - a_4 u_1 h_8 - a_4 & = & 0 \end{array} \quad (2)$$

These equations encode the connection of a local affine transformation and a homography.

As it is well-known, the relationship of a homography and a point correspondence $\mathbf{H}\mathbf{p}_1 \sim \mathbf{p}_2$ can be interpreted as an inhomogeneous linear system of equations. Note that operator $\sim$ means "equality up to an arbitrary scale". The system is as follows:

$$\begin{array}{rcl} u_1 h_1 + v_1 h_2 + h_3 - u_1 u_2 h_7 - v_1 u_2 h_8 & = & u_2 \\ u_1 h_4 + v_1 h_5 + h_6 - u_1 v_2 h_7 - v_1 v_2 h_8 & = & v_2 \end{array} \quad (3)$$

Combining Eqs. 2 and 3, an affine correspondence yields six linear equations on total. Thus each of them reduces the DoF of homography estimation by six.

Affine Transformation Model. Although the relationship of full affine correspondences and homographies are well-defined, the current problem is the exploitation of features containing only a part of $\mathbf{A}$ – the rotation. Therefore,

let us define an affine transformation model as a combination of linear transformations as follows:

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} s_u & w \\ 0 & s_v \end{bmatrix} = \begin{bmatrix} s_u \cos(\alpha) & w\cos(\alpha) - s_v \sin(\alpha) \\ s_u \sin(\alpha) & w\sin(\alpha) + s_v \cos(\alpha) \end{bmatrix}, \quad (4)$$

where α, s_u, s_v, and w are the rotational angle, scales along axes u and v, and shear parameter, respectively.

Substituting the components of the matrix defined in Eqs. 4 into Eqs. 2, the following system is given:

$$\begin{aligned} h_1 - u_2 h_7 - u_1 c_\alpha s_u h_7 - v_1 c_\alpha s_u h_8 - c_\alpha s_u &= 0, \\ h_2 - u_2 h_8 + v_1 c_\alpha w h_8 - v_1 s_\alpha s_v h_8 - & \\ u_1 c_\alpha w h_8 + u_1 s_\alpha s_v h_8 - c_\alpha w + s_\alpha s_v &= 0, \\ h_4 - v_2 h_7 - u_1 s_\alpha s_u h_7 - v_1 s_\alpha s_u h_8 - s_\alpha s_u &= 0, \\ h_5 - v_2 h_8 - v_1 s_\alpha w h_8 - v_1 c_\alpha s_v h_8 - & \\ u_1 s_\alpha w h_8 - u_1 c_\alpha s_v h_8 - s_\alpha w - c_\alpha s_v &= 0, \end{aligned} \quad (5)$$

where $c_\alpha = \cos(\alpha)$ and $s_\alpha = \sin(\alpha)$. Note that this system shows the general way of the affine parameters affecting the related homography. Even though we will consider exclusively α to be known in the subsequent sections, one can easily exploit these equations to solve for different features containing e.g. scales or shear besides the rotation.

3.2. Homography Estimation

Assume three co-planar point correspondences $\mathbf{p}_{1,i} = [u_{1,i} \quad v_{1,i} \quad 1]^\mathrm{T}$, $\mathbf{p}_{2,i} = [u_{2,i} \quad v_{2,i} \quad 1]^\mathrm{T}$ ($i \in [1, 3]$) and the related rotation components α_i, obtained by e.g. SIFT, to be known. The objective is to find homography $\mathbf{H}$ for which $\mathbf{H}\mathbf{p}_{1,i} \sim \mathbf{p}_{2,i}$ and also satisfies Eqs. 5.

In the first part of the algorithm, only the coordinates are used to reduce the number of unknown parameters. We form $\mathbf{H}\mathbf{p}_{1,i} \sim \mathbf{p}_{2,i}$ (Eq. 3) for all correspondences as a homogeneous linear system $\mathbf{B}\mathbf{h} = 0$. Since each point pair yields two equations for the nine unknowns, coefficient matrix $\mathbf{B}$ is of size 6×9 and $\mathbf{h} = [h_1 \ h_2 \ h_3 \ h_4 \ h_5 \ h_6 \ h_7 \ h_8 \ h_9]^\mathrm{T}$ is the vector of unknown parameters. The null-space of $\mathbf{B}$ is three-dimensional, therefore the final solution is calculated as a linear combination of the three null-vectors as follows:

$$\mathbf{h} = \beta\mathbf{b} + \gamma\mathbf{c} + \delta\mathbf{d}, \quad (6)$$

where $\mathbf{b} = [b_1 \ ... \ b_9]^\mathrm{T}$, $\mathbf{c} = [c_1 \ ... \ c_9]^\mathrm{T}$ and $\mathbf{d} = [d_1 \ ... \ d_9]^\mathrm{T}$ are the null-vectors, and β, γ, δ are unknown scalars. Due to the scale ambiguity of $\mathbf{H}$ one of them can be set to an arbitrary value, thus in our algorithm, $\delta = 1$.

Remember, that three rotation components are given, each providing four equations and three unknowns via Eqs. 5. Two rotations yield eight equations and six unknowns, therefore, they are enough for estimating β and γ. To exploit them, Eqs. 6 have to be substituted into Eqs. 5 replacing each h_j by $\beta b_j + \gamma c_j + d_j$ ($j \in [1, 9]$). Since the scale along axis v and shear w are not known, the 2nd and 4th equations of Eqs. 5 yield no additional information, they are removed from the system. Without them, the two rotations lead to a multivariate polynomial system consisting of four equations with monomials $[\beta \quad \gamma \quad s_{u,1} \quad s_{u,1}\beta \quad s_{u,1}\gamma \quad s_{u,2} \quad s_{u,2}\beta \quad s_{u,2}\gamma]^\mathrm{T}$. Coefficient matrix $\mathbf{C}$ is visualized in Table 1. Since four equations are given for four unknowns ($s_{u,1}$, $s_{u,2}$, β, and γ), and there are no higher order monomials, the system can straightforwardly be rearranged, then solved. The final formulas for β and γ are shown in Appendix A. Finally, homography $\mathbf{H}$ is recovered through Eq. 6.

β	γ	$s_{u,1}$	$s_{u,1}\beta$	$s_{u,1}\gamma$	$s_{u,2}$	$s_{u,2}\beta$	$s_{u,2}\gamma$
c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	c_{17}	c_{18}
				...			
c_{41}	c_{42}	c_{43}	c_{44}	c_{45}	c_{46}	c_{47}	c_{48}

Table 1: Homography estimation. Coefficient matrix $\mathbf{C}$ of the multivariate polynomial system to which the rotation components lead. Each column represents the coefficients of a monomial (1st row) in the four equations (rows).

Note that assuming that close points more likely belong to the same homography, we choose the rotations of the two closest points. Although this is a heuristics, it worked well in our experiments and does not require much computation. For problems, where the time is not critical, it is a possible choice to estimate the three homographies which the three rotations induce and select the one with the most inliers. Also note that all minimal samples, i.e. the selected five correspondences, can be rejected for which the two points in general positions also lie on the plane, thus leading to degenerate configuration. This can be checked by simply thresholding the re-projection error implied by $\mathbf{H}$ and each point pair.

4. Fundamental Matrix Estimation from Five Correspondences

Suppose that homography $\mathbf{H}$, estimated in the previously described way, and two additional point correspondences are given. The objective is to estimate fundamental matrix $\mathbf{F}$ compatible both with $\mathbf{H}$ and the two correspondences and $\det(\mathbf{F}) = 0$ holds. The compatibility with $\mathbf{H}$ could be ensured through the well-known formula [10]: $\mathbf{H}^\mathrm{T}\mathbf{F} + \mathbf{F}^\mathrm{T}\mathbf{H} = 0$. However, the *direct linear method* solving this system is unstable for inaccurate homographies, sometimes leading to completely meaningless results. The reason is that the samples are far from the normal distribution required for least squares fitting to work reasonably well [24]. Zhou et al. [26] proposed a normalization tech-

nique solving this problem, even so, this method needs at least three homographies to be known and do not consider the case when additional correspondences are given. Thus we chose the *hallucinated point* technique generating five point correspondences using $\mathbf{H}$. The five generated and two given point pairs yield seven linear equations through $\mathbf{p}_{2,i}^{\mathrm{T}}\mathbf{F}\mathbf{p}_{1,i} = 0$ ($i \in [1, 7]$). Combining them, the following homogeneous linear system is given: $\mathbf{Df} = 0$, where $\mathbf{D}$ is the coefficient matrix and $\mathbf{f} = [f_1\ f_2\ f_3\ f_4\ f_5\ f_6\ f_7\ f_8\ f_9]^{\mathrm{T}}$ is the vector of unknown parameters. Matrix $\mathbf{D}$ is as

$$\mathbf{D} = \begin{bmatrix} u_{1,1}u_{2,1} & v_{1,1}u_{2,1} & u_{2,1} & u_{1,1}v_{2,1} & v_{1,1}v_{2,1} & v_{2,1} & u_{1,1} & v_{1,1} & 1 \\ & & & & \dots & & & & \\ u_{1,7}u_{2,7} & v_{1,7}u_{2,7} & u_{2,7} & u_{1,7}v_{2,7} & v_{1,7}v_{2,7} & v_{2,7} & u_{1,7} & v_{1,7} & 1 \end{bmatrix}.$$

Note that making the estimator more stable, the normalization proposed by Hartley [11] is applied and the equations from the three co-planar points are also added. The null-space of matrix $\mathbf{D}$ is two-dimensional and the solution is calculated as the linear combination of the two null-vectors:

$$\mathbf{F} = \epsilon\mathbf{e} + \eta\mathbf{g}, \tag{7}$$

where ϵ and η are unknown scalars, $\mathbf{e} = [e_1\ \dots\ e_9]^{\mathrm{T}}$ and $\mathbf{g} = [g_1\ \dots\ g_9]^{\mathrm{T}}$ are the null-vectors. Due to the scale ambiguity of $\mathbf{F}$, η can be set to an arbitrary value. To achieve stability we use $\eta = 1 - \epsilon$, thus keeping the sum of the weights to be one. Substituting Eq. 7 into $\det(\mathbf{F}) = 0$ leads to a cubic polynomial equation. The possible solutions for ϵ (their number is $\in \{1, 2, 3\}$, similarly to the seven-point algorithm) are obtained as the real roots of the polynomial. The resulting fundamental matrices are finally calculated by substituting each ϵ to Eq. 7. Note that all fundamental matrices are discarded for which the *oriented* epipolar constraint [8] does not hold.

Concluding the current and the previous sections, fundamental matrix $\mathbf{F}$ can be estimated from three co-planar and two arbitrary correspondences of rotation invariant features.

5. Experimental Results

In this section, we compare the proposed method with the widely used seven- and eight-point algorithms [10] both on synthesized and real worlds tests.

5.1. Synthesized Tests

For synthesized testing, two perspective cameras were generated by their projection matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{3\times4}$ and five random planes were sampled, each at four locations. The generated 20 points were then projected into the cameras and the ground truth affine transformations were computed from the image points and plane parameters. Zero-mean Gaussian-noise were added to the point coordinates, thus contaminating the affine components as well.

Noise σ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
RU	0.0	0.3	0.6	0.9	1.2	1.6	1.9	2.3	2.8	3.1	3.6
UR	0.0	0.3	0.6	0.9	1.2	1.6	1.9	2.3	2.8	3.1	3.6

Table 2: Re-projection error of the estimated homographies using **RU** and **UR** decompositions.

Fig. 2 shows the results of the proposed, eight- and seven-points algorithms applied to view pairs with specific camera motions (left – random motion, middle – pure sideways motion, right – pure forward motion). The error is plotted as the function of the noise σ (horizontal axis; in pixels). It is the mean symmetric epipolar distance from the correspondences not used for the estimation. For random motion, both cameras were located at a random point of a 10-radius sphere and look towards the origin. For sideways and forward motions, the distance of the cameras was 10 unit and a small perturbation, i.e. zero-mean Gaussian-noise with 0.1 standard deviation, was added to the coordinates.

It can be seen, that the proposed method leads similar accuracy to the seven-point algorithm for general movement. However, for purely sideways motion, the method is significantly less sensitive to the noise than the other competitors. For forward motion, if the noise σ does not exceed 0.5, the five-point technique is most accurate. After that point, the seven-point algorithm outperforms it.

Decompositions. In this paper, we chose to decompose $\mathbf{A}$ to **RU** where $\mathbf{R}$ is a 2D rotation by α degrees and $\mathbf{U}$ is an upper-triangle matrix applying the shear and scales along the image axes. It can nevertheless be decomposed in other ways as well, for instance, as **UR** instead of **RU**. Table **??** shows the re-projection error of the estimated homographies using these decompositions. They lead to identical results.

Homography estimation. We compare the proposed homography estimation with normalized DLT (Direct Linear Transform) and HA (Homography from Affine transformation) methods. DLT [10] solves a linear system, induced by formula $\mathbf{Hp}_1 \sim \mathbf{p}_2$, if at least four point correspondences are given. HA [2] estimates the homography from two ACs. Reflecting the fact that only angle α and scale s are given for SIFT correspondences, we approximated each affine transformation as $\mathbf{A} \approx \mathbf{R}_\alpha \mathrm{diag}(s, s)$, where $\mathbf{R}_\alpha$ is a 2D rotation matrix rotating by α degrees and $\mathrm{diag}(s, s)$ is a 2×2 diagonal matrix containing the SIFT scale. Note that due to this rough approximation, the error of HA is not zero even in the noise-free case. The left plot of Fig. 3 shows the re-projection error (in pixels; vertical axis) plotted as the function of the noise σ (in pixels; horizontal). Due to the approximation, HA is very sensitive to the noise, and thus not applicable to real world problems if not the full affine

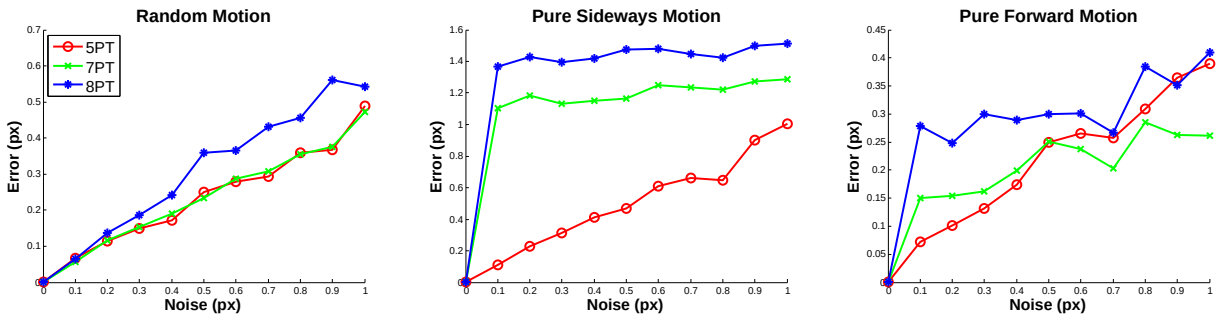


Figure 2: The mean error (in pixels; plotted as the function of the noise σ) of the proposed, seven- and eight-point algorithms on cameras motions: random (left), sideways (middle) and forward (right). For random motion, both cameras are placed at a random point of a 10-radius sphere and look towards the origin. For sideways and forward motions, the distance of the cameras was 10 unit and a small zero-mean Gaussian-noise (with standard deviation set to 0.1) is added to each coordinate.

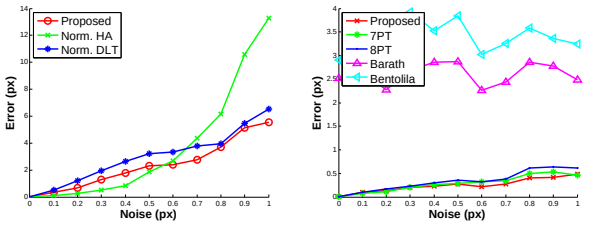


Figure 3: *(Left)* Comparison of the proposed homography estimation with normalized HA [2] and DLT methods. *(Right)* Comparison of the 5PT method with point (7PT, 8PT) and the affine correspondence-based **F** estimators of Barath et al. [4] and Bentolila et al. [6].

correspondences are known. The proposed homography estimation slightly outperforms normalized DLT.

AC-based methods. Techniques exploiting affine correspondences are not applicable to the current problem, i.e. when partially affine invariant features are given, due to the roughness of the approximation of **A**. The right plot of Fig. 3 shows the comparison of the five-, seven- and eight-point algorithms with the methods of Barath et al. [4] and Bentolila et al. [6]. Even though [4] estimates **F** and a common focal length, the linear relationship which they proposed can be straightforwardly modified to solve **F** from three affine correspondences. Bentolila et al. [6] obtains **F** from three ACs using conic constraints. Both methods got the approximated affinities, i.e. $\mathbf{A} \approx \mathbf{R}_\alpha \text{diag}(s, s)$, as input. The figure reports the mean symmetric epipolar error (in pixels; vertical axis) of the estimated fundamental matrices plotted as the function of the noise σ (in pixels; horizontal). For the proposed, 7PT and 8PT algorithms, the same trend can observed as in the previous test cases. It can also be seen that the approximation of **A** is too rough for the AC-based method.

5.2. Real World Tests

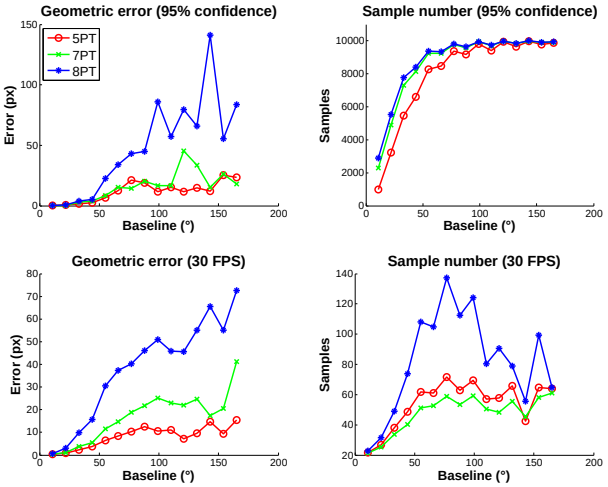


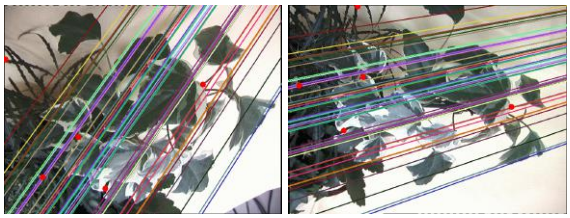
Figure 4: The mean error (left; in pixels) and sample number (right) plotted as the function of the baseline (in degrees; rotation around the object) for confidence 99% (top) and time limit $1/30$ secs (bottom). Results are computed from 100 runs on each image pair (#515) in the `Strecha` dataset.

To test the proposed method on real world data, we used the `AdelaideRMF`[2], `Kusvod2`[3], `Multi-H`[4], and `Strecha`[5] datasets (see Fig. 5 for examples). `AdelaideRMF`, `Kusvod2` and `Multi-H` consist of image pairs of resolution from 455×341 to 2592×1944 and manually annotated (assigned to outlier or inlier classes) correspondences. Since the reference points do not contain rotation components we detected and matched points applying SIFT detector.

`Strecha` dataset consists of image sequences (each im-

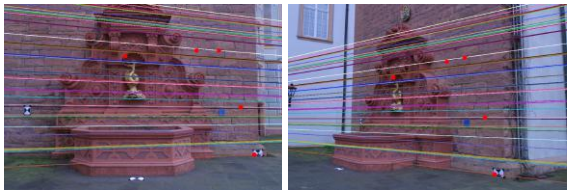
[2] cs.adelaide.edu.au/ hwong/doku.php?id=data
[3] cmp.felk.cvut.cz/data/geometry2view
[4] web.eee.sztaki.hu/ dbarath
[5] cvlab.epfl.ch/data/strechamvs

(a) AdelaideRMF



(b) Kusvod2

(c) Multi-H



(d) Strecha

Figure 5: The results of the proposed method combined with Graph-Cut RANSAC. An image pair from each dataset with the corresponding epipolar lines of 50 random inliers drawn by colors. The five point pairs which are used as the minimal sample are visualized by red dots.

age is of size 3072×2048) and a projection matrix for every image. Therefore, we paired the images in each sequence in every possible way. The ground truth $\mathbf{F}$ was estimated from the projection matrices [10] and SIFT was used to get correspondences. Every detected point pair was considered as a reference point for which the symmetric epipolar distance [10] from the ground truth $\mathbf{F}$ was smaller than 1.0 pixels. If less then 20 reference points were kept, the pair was not used in the latter evaluation.

We chose Graph-Cut RANSAC [3] as a robust estimator since it can be considered as state-of-the-art and its source code is publicly available[6]. In brief, it is a locally optimized RANSAC using graph-cut to achieve efficiency and global optimality w.r.t. the current so-far-the-best model. Validating the estimated fundamental matrices, we used the reference point sets. The geometric error was computed as the mean symmetric epipolar distance. The competitor methods, i.e. the minimal solvers combined with GC-RANSAC, were the normalized eight- and seven-point algorithms[7]. In the LSQ re-fitting step of GC-RANSAC, the normalized eight-point method was applied using the current inlier set.

Blocks (a–f) of Table 3 reports the mean result of 100 runs on each pair from the Strecha dataset. The first column is the name of the sequence, the second is the number of the image pairs – the ones having more than 20 reference points. The next two blocks, each consisting of three columns, show the results of the methods if the confidence is set to 99% (1st block) and for a strict 30 FPS time limit (interrupted after $1/30$ secs; 2nd block). The reported properties are the geometric error of the estimated fundamental matrices w.r.t. the reference point sets, and the number of the samples drawn by GC-RANSAC. It can be seen that using the proposed method leads to *more accurate model estimates using less samples* than the competitor algorithms. However, this test is slightly unfair since Strecha consists of images of buildings with planar facades. Thus finding three co-planar points is not challenging. Blocks (g–i) show the mean results on AdelaideRMF, Kusvod2 and Multi-H datasets (1st col) if the confidence is set to 99% (4th – 6th) and for a strict $1/30$ seconds time limit (7th – 9th). It can be seen that for both cases, the proposed method achieved the lowest mean errors in all but one test cases.

Fig. 4 shows the error (in pixels) and the sample number plotted as the function of the baseline (in degrees). The results are the mean of 100 runs on each image pair, #515 on total, of the Strecha dataset. Since the cameras in the sequences move around a building with approx. 180°, the baseline is indicated by the current angle.

Fig. 5 shows example image pairs from each dataset with the epipolar lines of 50 random inliers and five correspondences used as a minimal sample in the proposed method (red dots). It can be seen, that the results seem good: the epipolar lines go through the same pixels in the first (left) and second (right) images. Pairs (a) and (b) show an interesting effect: there are no entirely co-planar three points. Nevertheless, the initially estimated fundamental matrix was precise enough to be accurately refined by the local optimization step of GC-RANSAC.

5.3. Application: Rigid Motion Segmentation

In this section, we show an possible application where estimating a fundamental matrix using fewer points than the state-of-the-art is beneficial. Multiple rigid motions in two views can be interpreted as a set of fundamental matrices. Typically, they are estimated by applying a multi-

[6] https://github.com/danini/graph-cut-ransac

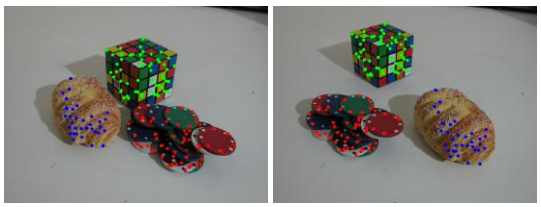
[7] OpenCV implementation.

			Confidence 99%			30 FPS		
Minimal methods →			**5**	7	8	**5**	7	8
(a)	53	Avg Err (px)	**3.06**	4.34	16.21	**4.31**	7.29	17.15
		Samples	**3 692**	5 084	5 471	42	**38**	59
(b)	45	Avg Err (px)	**1.42**	1.63	3.10	**2.33**	3.93	8.95
		Samples	**4 953**	6 621	7 045	40	**36**	57
(c)	81	Avg Err (px)	**6.71**	9.52	20.54	**6.80**	10.75	23.92
		Samples	**6 450**	7 394	7 586	30	**29**	33
(d)	196	Avg Err (px)	**5.40**	8.71	20.51	**6.78**	8.82	19.01
		Samples	**6 720**	7 780	8 094	49	**42**	82
(e)	26	Avg Err (px)	**2.86**	6.08	19.85	7.36	**6.54**	19.38
		Samples	**5 432**	6 545	7 088	45	**40**	74
(f)	114	Avg Err (px)	**4.84**	9.14	16.21	**7.69**	10.06	27.83
		Samples	**5 881**	7 100	7 434	58	**47**	103
(g)	18	Avg Err (px)	0.63	**0.52**	0.53	0.70	**0.56**	0.59
		Samples	**523**	1 178	1 656	**153**	232	413
(h)	24	Avg Err (px)	**6.11**	6.93	9.08	**7.44**	7.55	10.94
		Samples	**1 353**	2 273	2 859	**100**	182	285
(i)	4	Avg Err (px)	**0.34**	0.37	0.38	**0.79**	0.97	5.46
		Samples	**1 985**	3 299	4 991	42	**33**	68
(all)	561	Avg Err (px)	**3.47**	7.41	16.53	**4.90**	8.33	19.51
		Samples	**5 560**	6 276	7 055	**52**	**52**	93

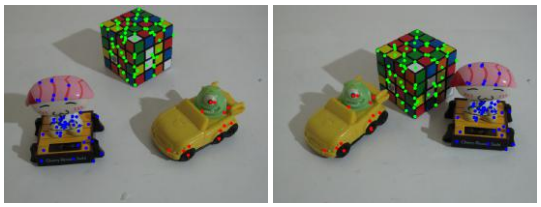
Table 3: Fundamental matrix estimation using GC-RANSAC [3] with minimal methods (2nd row) applied to the sequences of the `Strecha` dataset. The 1st column shows the sequences: (a) Fountain-P11, (b) `Entry-p10`, (c) `Castle-p19`, (d) `Castle-p30`, (e) Herzjesus-p8, and (f) `Herzjesus-p25`, (g) `Kusvod2`, (h) `AdelaideRMF`, and (i) `Multi-H`. The number of the image pairs and the tested properties are reported in the 2nd and 3rd columns. The next three report the results at 99% confidence. For the remaining columns, there was a time limit set to 30 FPS, i.e. the run is interrupted after $1/30$ secs. Values are the means of 100 runs. The mean geometric error (in pixels) of the results w.r.t. the manually annotated inliers are written in each 1st row; the required number of samples are reported in every 2th row. The error is the symmetric epipolar distance.

model fitting algorithm like PEARL [12]. State-of-the-art fitting algorithms generate a set of initial fundamental matrices using a RANSAC-like sampling combined with a minimal method. Then an optimization is applied assigning the points to motion clusters and selecting the motions best interpreting the scene.

The methods were evaluated on the `AdelaideRMF` motion dataset (see Fig. 6 for examples) consisting of 18 image pairs and the ground truth – correspondences assigned to their motion clusters or outlier class. Table 4 reports the result of PEARL combined with minimal methods (rows). The error is the misclassification error, which is the ratio of the points not assigned to the desired motion cluster. PEARL used the same initial model number for all methods, i.e. twice the input point number. The inlier-outlier threshold was tuned for each problem and each method separately. It can be seen that by using the five-point algorithm, the *obtained clusterings are the most accurate*.



(a) breadcubechips



(b) toycubecar

Figure 6: Example results of PEARL [12] combined with the proposed algorithm applied to the `AdelaideRMF` motion dataset. Colors denote motions, black dots are outliers.

5.4. Processing Time

The proposed method consists of three main steps: (i) the null-space computation of a matrix of size 6×9, then the homography parameters are calculated in closed form. (ii) Using the estimated $\mathbf{H}$ and two additional correspondences, a coefficient matrix of size 7×9 is built and its null-space is computed. (iii) Finally, the roots of a cubic polynomial are estimated. The average processing time of 100 runs of our

	5 PT	7 PT	8 PT
Avg	**4.5**	4.9	**4.5**
Med	**2.7**	3.8	3.6

Table 4: Mean and median misclassification error of PEARL combined with minimal methods (2th – 4th cols) on the AdelaideRMF motion dataset.

C++ implementation using OpenCV was 0.16 milliseconds.

Combining RANSAC-like *hypothesize-and-verify* robust estimators with the proposed method is beneficial since their processing time highly depends on the size of the minimal sample required. For instance, the theoretical iteration number of RANSAC for outlier ratio 0.95 and confidence 0.95 is $\approx 10^7$ if five and $\approx 10^9$ if seven correspondences are needed for the estimation.

6. Conclusion

In this paper, we proposed a method for estimating the fundamental matrix between two non-calibrated cameras from five correspondences of rotation invariant features. Three of the points have to be co-planar and two of them be in general position. The solver, combined with Graph-Cut RANSAC, was superior to the seven- and eight-point algorithms both in terms of accuracy and needed sample number on the evaluated 561 publicly available real image pairs. It is demonstrated that the co-planarity of three points is not a too restrictive constraint in real world (e.g. in urban environment) and can be weakened by state-of-the-art robust estimators. Moreover, we showed that the method makes multi-motion fitting more accurate than using the eight- or seven-point algorithms.

Acknowledgement

The project was supported by ÚNKP-17-3 new national excellence program of the ministry of human capacities and the Hungarian National Research, Development and Innovation Office grant VKSZ 14-1-2015-0072.

A. Calculation of the Homography Parameters

In this section, we show how parameters β and γ in Eqs. 6 are calculated. Replacing each h_j with $\beta b_j + \gamma c_j + d_j$ ($j \in [1, 9]$) in the 1st and 3rd equations of Eqs. 5 leads to the following system:

$$\begin{aligned}(\beta b_1 + \gamma c_1 + d_1) - u_2(\beta b_7 + \gamma c_7 + d_7) - \\ u_1 c_\alpha s_u(\beta b_7 + \gamma c_7 + d_7) - \\ v_1 c_\alpha s_u(\beta b_8 + \gamma c_8 + d_8) - c_\alpha s_u = 0, \\ (\beta b_4 + \gamma c_4 + d_4) - v_2(\beta b_7 + \gamma c_7 + d_7) - \\ u_1 s_\alpha s_u(\beta b_7 + \gamma c_7 + d_7) - \\ v_1 s_\alpha s_u(\beta b_8 + \gamma c_8 + d_8) - s_\alpha s_u = 0.\end{aligned}$$

After expanding and rearranging the expressions, the first equation becomes

$$\begin{aligned}(b_1 - u_2 b_7)\beta + (c_1 - u_2 c_7)\gamma - (u_1 c_\alpha b_7 + v_1 c_\alpha b_8)s_u\beta - \\ (u_1 c_\alpha d_7 + v_1 c_\alpha d_8 + c_\alpha)s_u + (u_1 c_\alpha c_7 + v_1 c_\alpha c_8)s_u\gamma - \\ d_1 - u_2 d_7 = 0,\end{aligned}$$

and the second one is as follows:

$$\begin{aligned}(b_4 - v_2 b_7)\beta + (c_4 - v_2 c_7)\gamma - (u_1 s_\alpha b_7 + v_1 s_\alpha b_8)s_u\beta - \\ (u_1 s_\alpha d_7 + v_1 s_\alpha d_8 + s_\alpha)s_u - (u_1 s_\alpha c_7 + v_1 s_\alpha c_8)s_u\gamma - \\ d_4 - v_2 d_7 = 0.\end{aligned}$$

The monomials of this polynomial system are $[\beta \quad \gamma \quad s_u \quad s_u\beta \quad s_u\gamma]^{\mathrm{T}}$.

Having two rotations α_1 and α_2 doubles the equations and introduces another unknown (each correspondence has different s_u). Thus the monomials of the polynomial equation system to which the two rotations lead are $[\beta \quad \gamma \quad s_{u,1} \quad s_{u,1}\beta \quad s_{u,1}\gamma \quad s_{u,2} \quad s_{u,2}\beta \quad s_{u,2}\gamma]^{\mathrm{T}}$, where $s_{u,i}$ is the scale along axis u of the ith correspondence ($i \in \{1, 2\}$). Since four equations are given for four unknowns and there is no higher-order term, the system can straightforwardly be rearranged and solved. The formulas for β and γ are as follows:

$$\begin{aligned}\beta = \quad & (-c_{\alpha_2}c_1d_7v_{2,2}s_{\alpha_1} + c_{\alpha_2}c_4d_7u_{2,1}s_{\alpha_1} + c_{\alpha_2}c_7d_1v_{2,2}s_{\alpha_1} \\ & -c_{\alpha_2}c_7d_4u_{2,1}s_{\alpha_1} - c_{\alpha_2}c_{\alpha_1}c_4d_7v_{2,1} + c_{\alpha_2}c_{\alpha_1}c_4d_7v_{2,2} \\ & +c_{\alpha_2}c_{\alpha_1}c_7d_4v_{2,1} - c_{\alpha_2}c_{\alpha_1}c_7d_4v_{2,2} - c_1d_7u_{2,1}s_{\alpha_2}s_{\alpha_1} \\ & +c_1d_7u_{2,2}s_{\alpha_2}s_{\alpha_1} + c_7d_1u_{2,1}s_{\alpha_2}s_{\alpha_1} - c_7d_1u_{2,2}s_{\alpha_2}s_{\alpha_1} \\ & +c_{\alpha_1}c_1d_7v_{2,1}s_{\alpha_2} - c_{\alpha_1}c_4d_7u_{2,2}s_{\alpha_2} - c_{\alpha_1}c_7d_1v_{2,1}s_{\alpha_2} \\ & +c_{\alpha_1}c_7d_4u_{2,2}s_{\alpha_2} + c_{\alpha_2}c_1d_4s_{\alpha_1} - c_{\alpha_2}c_4d_1s_{\alpha_1} \\ & -c_{\alpha_1}c_1d_4s_{\alpha_2} + c_{\alpha_1}c_4d_1s_{\alpha_2})/ \\ & (c_{\alpha_2}b_1c_7v_{2,2}s_{\alpha_1} + c_{\alpha_2}b_4c_7u_{2,1}s_{\alpha_1} + c_{\alpha_2}b_7c_1v_{2,2}s_{\alpha_1} \\ & -c_{\alpha_2}b_7c_4u_{2,1}s_{\alpha_1} - c_{\alpha_2}c_{\alpha_1}b_4c_7v_{2,1} + c_{\alpha_2}c_{\alpha_1}b_4c_7v_{2,2} \\ & +c_{\alpha_2}c_{\alpha_1}b_7c_4v_{2,1} - c_{\alpha_2}c_{\alpha_1}b_7c_4v_{2,2} - b_1c_7u_{2,1}s_{\alpha_1}s_{\alpha_2} \\ & +b_1c_7u_{2,2}s_{\alpha_1}s_{\alpha_2} + b_7c_1u_{2,1}s_{\alpha_1}s_{\alpha_2} - b_7c_1u_{2,2}s_{\alpha_1}s_{\alpha_2} \\ & +c_{\alpha_1}b_1c_7v_{2,1}s_{\alpha_2} - c_{\alpha_1}b_4c_7u_{2,2}s_{\alpha_2} - c_{\alpha_1}b_7c_1v_{2,1}s_{\alpha_2} \\ & +c_{\alpha_1}b_7c_4u_{2,2}s_{\alpha_2} + c_{\alpha_2}b_1c_4s_{\alpha_1} - c_{\alpha_2}b_4c_1s_{\alpha_1} \\ & -c_{\alpha_1}b_1c_4s_{\alpha_2} + c_{\alpha_1}b_4c_1s_{\alpha_2}), \\ \gamma = \quad & -(-c_{\alpha_2}b_1d_7v_{2,2}s_{\alpha_1} + c_{\alpha_2}b_4d_7u_{2,1}s_{\alpha_1} + c_{\alpha_2}b_7d_1v_{2,2}s_{\alpha_1} \\ & -c_{\alpha_2}b_7d_4u_{2,1}s_{\alpha_1} - c_{\alpha_2}c_{\alpha_1}b_4d_7v_{2,1} + c_{\alpha_2}c_{\alpha_1}b_4d_7v_{2,2} \\ & +c_{\alpha_2}c_{\alpha_1}b_7d_4v_{2,1} - c_{\alpha_2}c_{\alpha_1}b_7d_4v_{2,2} - b_1d_7u_{2,1}s_{\alpha_1}s_{\alpha_2} \\ & +b_1d_7u_{2,2}s_{\alpha_1}s_{\alpha_2} + b_7d_1u_{2,1}s_{\alpha_1}s_{\alpha_2} - b_7d_1u_{2,2}s_{\alpha_1}s_{\alpha_2} \\ & +c_{\alpha_1}b_1d_7v_{2,1}s_{\alpha_2} - c_{\alpha_1}b_4d_7u_{2,2}s_{\alpha_2} - c_{\alpha_1}b_7d_1v_{2,1}s_{\alpha_2} \\ & +c_{\alpha_1}b_7d_4u_{2,2}s_{\alpha_2} + c_{\alpha_2}b_1d_4s_{\alpha_1} - c_{\alpha_2}b_4d_1s_{\alpha_1} \\ & -c_{\alpha_1}b_1d_4s_{\alpha_2} + c_{\alpha_1}b_4d_1s_{\alpha_2})/ \\ & (-c_{\alpha_2}b_1c_7v_{2,2}s_{\alpha_1} + c_{\alpha_2}b_4c_7u_{2,1}s_{\alpha_1} + c_{\alpha_2}b_7c_1v_{2,2}s_{\alpha_1} \\ & -c_{\alpha_2}b_7c_4u_{2,1}s_{\alpha_1} - c_{\alpha_2}c_{\alpha_1}b_4c_7v_{2,1} + c_{\alpha_2}c_{\alpha_1}b_4c_7v_{2,2} \\ & +c_{\alpha_2}c_{\alpha_1}b_7c_4v_{2,1} - c_{\alpha_2}c_{\alpha_1}b_7c_4v_{2,2} - b_1c_7u_{2,1}s_{\alpha_1}s_{\alpha_2} \\ & +b_1c_7u_{2,2}s_{\alpha_1}s_{\alpha_2} + b_7c_1u_{2,1}s_{\alpha_1}s_{\alpha_2} - b_7c_1u_{2,2}s_{\alpha_1}s_{\alpha_2} \\ & +c_{\alpha_1}b_1c_7v_{2,1}s_{\alpha_2} - c_{\alpha_1}b_4c_7u_{2,2}s_{\alpha_2} - c_{\alpha_1}b_7c_1v_{2,1}s_{\alpha_2} \\ & +c_{\alpha_1}b_7c_4u_{2,2}s_{\alpha_2} + c_{\alpha_2}b_1c_4s_{\alpha_1} - c_{\alpha_2}b_4c_1s_{\alpha_1} \\ & -c_{\alpha_1}b_1c_4s_{\alpha_2} + c_{\alpha_1}b_4c_1s_{\alpha_2}).\end{aligned}$$

References

[1] D. Barath. P-HAF: Homography estimation using partial local affine frames. In *International Conference on Computer Vision Theory and Applications*, 2017.

[2] D. Barath and L. Hajder. A theory of point-wise homography estimation. *Pattern Recognition Letters*, 2017.

[3] D. Barath and J. Matas. Graph-Cut RANSAC. *Conference on Computer Vision and Pattern Recognition*, 2018.

[4] D. Barath, T. Toth, and L. Hajder. A minimal solution for two-view focal-length estimation using two affine correspondences. In *Conference on Computer Vision and Pattern Recognition*, 2017.

[5] D. Batra, B. Nabbe, and M. Hebert. An alternative formulation for five point relative pose problem. In *Workshop on Motion and Video Computing*.

[6] J. Bentolila and J. M. Francos. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding*, 2014.

[7] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Joint Pattern Recognition Symposium*, 2003.

[8] O. Chum, T. Werner, and J. Matas. Epipolar geometry estimation via RANSAC benefits from the oriented epipolar constraint. In *International Conference on Pattern Recognition*, 2004.

[9] R. Hartley and H. Li. An efficient hidden variable approach to minimal-case camera motion estimation. *Pattern Analysis and Machine Intelligence*, 2012.

[10] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[11] R. I. Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence*, 1997.

[12] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 2012.

[13] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *British Machine Vision Conference*, 2008.

[14] H. Li. A simple solution to the six-point two-view focal-length problem. In *European Conference on Computer Vision*, 2006.

[15] H. Li and R. Hartley. Five-point motion estimation made easy. In *International Conference on Pattern Recognition*, 2006.

[16] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer vision*, 1999.

[17] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 2004.

[18] J. Molnár and D. Chetverikov. Quadratic transformation for planar mapping of implicit surfaces. *Journal of Mathematical Imaging and Vision*, 2014.

[19] D. Nistér. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence*, 2004.

[20] M. Perdoch, J. Matas, and O. Chum. Epipolar geometry from two correspondences. In *International Conference on Pattern Recognition*, 2006.

[21] C. Raposo and J. P. Barreto. Theory and practice of structure-from-motion using affine correspondences. In *Computer Vision and Pattern Recognition*, 2016.

[22] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *International Journal of Computer Vision*, 2011.

[23] H. Stewénius, D. Nistér, F. Kahl, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. *Image Vision Computing*, 2008.

[24] R. Szeliski and P. Torr. Geometrically constrained structure from motion: Points on planes. *3D Structure from Multiple Images of Large-Scale Environments*, 1998.

[25] A. Torii, Z. Kukelova, M. Bujnak, and T. Pajdla. The six point algorithm revisited. In *Asian Conference on Computer Vision*, 2010.

[26] Y. Zhou, L. Kneip, and H. Li. A revisit of methods for determining the fundamental matrix with planes. In *International Conference on Digital Image Computing: Techniques and Applications*, 2015.