

Separable Four Points Fundamental Matrix

Gil Ben-Artzi

Computer Science Department

Ariel University

gilba@g.ariel.ac.il

Abstract

We present a novel approach for RANSAC-based computation of the fundamental matrix based on epipolar homography decomposition. We analyze the geometrical meaning of the decomposition-based representation and show that it directly induces a consecutive sampling strategy of two independent sets of correspondences. We show that our method guarantees a minimal number of evaluated hypotheses with respect to current minimal approaches, on the condition that there are four correspondences on an image line. We validate our approach on real-world image pairs, providing fast and accurate results.

1. Introduction

One of the basic building blocks in computer vision is the computation of epipolar geometry given a set of putative image point correspondences. Often, such correspondences include mismatches, therefore a robust estimation method needs to be carried out. The most common method used is RANSAC [9]. RANSAC-based computation of the fundamental matrix iteratively samples a minimal set of putative correspondences and hypothesizes the fundamental matrix parameters. The goodness-of-fit of each hypothesis is evaluated with respect to the whole set of the putative correspondences. The number of evaluations at each iteration is between one to three, depending on the number of possible solutions. This process is repeated until a predefined number of iterations is exceeded. One of its key limitations is that in the presence of a considerable amount of mismatched points' correspondences, a large number of hypotheses' evaluations is needed in order to obtain a reliable model. The number of the required hypotheses is directly related to the number of points needed to be sampled in order to hypothesize the fundamental matrix parameters. From a geometrical point of view, the minimal sample size is at least seven [11]. Minimal sample approaches have been proposed requiring only six [28] or five points' correspondences [24, 1], with additional knowledge regarding the pa-



Figure 1. Our approach matches points on line segments in the images without the need for the full computation of the fundamental matrix. Based on the matched points and line segments, we use our reformulation to compute the epipolar homography and reduce the required number of RANSAC iterations for the full computation. The red points are the matched putative point correspondences which are aligned on line segments (blue lines) in the image planes. The green lines are epipolar lines which are not known a-priori and do not take part in the matching process. The matching process is based on the computation of the epipolar homography which, unlike commonly used methods, does not require the knowledge of epipolar line correspondences.

rameters or the scene.

We present a method that can reduce the minimal number of RANSAC hypotheses' evaluations beyond the current known minimum. Our approach, similar to approaches that require less than seven correspondences, assumes additional knowledge. We require that a single line segment exists in the image with four correspondences and show experimentally that such a configuration is common. Given such

a configuration, we show that our approach guarantees a minimal number of hypotheses' evaluations. In cases where such a configuration does not exist, standard RANSAC can be applied with overall small overhead. Hereafter, when referring to RANSAC iterations, we will refer to the actual number of hypotheses' evaluations which depends upon the number of solutions per hypothesis.

We compute the fundamental matrix in two steps. In the first step, we sample three points on a line segment. We compute the epipolar homography and validate its goodness of fit using at least one additional point, without the full computation of the fundamental matrix. In the second step, the recovered matches are fixed and the remaining points' matches are recovered by sampling, without any restriction. Our approach is based on the computation of the epipolar homography. Existing methods for the computation of the epipolar homography [13, 23, 5, 10, 11] require the knowledge of (at least) three corresponding epipolar lines and, as a consequence, are rarely used in practice. We show how to compute the epipolar homography without a-priori knowledge of epipolar lines. Figure 1 shows the recovered four matches in the first step of our approach and the extracted epipolar lines based on our epipolar homography computation.

This paper therefore contributes by presenting: (a) a reformulation of the fundamental matrix based on epipolar homography decomposition and an analysis of its geometrical meaning, (b) a novel two steps RANSAC-based approach for the computation of the fundamental matrix that can markedly reduce the required number of hypotheses' evaluations and (c) an efficient implementation and a validation of our approach on real-world image pairs showing that it is accurate and applicable.

2. Related Work

The most common approaches for the computation of the fundamental matrix are the seven or eight point algorithms [11]. The eight-point algorithm [15] which was adapted for the fundamental matrix, was made practical by [12]. It is based on the normalization of the points' correspondences, and computes the parameters based on the Direct Linear Transform (DLT) by enforcing the rank 2 constraint [17]. The seven points algorithm relaxes the eight points requirement by an additional zero determinant constraint of the matrix, resulting in a cubic equation with one or three real solutions.

Various minimal methods assume additional knowledge in order to reduce the required number of correspondences. If the camera parameters are known, the five-points algorithm [18] can be used. Recently, Barath [1] proposed an approach for the estimation of the fundamental matrix based on five correspondences, in cases where three coplanar point correspondences and the rotation of features

are known. Ben-Artzi et al. [4] showed that in case the intensity of the epipolar lines are similar across views, two corresponding points are sufficient for the computation of matching epipolar lines, and can be used for recovery of the fundamental matrix.

Methods attempting to empirically reduce the number of RANSAC iterations have also been introduced [3, 6, 19]. Unlike such methods, our approach, as well as other minimal methods, guarantees to reduce the number of required samples and this number can be calculated in advance in the same way as in standard RANSAC. In addition, our approach has a true geometrical meaning (See Sec. 3), whereas no geometrical analysis is possible in such methods.

The fundamental matrix can be computed based on the epipolar homography [11, 8]. These methods require the knowledge of epipolar line correspondences. Sinha [23] introduced an approach for camera calibration based on the computation of the epipolar homography and epipoles hypothesizing. His approach was later improved by [5], who directly recovered the epipolar lines required for the computation of the epipolar homography by using the motion-barcode descriptor. Similar approaches were introduced by [13, 10] for the computation of the epipolar homography by directly estimating the epipolar lines. However, all of the above methods are only applicable to videos of dynamic scenes. Wurfl et. al. [30] presented an approach for estimating the fundamental matrix transmission imaging, based on the epipolar homography formulation. They do not require explicit correspondences but rather sampling of all edge pixels of the image. The computation of the relative pose by general line homography based on line segments was demonstrated by [20] but they required the matching points to be the projection of existing 3D lines in the scene and required at least two such lines.

3. Theoretical Background

Epipolar-Lines Based Parametrization. Let l, l' be corresponding epipolar lines, in the first and second image, respectively. Let l_p be a line that intersects l in the point p and does not include the epipole. It follows that $p = l_p \times l$, where \times is the cross product. Since p is on l , it follows that $l' = Fp$. Denote $[\cdot]_x$ as the skew symmetric matrix associated with the cross product, we have the following mapping:

$$l' = F[l_p]_x l$$

where $H_l = F[l_p]_x$ is the epipolar homography. The seven degrees of freedom of the fundamental matrix is constructed by the four degrees of freedom of the two epipoles and the three degrees of freedom of the epipolar homography [11]. The epipolar homography is obtained by three corresponding epipolar lines.

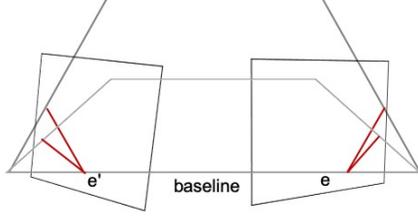


Figure 2. The pencil of epipolar planes. As the epipolar lines (red lines) rotate around the epipoles (e, e'), their associated planes rotate around the baseline.

Image-Points' Based Parametrization. Consider the set of epipolar lines passing through the epipole which is denoted as the *pencil of epipolar lines*. As shown in Fig. 2, the corresponding epipolar lines across images define a plane which intersects the two retinal planes with its axis as the baseline. When the corresponding epipolar lines rotate around the epipoles, their corresponding planes rotate around the baseline, defining the *pencil of epipolar planes*. The two corresponding pencils of epipolar lines are related by the epipolar homography based on the pencil of epipolar planes [8]. Let (l_i, l'_i) be corresponding epipolar lines in the first and second image respectively. Let \bar{l}, \bar{l}' be arbitrary lines in the first and second image, such that the epipoles e, e' are not on the lines. The 2×2 epipolar homography H_e maps between (l_i, l'_i) based on their intersection points with \bar{l}, \bar{l}' as follows. Let $x_1, x_2 \in \bar{l}$ and $x'_1, x'_2 \in \bar{l}'$, denoted as *control points*. The intersection points of (l_i, l'_i) with \bar{l}, \bar{l}' are $x_p = (x_1 \times x_2) \times l_i$ and $x'_p = (x'_1 \times x'_2) \times l'_i$, where \times is the cross product and using homogeneous coordinates of the control points from now on. Using the following identity

$$v \times (u \times w) = (v^T w)u + (-v^T u)w$$

for $u, v, w \in \mathbb{R}^{3 \times 1}$, the intersection points can be written as a linear combination of the control points as $x_p = \alpha x_1 + \beta x_2$ and $x'_p = \alpha' x'_1 + \beta' x'_2$, where

$$\alpha = l_i^T x_2, \quad \beta = -l_i^T x_1, \quad \alpha' = l'_i{}^T x'_2, \quad \beta' = -l'_i{}^T x'_1,$$

$\alpha, \beta, \alpha', \beta' \in R$. The 2×2 epipolar homography is given by:

$$\begin{bmatrix} \alpha' \\ \beta' \end{bmatrix} = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{H_e} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (1)$$

Rewriting the corresponding epipolar lines as the cross product of the epipoles and points on the epipolar lines, and expanding the intersection points' coefficients, we have the following parametrization of the fundamental matrix:

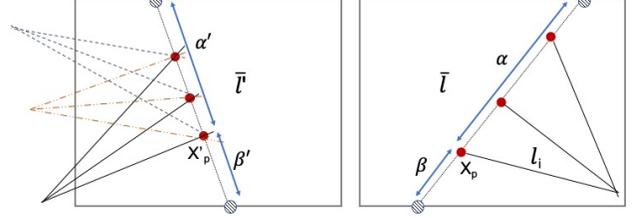


Figure 3. The epipolar homography has three degrees of freedoms. Existing methods compute the epipolar homography based on epipolar lines. The points-based parametrization enables the epipolar homography's computation regardless of the epipole positions. On the left image, various epipole positions are valid for the same epipolar homography. The parametrization of the intersection point x_p of the epipolar line l_i is based on a linear combination α, β of selected fixed points on the lines, marked by the striped points, intersecting the boundaries of the image. x'_p is the corresponding intersection point of x_p and is parametrized by $\bar{l}', \alpha', \beta'$. See text for details.

$$F \approx [e_2]_x [x'_1, x'_2] \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_2^T \\ -x_1^T \end{bmatrix} [e_1]_x \quad (2)$$

Observation 1. The epipolar homography H_e maps between the corresponding points x_p, x'_p based on their representation as a linear combination of their respective control points. It can be computed without knowledge of the corresponding epipolar lines. Given the line segments of the points, we select (fixed) control points on the lines and represent the points as their linear combination. The points-based parametrization of the pencil of epipolar lines by their intersection with the lines \bar{l}, \bar{l}' is shown in Fig. 3.

Observation 2. The required number of RANSAC iterations for the full computation of the fundamental matrix can be greatly reduced. First, we recover H_e which is a three parameters model without the need for the full computation of the fundamental matrix. We can then recover four additional arbitrary correspondences by relying on the recovered H_e . Let $F(k, r)$ be the required number of RANSAC iterations for a model with k parameters, outlier rate r and success probability of 0.99. The required number of RANSAC iterations is now $F(3, r) + F(4, r)$ instead of $F(7, r)$. Fig. 4 shows the required number of RANSAC iterations for the computation of the fundamental matrix by our approach and existing minimal methods, as a function of the outlier rate. Our approach can be implemented based on the seven-point or eight-point algorithm; both are presented in the figure. The number of RANSAC iterations is computed assuming one solution per sample and no additional overhead for any of the methods. In Sec. 6, we present

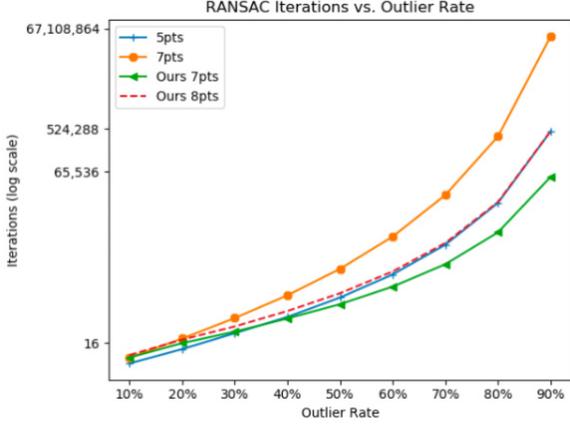


Figure 4. The theoretical number of RANSAC iterations for the computation of the fundamental matrix by minimal methods. 7pts represents the seven-point algorithm and 5pts represents a RANSAC-based approach requiring a sample size of five (e.g. [1]). Ours-7pts represents our approach based on the seven-point algorithm and Ours-8pts represents our approach based on the eight-point algorithm.

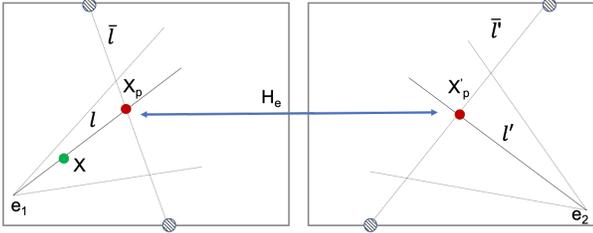


Figure 5. The mapping of a point on the left image to the corresponding epipolar line on the right image by the fundamental matrix, parametrized by points-based epipolar homography H_e : (a) from a point x to epipolar line l based on the known epipole e_1 , (b) from the epipolar line l to the intersection point x_p with a fixed arbitrary image line \bar{l} , (c) from the intersection point x_p to the corresponding intersection point x'_p with a fixed arbitrary line \bar{l}' and (d) from the point of intersection x'_p to the epipolar line l' based on the known epipole e_2 .

the expected number of RANSAC iterations based on the actual number of solutions per sample in the tested datasets.

Observation 3. The lines \bar{l}, \bar{l}' with the corresponding points are arbitrary line segments in the images. Figure 5 illustrates the mapping of points in the first image to lines in the second image, based on the parametrization of the fundamental matrix using the points-based epipolar homography (Eq. 2), given \bar{l}, \bar{l}' .

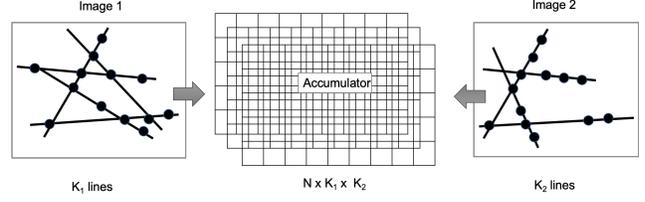


Figure 6. We match points on line segments without the need to explicitly compare between lines across images. In each image, Hough transform is carried out and a shared accumulator of size $N \times K_1 \times K_2$ is updated independently, where N is the number of correspondences, K_1 is the number of lines in the first image and K_2 is the number of lines in the second image.

4. Line Segments with Corresponding Points

In this section we describe how to find the lines \bar{l}, \bar{l}' across images with the maximum number of putative correspondences. For each image, our approach does not require knowledge of the points and lines in the other image, which allows parallel implementation. The key idea is to use Hough transform [7] to detect the lines independently in each image and update a joint accumulator which can be queried only after the process is completed. Fig. 6 illustrates the matching process of corresponding lines across images.

Let I_1, I_2 denote a pair of images. Let $X = \{(x_i, x'_i)\}_{i=1..N}$ denote the set of ordered pairs of unique putative corresponding image points where $x_i, x'_i \in \mathbb{R}^2$ are image points in I_1, I_2 , respectively. Let $\bar{X} = \{x_i\}_{i=1}^N$, $\bar{X}' = \{x'_i\}_{i=1}^N$. We create binary images B_1, B_2 such that for each pixel y_i in I_1 , $B_1(y_i) = 1$ if $y_i \in \bar{X}$ otherwise $B_1(y_i) = 0$, and similarly for B_2 . We use Hough transform [7] to extract the ordered set of lines $L_1 = \{l_i\}_{i=1}^{K_1}$, $L_2 = \{l'_i\}_{i=1}^{K_2}$ in B_1, B_2 , respectively. We consider only lines with at least four correspondences. This requirement is due to the robust estimation process of the homography (See Sec. 5), where we sample three points and need at least one additional point to estimate the quality of the hypothesized homography.

We define a multidimensional array, an accumulator A of size $N \times K_1 \times K_2$, where K_1 and K_2 are the number of lines in L_1 and L_2 , respectively, and N is the number of putative corresponding points. The accumulator is initialized to zero. Let $D_1(j) \subset \{1..K_1\}$ be the indices of nearby lines in L_1 for a point x_j in image I_1 ,

$$D_1(j) = \{i | l_i \in L_1, x_j \in \bar{X}, d(x_j, l_i) < C\}, \quad (3)$$

where d is the point to line distance and C is the constant representing the required distance between the lines and points. For each point x_j in the first image, we increment the accumulator according to its nearby lines in

the first image and all existing lines in the second image, $A(j, m, n)_+ = 1$ s.t.

$$\{(j, m, n) | m \in D_1(j), n \in \{1 \dots K_2\}\}, \quad (4)$$

For each point x'_j in the second image, $D_2(j) \subset \{1 \dots K_2\}$ is defined similarly and the accumulator entries (j, m, n) are incremented according to

$$\{(j, m, n) | m \in \{1 \dots K_1\}, n \in D_2(j)\}, \quad (5)$$

The pair of matching lines with the maximum number of putative correspondences are given by $(l_{\mathbf{m}^*}, l'_{\mathbf{n}^*})$ where

$$(\mathbf{m}^*, \mathbf{n}^*) = \operatorname{argmax}_{m, n} \sum_j A(j, m, n), \quad (6)$$

In case there is more than a single line we randomly select one.

5. Separable Four Points Fundamental Matrix

Given a set of putative correspondences $\{(x_i, x'_i)\}_{i=1 \dots N}$, we compute the fundamental matrix by two steps. Our computation is based on RANSAC. We assume inlier ratio R , epipolar homography inlier threshold T_1 and fundamental matrix inlier threshold T_2 .

Step One. Based on Sec. 4, we recover $(l_{\mathbf{m}^*}, l_{\mathbf{n}^*})$, which are the corresponding image lines with the maximum number of K putative corresponding points $(x_i, x'_i)_{1 \dots K}$. We select fixed control points for each line $l_{\mathbf{m}^*}$ and $l_{\mathbf{n}^*}$. The points are selected on the intersection of the lines and the boundaries of the images. For each pair of putative corresponding points on the lines, we compute their representation by the control points with the two possible orientations and use the second one only if the first fails. We use RANSAC to recover the epipolar homography. We iteratively sample three putative corresponding points, compute the epipolar homography, transfer each point on the line to the corresponding line, and count the number of inliers based on the Euclidean distance and the threshold T_1 . The number of required iterations for the three parameter model is often small. For example, for an outlier rate of 60%, we need only 71 iterations to recover the correct model with a confidence of 99%. Note that although we sample three points, we require at least four corresponding points on line segment; the additional points are used for validation of the hypothesized epipolar homography.

Step Two. Given the three points selected, we iteratively sample four additional points and compute the fundamental matrix based on the seven points algorithm. The number of iterations is computed based on R with respect to four parameters. We count the number of inliers based on T_2

	Pairs	Ground truth
Strecha	569	Given
Flickr	2000	COLMAP
Tanks and Temples	2000	COLMAP

Table 1. The datasets. For both Flickr and Tanks and Temples we obtained the ground truth fundamental matrices based on COLMAP reconstruction.

with respect to the symmetric epipolar distance [11]. For the same case as above of an outlier rate of 60%, for the four parameters' model we need only 178 iterations.

The above refers to the implementation of our approach based on the seven points algorithm. For implementation based on the eight points algorithm, we sample five points instead of four in the second step.

6. Experiments

6.1. Datasets

We conduct experiments on the following real world datasets: (1) The Strecha datasets [25] which include ground truth cameras and reconstruction, (2) The Tower of London sequence from the Flickr dataset [29], where the images were downloaded from Flickr based on geotag. It includes images captured by the community at random camera positions with a large variety of urban outdoor scenes, and (3) The Family sequence from the Tanks and Temples dataset [14], which includes wide baseline camera poses, with medium sized images captured by a hand-held camera in an urban yard scene.

For the last two datasets, we sampled 2000 image pairs with at least 20 matches with a symmetric epipolar distance of less than or equal to 1 pixel with respect to the ground truth, and used them for the evaluation. We show that our approach is able to successfully compute the fundamental matrices for these real-world diverse sets of image pairs, leading to an efficient estimation process. Table 1 shows the properties of the datasets.

6.2. Matching Co-Linear Points

In this subsection, we discuss the matching process of corresponding lines with at least four points, based on the Strecha dataset.

How Many Lines? On average, the Hough transform recovers 286.4 independent lines in each image. Out of all possible pairs of matched lines, the average number of lines with corresponding points across a pair of images is 3354.91, including lines with outliers. 45.8% (1538.84) of the matched lines are lines with at least four correspondences, where a corresponding point might be either inlier or outlier.

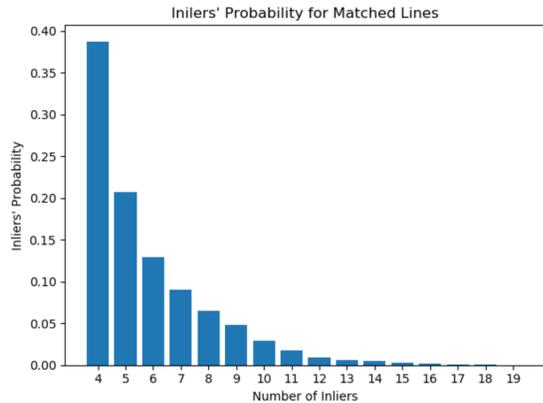


Figure 7. The probability for a given number of inliers in matched lines. The x-axis is the number of inliers and the y-axis is the probability. The average number of inliers is 5.83.

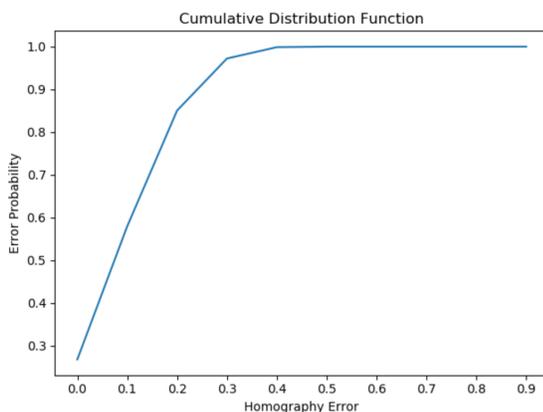


Figure 8. The cumulative distribution function (CDF) of the line homography error in matched lines. The x-axis is the error in pixels and the y-axis is the probability. 99% of the inliers are with a homography error of equal to or less than 0.5.

Quality of Lines. Overall, we would like to match lines with a high number of inliers and a low homography error. The average number of points, including outliers, on matched lines is 7.62. The average number of inliers on corresponding lines is 5.83, for a homography threshold of 1. Fig. 7 shows the probability of having a given number of inliers in matched lines, for such a homography error. Fig. 8 presents the probability for a given homography error on the matched lines. The homography error of 97% of the inliers is below 0.4 and 99% are below 0.5, which is used as our default threshold.

Scene Structure. Our approach often successfully matches lines in both near and far scenes and low and high inlier rates. Fig. 9 shows examples from Tower of London dataset of matched lines and corresponding points recovered by our approach, with inlier ratios of 0.18, 0.28 and



Figure 9. Examples of matched lines recovered by our approach for various scenes from Tower of London dataset. The red circles are the matched points on the corresponding lines and the cyan dots are the putative correspondences. The inliers ratio are 0.18, 0.28 and 0.85 from top to bottom.

0.85, for various scenes. For the first and second row from the top, there are 27 and 25 inliers out of 148 and 89 putative correspondences, respectively. For the last row, there are 498 inliers out of 582 putative correspondences.

Missing Inliers Relative to RANSAC, our approach failed to recover at least 20 inliers across images in 4.9% of the cases. Qualitatively, these cases occur when there is only a small number of inliers, both in the images and on the recovered lines. For the failure cases, the average number of inliers is 26.8 and the maximum is 34. Quantitative analysis indicates that in such cases the best possible matching of corresponding lines across images contains wrong or inaccurate matches and the homography could not be recovered. Fig. 10 shows such an example. The overall percentages of failure cases over all datasets are presented in Table 5.

6.3. Evaluation

Methods. We compared RANSAC-based [9] and LMEDS-based [21] computations of the fundamental matrix based on the following minimal solvers' methods:

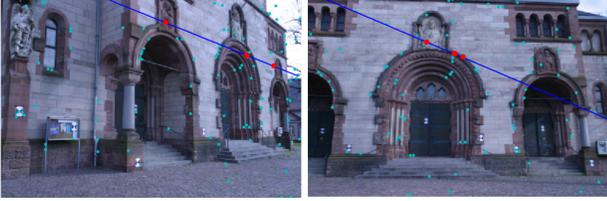


Figure 10. An example of a failure case. The the best possible corresponding lines are with wrong matches and the homography could not be recovered.

- RANSAC with the standard eight points algorithm, denoted as RANSAC.
- LMEDS with the standard eight points algorithm, denoted as LMEDS.
- State-of-the-art minimal solver of [1] combined with GC-RANSAC [2], which is based on homography and rotation information of five points, denoted as Hom-Rot-5.
- RANSAC with our minimal sampling solver, using the seven points algorithm based on two samples of three and four points, denoted as Ours-RANSAC-4.
- RANSAC with our minimal sampling solver, using the eight points algorithm based on two samples of three and five points, denoted as Ours-RANSAC-5.
- LMEDS with our minimal sampling solver, using the seven points algorithm based on two samples of three and four points, denoted as Ours-LMEDS-4.
- LMEDS with our minimal sampling solver, using the eight points algorithm based on two samples of three and five points, denoted as Ours-LMEDS-5.

Methodology. For the baseline methods we performed a grid-search over hyperparameters. For sequences with no ground truth fundamental matrices, we reconstructed the sequences using COLMAP [22]. COLMAP is a general purpose Structure-from-Motion (SfM) pipeline which globally infers the 3D structure, leading to accurate results. We obtained the camera poses and fundamental matrices, and used them as the ground truth.

For all sequences, we used the SIFT [16] descriptor to extract and match putative correspondences between all pairs of images. We used the ground truth fundamental matrices to evaluate the symmetric epipolar distance of the putative matching. Each image pair with less than 20 matches within a symmetric epipolar distance of one is discarded.

Metrics. We report the average symmetric epipolar distance resulting from our approach and the baselines. Unless otherwise stated, we used a threshold of 3 pixels in

	% inliers	F-Score	Mean
RANSAC	71.2	0.91 ± 0.05	0.61
LMEDS	71.1	0.88 ± 0.08	0.57
Hom-Rot-5	71.5	0.90 ± 0.06	0.60
Ours-LMEDS-4	62.4	0.81 ± 0.06	0.72
Ours-LMEDS-5	72.1	0.90 ± 0.05	0.59
Ours-RANSAC-4	61.7	0.82 ± 0.06	0.71
Ours-RANSAC-5	72.2	0.91 ± 0.05	0.58

Table 2. Results of the Strecha dataset, based on the ground truth camera poses.

	% inliers	F-Score	Mean
RANSAC	43.7	0.52 ± 0.08	0.63
LMEDS	43.2	0.53 ± 0.09	0.69
Hom-Rot-5	44.5	0.58 ± 0.08	0.61
Ours-LMEDS-4	41.6	0.51 ± 0.1	0.74
Ours-LMEDS-5	46.5	0.60 ± 0.09	0.57
Ours-RANSAC-4	42.9	0.51 ± 0.1	0.78
Ours-RANSAC-5	49.3	0.62 ± 0.09	0.54

Table 3. Results of Tower of London sequence from the Flickr dataset.

	% inliers	F-Score	Mean
RANSAC	45.4	0.72 ± 0.07	0.69
LMEDS	44.2	0.71 ± 0.08	0.68
Hom-Rot-5	45.5	0.78 ± 0.07	0.61
Ours-LMEDS-4	43.8	0.61 ± 0.09	0.71
Ours-LMEDS-5	45.7	0.71 ± 0.07	0.66
Ours-RANSAC-4	44	0.62 ± 0.08	0.69
Ours-RANSAC-5	48.1	0.73 ± 0.07	0.63

Table 4. Results of the Family sequence from the Tanks and Temples dataset.

our approach as the inlier threshold and 0.5 as the epipolar homography threshold. We use a fixed homography’s threshold over all the datasets, demonstrating its robustness. We report the percentage of inliers found by each method, and report the F-measure, where corresponding points with symmetric epipolar distance less than 1 pixel with respect to the ground truth fundamental matrix is considered as positive.

6.4. Accuracy

Strecha. The dataset contains eight multi-view collections of high-resolution images (3072×2048), provided with ground truth camera poses. For the computation of matching lines with putative corresponding points across images, we sampled 150 putative correspondences, and used these lines with respect to all existing points’ correspondences. We used SIFT ratio-test of 0.75. The binary images used for matching the lines (Sec. 4) are of a lower resolution, as suggested in [26, 27], and the points coordinates are resized accordingly. Unless noted otherwise, the width resolution of 512 was used for the line matching phase with the original aspect ratio. The reported results are with respect to the original resolutions. The sequences are

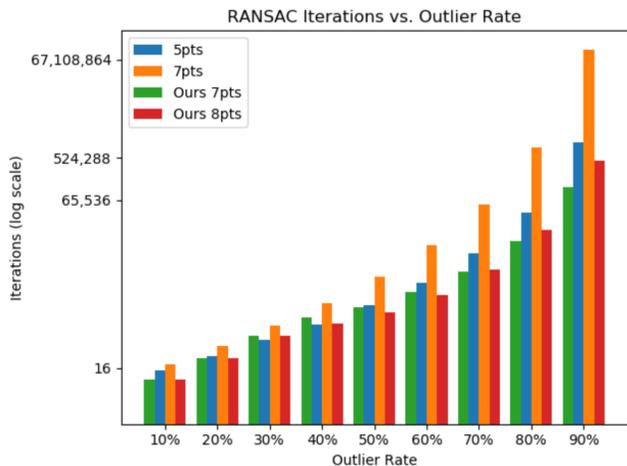


Figure 11. The expected number of RANSAC iterations based on the actual number of solutions per sample in the 7-point algorithm and the additional preprocessing iterations required by our approach.

	RANSAC Iterations	%Failure
RANSAC	2931	-
Hom-Rot-5	671	4.3
Ours-RANSAC-4	627	7.2
Ours-RANSAC-5	643	5.1

Table 5. The number of RANSAC iterations and failure cases relative to RANSAC for all three datasets.

of images with varying distances between focal points with a total of 569 valid image pairs. As suggested in [1], we used the five-points’ solver of [1] within the GC-RANSAC [2] estimator. Table. 2 presents our results.

Flickr. The dataset obtained by downloading images of specific scenes from Flickr represents the most challenging case. The images captured random scenes at different times, with various cameras from very different positions. We use the Tower of London sequence for comparison. We reconstruct the sequence using COLMAP[22] and use the reconstruction as ground truth. The results are presented in Table 3, where it can be seen that our RANSAC-based eight points approach clearly outperforms all other methods.

Tanks and Temples. The dataset includes medium-resolution images (1920×1080). For comparison, we used the Family sequence. As ground truth, we use the COLMAP Reconstruction. As shown in Table 4, our RANSAC-based eight points approach recovered the highest number of inliers, with mean error similar to [1]. Both outperform the baselines.

6.5. Efficiency

We measured the number of RANSAC iterations required by each method and the number of failure cases. For the total number of RANSAC iterations, we also consider all validation iterations. For methods which are based on the seven points algorithm, both ours and the baseline, there might be more iterations than the theoretical number of required samples. A failure is considered if the solver has not been able to accurately recover at least 20 matches, in case there are indeed such ground truth matches.

The solver of [1] is combined within another robust estimator framework, the GC-RANSAC [2]. We execute only the solver itself with the optimal number of iterations as an initialization. Our approach starts by finding the matching lines with the maximum number of putative correspondences (Sec. 5). Based on our experiments the runtime of our lines’ matching step is equivalent to 71 RANSAC iterations and the number of solutions per sample for the 7-point algorithm is 2.43. To reduce the additional overhead associated with our matching step we use instead the standard eight points’ RANSAC for inlier ratio less than 0.3. For such inlier ratio the 8-point algorithm requires only one hypothesis’ evaluation per sample, with overall 9-78 iterations instead of 20-132 for the 7-point algorithm. Fig. 11 presents the required number of iterations by the different approaches based on the actual number of solutions per sample including the additional iterations required by our matching lines phase.

Table 5 shows the percentage of cases that failed for each approach and the average number of iterations, over all the three datasets, the Strecha, Flickr and Tanks and Temples. Our approach significantly reduces the required number of hypotheses’ evaluations.

7. Conclusion

We presented a novel approach for the computation of the fundamental matrix based on epipolar homography decomposition. Our approach can reduce, both theoretically and practically, the number of required iterations to a new minimal number. The standard RANSAC procedure can be incorporated into our approach, providing an overall robust and efficient solution which is well suited for Structure from Motion (SfM) pipelines.

References

- [1] Daniel Barath. Five-point fundamental matrix estimation for uncalibrated cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 235–243, 2018.
- [2] Daniel Barath and Jiří Matas. Graph-cut ransac. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6733–6741, 2018.

- [3] Daniel Barath, Jiri Matas, and Jana Noskova. Magsac: marginalizing sample consensus. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10197–10205, 2019.
- [4] Gil Ben-Artzi, Tavi Halperin, Michael Werman, and Shmuel Peleg. Epipolar geometry based on line similarity. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 1864–1869. IEEE, 2016.
- [5] Gil Ben-Artzi, Yoni Kasten, Shmuel Peleg, and Michael Werman. Camera calibration from dynamic silhouettes using motion barcodes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4095–4103, 2016.
- [6] Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 220–226. IEEE, 2005.
- [7] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. Technical report, Sri International Menlo Park Ca Artificial Intelligence Center, 1971.
- [8] Olivier Faugeras, Quang-Tuan Luong, and Theo Papadopoulos. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2001.
- [9] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- [10] Tavi Halperin and Michael Werman. An epipolar line from a single pixel. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 983–991. IEEE, 2018.
- [11] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [12] Richard I Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):580–593, 1997.
- [13] Yoni Kasten, Gil Ben-Artzi, Shmuel Peleg, and Michael Werman. Fundamental matrices from moving objects using line motion barcodes. In *European Conference on Computer Vision*, pages 220–228. Springer, 2016.
- [14] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017.
- [15] H Christopher Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, MA Fischler and O. Firschein, eds, pages 61–62, 1987.
- [16] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [17] Quan-Tuan Luong and Olivier D Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International journal of computer vision*, 17(1):43–75, 1996.
- [18] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):0756–777, 2004.
- [19] David Nistér. Preemptive ransac for live structure and motion estimation. *Machine Vision and Applications*, 16(5):321–329, 2005.
- [20] Irene Reiser-Kollmann, Andreas Reichinger, and Werner Purgathofer. 3d camera pose estimation using line correspondences and 1d homographies. In *International Symposium on Visual Computing*, pages 41–52. Springer, 2010.
- [21] Peter J Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984.
- [22] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [23] Sudipta N Sinha and Marc Pollefeys. Camera network calibration and synchronization from silhouettes in archived video. *International journal of computer vision*, 87(3):266–283, 2010.
- [24] Henrik Stewénus, David Nistér, Fredrik Kahl, and Fredrik 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.
- [25] Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool, Pascal Fua, and Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, 2008.
- [26] Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):815–830, 2009.
- [27] Eduard Trulls, Iasonas Kokkinos, Alberto Sanfeliu, and Francesc Moreno-Noguer. Dense segmentation-aware descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2890–2897, 2013.
- [28] Wei Wang and Chengke Wu. Six-point synthetic method to estimate fundamental matrix. *Science in China Series E: Technological Sciences*, 40(2):199–205, 1997.
- [29] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *European Conference on Computer Vision*, pages 61–75. Springer, 2014.
- [30] Tobias Wurfl, Andre Aichert, Nicole Maass, Frank Dennerlein, and Andreas Maier. Estimating the fundamental matrix without point correspondences with application to transmission imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1072–1081, 2019.