Recovery of 3D Urban Scenes: Epipolar Geometry of Two Views and Photo-Sequencing

Álvaro Budria, Alex Carrillo, Sergi Masip and Adrià Molina

Universitat Pompeu Fabra

January 2022

1 Introduction

In this document, we present our results after implementing the estimation of the fundamental matrix relating two image views of the same scene via the *normalized 8-point algorithm*, using two methods: an algebraic one and a more robust one using RANSAC. Furthermore, the correct estimation of the fundamental matrix is verified by computing the corresponding epipolar lines of some sample points. Finally, the application of these algorithms is evaluated in the *photo-sequencing problem* where the goal is to temporally order a set of still images taken asynchronously by a set of uncalibrated cameras.

2 Estimation of the fundamental matrix

In epipolar geometry, the fundamental matrix F is a 3x3 matrix that relates corresponding points between two images in an uncalibrated camera setting. More precisely, for any pair of corresponding points $(x \Leftrightarrow x') \in \mathbb{P}^2$ on two images I and I' expressed in homogeneous coordinates, such that x and x' are the projection of the same 3D point onto each image plane, the fundamental matrix F satisfies Equation 1.

$$x'^T F x = 0 \tag{1}$$

With the given definition of Equation 1, one can derive a set of constraints to determine the parameters of the Fundamental matrix. Given a pre-determined set of correspondences (e.g. established by matching feature points between images $(p \Leftrightarrow p')$), a homogeneous system of linear equations can be formulated to calculate the fundamental matrix F. Hence, by setting the last coordinate of p and p' to zero, the following Equation 2 can be obtained. Moreover, the system of equations can be expressed in Equation 3 as a 1×9 vector multiplication of points coordinates by a 9×1 vector of unknown entries of the matrix F by rearranging and isolating the terms. It can be observed that only 8 point correspondences are required: F has only 7 degrees of freedom as it is defined up to scale and $det(F) = 0$. Note that each point correspondence $(p_i \Leftrightarrow p'_i)$ yields different vectors w_i.

$$p'^T F p = \begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \tag{2}$$

$$w_i f = \begin{bmatrix} uu' & vu' & u' & uv' & vv' & v' & u & v & 1 \end{bmatrix}_i \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0 \tag{3}$$

Hence, one can build a matrix W by stacking eight row vectors w_i to find a solution for the vector of unknown entries of the matrix F by finding the nullspace of W. Bear in mind that, analogous to the DLT algorithm for homography estimation, the trivial solution $f = 0$ satisfies the equations. Therefore, we approach the problem as the least squares minimization problem shown in Equation 4.

$$\begin{aligned} \min_{f} \quad & \|Wf\|_2 \\ \text{s.t.} \quad & \|f\| = 1 \end{aligned} \tag{4}$$

2.1 Implementation of the normalized 8-point algorithm

The implementation of the 8-point algorithm is performed in the function called `fundamental_matrix(p,p')`. To begin with, the points are normalized, analogous to the DLT algorithm for homography estimation. The fact is that the algorithm is highly numerically unstable as it involves two SVD decompositions and may also suffer from measurement errors in the correspondences. Moreover, further instability is caused by the varying scale of the elements in W: for example, different entries entries such u may be in the order of 10 and vv' may be in the order of 10^6. To mitigate this, points are normalized to be centered at the origin of the coordinate frame and have an average distance to the origin of $\sqrt{2}$. On account of that, two homographies H and H' of the form of Equation 5 are set up to normalize the set of points p and p' respectively. Note that $c = [c_x, c_y]^T$ is the mean vector of the set of points and s is the average distance to the centroid. This process is defined in the `normalize_points(p)` function. Then, with the normalized points, the matrix W is built.

$$H = \begin{bmatrix} s & 0 & -sc_x \\ 0 & s & -sc_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

An initial estimation $\hat{F}$ of the fundamental matrix is found by taking the SVD of $W = UDV^T$, picking the last column of V as the vector of coefficients f, and rearranging them into $\hat{F}$. However, in general, $\hat{F}$ will not be of rank 2 but 3 due to the noise introduced by the point correspondences. This means that epipolar lines will not coincide in the epipole. Consequently, and to impose this property, one can decompose matrix $\hat{F}$ via SVD as $F_{R_3} = \bar{U}\bar{D}\bar{V}^T$ and set the last eigenvalue of D to 0 in order to find the closest rank 2 matrix to $\hat{F}$ in Frobenius norm, denoted as D_{R_2}. Hence, the fundamental matrix of rank 2 can be derived as $F_{R_2} = \bar{U}\bar{D}_{R_2}\bar{V}^T$. Finally, an additional step of denormalization is needed, as F_{R_2} was estimated using normalized points. The final denormalized fundamental matrix resolves into $F = H'^T F_{R_2} H$ as shown in Equation 6 with $\tilde{p}'$ and $\tilde{p}$ denoting the normalized points p' and p, respectively.

$$p'^T F p = (p'H')^T F_{R_2}(Hp) = \tilde{p}'^T F_{R_2} \tilde{p} = 0 \tag{6}$$

To verify the correctness of the implemented function for estimating the fundamental matrix, a comparison with a known ground truth fundamental matrix can be made. First, given the parameters of the camera, the fundamental matrix can be defined as Equation 7. However, since the intrinsic camera parameters K are not provided in the example (only T' and R are given, having $P_1 = [I|0]$ and $P_2 = [R|t]$), and knowing that the camera matrix is defined as $P = K[R|t]$, it can be assumed that camera parameters are equal to the identity matrix $K = K' = I$. This leads to the conclusion that the ground truth fundamental matrix can be calculated as $F = E$. Therefore, to compute the ground truth fundamental matrix one just needs to algebraically compute the essential matrix E. Results in Equation 8 show that the ground truth fundamental matrix and the matrix estimated with the normalized 8 point algorithm is practically the same, with a difference in norm near 0 ($\sim 1.31^{-14}$).

$$F = K'^{-T} E K^{-1} = K'^{-T} [T'_x] R K^{-1} \tag{7}$$

$$F_{GT} = \begin{bmatrix} 0.097 & 0.365 & -0.188 \\ -0.365 & 0.097 & 0.566 \\ 0.035 & -0.596 & 0 \end{bmatrix} \approx \hat{F} \tag{8}$$

As an additional check to ensure that the fundamental matrix F has been properly estimated, we also randomly pick two points p' and p, and compute $p'\tilde{F}p$. This should be equal to 0. With our estimated fundamental matrix, the result is $2.08 \cdot 10^{-16} \approx 0$.

2.2 Robust estimation of the fundamental matrix

Our goal in this section is to estimate the fundamental matrix in a real-world setting in which the image correspondences present outliers. To do so, we use a robust version of the previous algorithm.

Note that the normalized 8 point algorithm presented in Section 2 is sensitive to outliers in the points correspondences, which is a quite common phenomenon when matching computed keypoints between images. To address this issue, the algorithm is extended using RANSAC to be a robust estimate of the matrix F and determine the set of inliers corresponding to the estimated matrix. At each iteration, we randomly sample 8 point correspondences, which is the minimum required to estimate F, and use this estimation to identify points that fit well with the matrix, i.e. the inliers. When a certain number of inliers is reached or the maximum number of iterations is exceeded, the algorithm stops, and the matrix with the maximum number of inliers is chosen.

Moreover, to classify the points as inliers or outliers, the Sampson distance shown in Equation 9 is used, which provides a first-order approximation of the geometric distance.

$$\frac{\left(x_i'^T F x_i\right)^2}{(Fx_i)_1^2 + (Fx_i)_2^2 + (F^T x_i')_1^2 + (F^T x_i')_2^2} \tag{9}$$

Thus, given pairs of correspondences $(x \Leftrightarrow x')$ and an estimate of a fundamental matrix F, the Sampson distance can be computed for each points and a threshold t is used, where points with a smaller distance than t are considered inliers. The higher t is, the faster the algorithm will run at the expense of less quality. Note that minimizing this distance also minimizes the distance between the epipolar line of a point and its correspondence. Furthermore, the algorithm stops when either the maximum number of iterations set by default is reached, or when an estimation of the number of iterations required to find at least one set of points with all inliers is made. The number of iterations k can be dynamically computed while iterating and is derived by Equation 10, where n is the number of model parameters, w is the percentage of inliers, and p is the probability that RANSAC selects in some iteration only inliers (i.e. the probability to produce a useful result).

$$k = \frac{\log 1 - p}{\log 1 - w^n} \tag{10}$$

In function `ransac_fundamental_matrix()`, we set $n = 8$ and $p = 0.99$. Therefore, as the algorithm iterates, the fraction of inliers is computed and, if the computed number of iterations k yields a number smaller than the ones already performed, it suggests that with a probability of 99% the largest set of inliers has been seen and the algorithm can be stopped.

Figure 1 illustrates the use of RANSAC to filter outliers of keypoint matches between views related by a fundamental matrix F. The image on top shows all the matches found, while the image at the bottom shows the inliers matches after estimating F with RANSAC. This algorithm is used to robustly estimate the fundamental matrix by removing outliers in the correspondences and resulting in a more accurate representation of the inlier matches. This can be seen by comparing the number of matches shown in both images, with the image at the bottom showing a smaller number of matches (458) that are more likely to be accurate correspondences, and resulting in a geometric error of 302.60. This geometric error is a measure of the quality of the correspondences, as it is the average distance between the epipolar lines and the corresponding points in the other image. A low geometric error means that the correspondences are accurate and that the fundamental matrix estimated is a good representation of the true fundamental matrix.

2.3 Epipolar lines

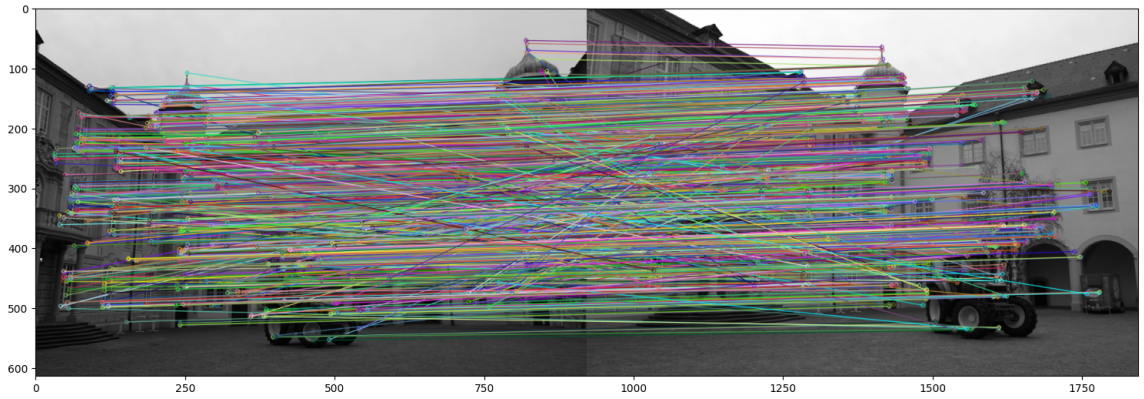
In this section we compute epipolar lines of points via the Fundamental matrix estimated with the robust 8-point algorithm. Epipolar lines are defined as the intersection of the epipolar plane with the image plane. A way to think of it is that a 3D point in the world is projected onto a line in each of the image planes. Given a matrix F relating two images, we can compute epipolar lines for points p and p' in projective space as

$$l = F\tilde{p}' \tag{11}$$

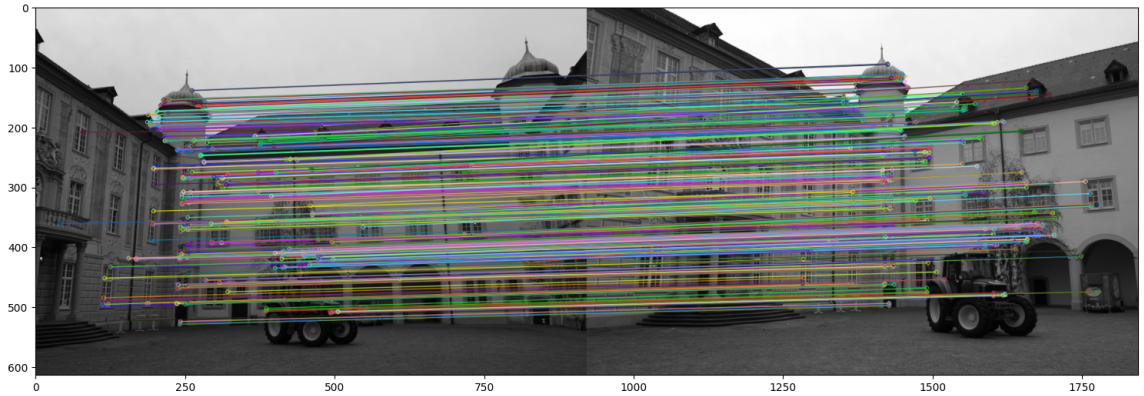
$$l' = F\tilde{p} \tag{12}$$

where l (l') is the epipolar line corresponding to p' (p).

Figure 2 illustrates the relationship between epipolar lines and corresponding points. The image shows epipolar lines in one image and corresponding points in another. We can see how the epipolar lines pass through the marked corresponding points. If the epipolar lines did not cross these points, then the fundamental matrix would not be correct.



(a) All keypoint matches found between images `Data/0000_s.png` and `Data/0001_s.png`.



(b) Inlier matches between images `Data/0000_s.png` and `Data/0001_s.png` after estimating matrix F with RANSAC.

Figure 1: Example of the use of RANSAC algorithm to remove outliers of keypoint matches between image views related by fundamental matrix.

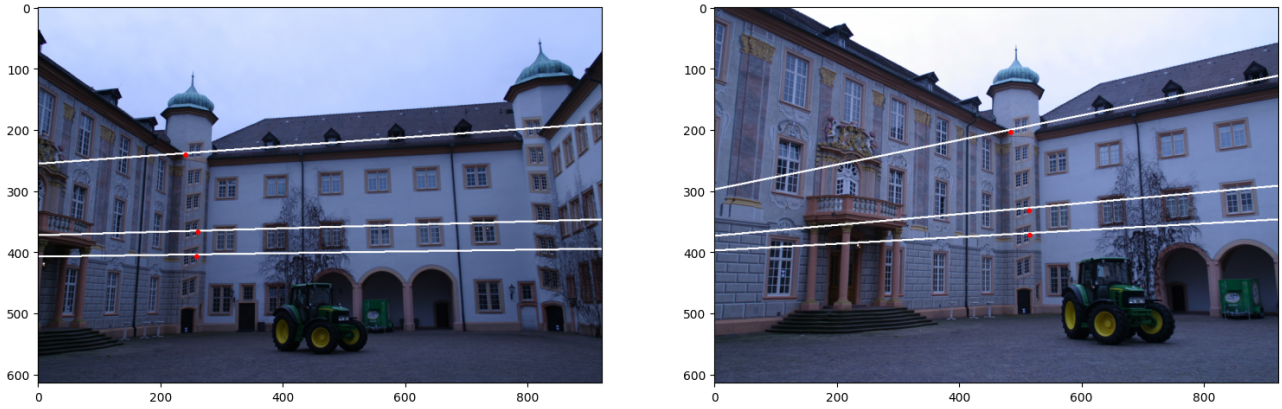


Figure 2: Relationship between epipolar lines Fx in one image and corresponding points x' in the other (images `Data/0000_s.png` and `Data/0001_s.png`).

3 Photo Sequencing

An application of the concepts explained above in the photo-sequencing problem is proposed. The solution relies on robust keypoint matching for recovering the temporal ordering for a series of photos taken from different uncalibrated cameras[1]. For We utilize two frames plus an initial one that we use as reference.

We start by computing the keypoint matching from a chosen reference frame and any other frame. We observe the there is a huge number of available keypoints (Fig. 4), and we need to find which one corresponds to the moving object.

[1] Photo Sequencing, Basha et. al, European Conference on Computer Vision 2012

Figure 3: Zoom on the moving van from the provided data in the reference image (left) and the following frames (right). The main objective for the task will be ordering the sequence (right) with respect to the moving object.

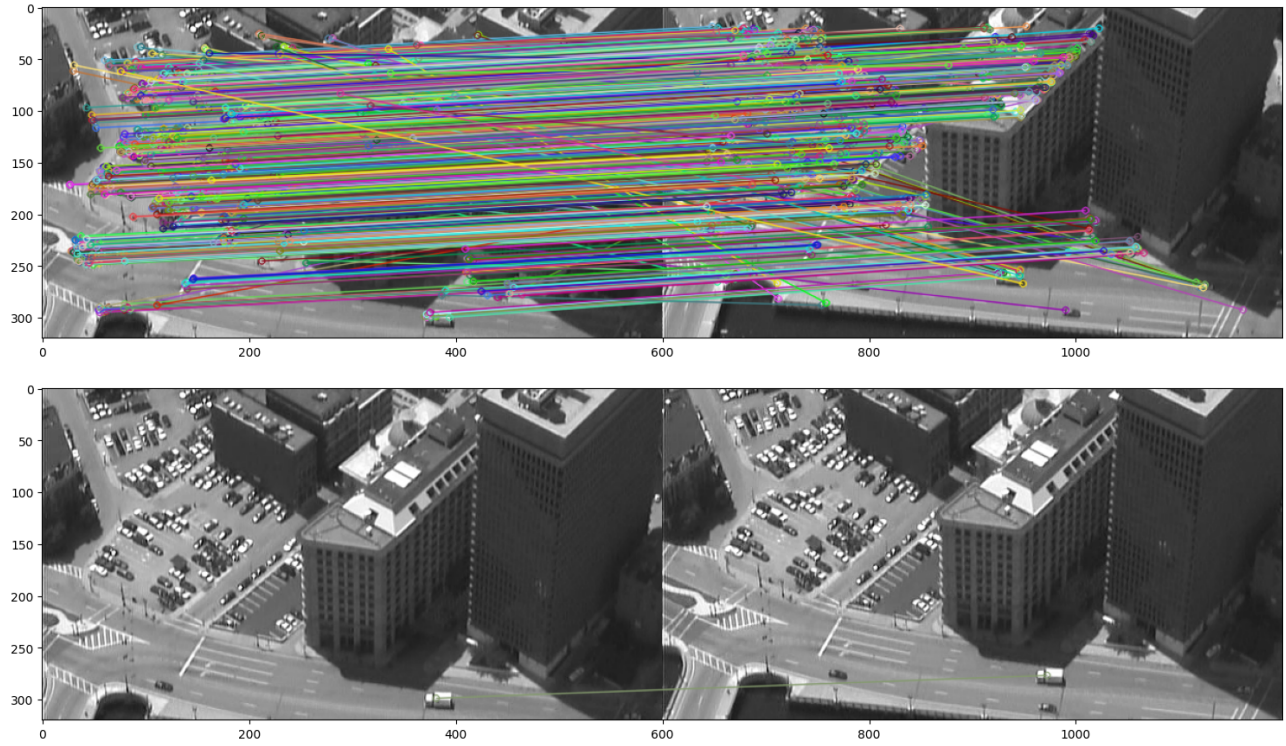


Figure 4: All available matches (top), most of which are not dynamic but static correspondences, and the target match (bottom) which will be used to track the moving object across the road.

To do so, we manually look for the target point by constraining the search to the small region where it has to be found: $height \in [250, 280]$ and $width \in [370, 400]$. See Figure 4, bottom right frame.

3.1 Computing the Trajectory and Estimating Positions

From the position of both matched keypoints belonging to the van, we can infer a line specified by both points. This will be considered as the trajectory of the moving object from $frame_{ref}$ to $frame_1$ (see Figure 5). The line approximating the straight trajectory is obtained as the cross product between the interest keypoints on the reference and a manually selected point on the second image which belongs to the trajectory.

With this, we can estimate epipolar lines from frames 2...n by solving the fundamental matrix through the correspondences as defined in Sec. 2 and projecting it back to the reference frame. Then we can estimate the different locations of the moving object in the reference image as the intersection of the (straight) trajectory and the corresponding epipolar lines.

The sequential ordination of the different frames can be resolved as the ordination of the intersection points through the trajectory; i.e. we can transform this trajectory into a 1-D temporal axis where points are placed along. This is the case because the van is being displaced along a straight line (both in the 3D world and in the image projection), and is not moving back and forth but only in one direction.

Additionally, we tested the fundamental matrix estimation both on SIFT and ORB with observable improvements Figure 6 when using SIFT. We can precisely approximate the location of the van in the reference image at each time instance.

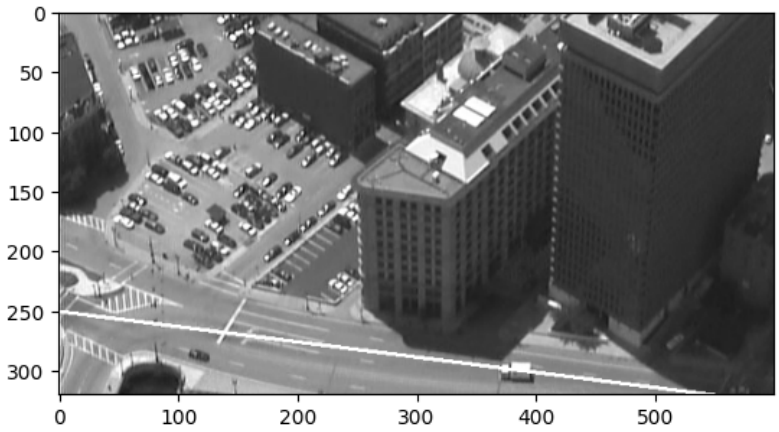


Figure 5: Estimated trajectory (white) for the moving object computed as the vector formed by both keypoints.

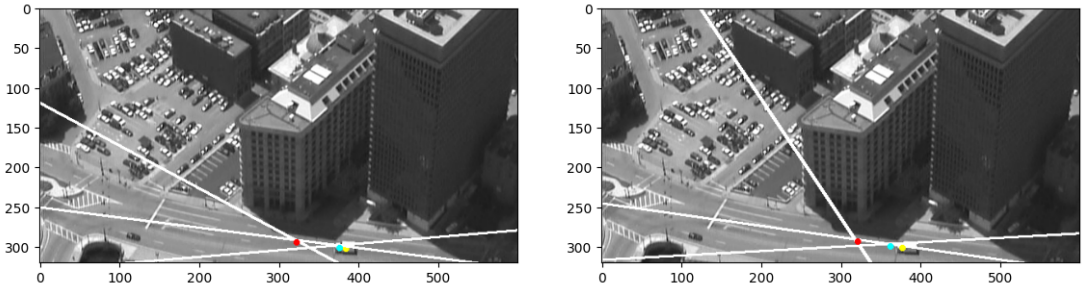


Figure 6: Intersection of the trajectory with the epipolar lines. We can observe that results are less noisy on SIFT considering our prior knowledge from Figure 3 and human intuition. As shown in the original images, the van should be closer to the first 2 time instances than is the case with ORB keypoints. Yellow, blue, and red points correspond to frames 1, 2, and 3 respectively.

4 Conclusions

In this practicum, we have delved into the topic of estimating the fundamental matrix between a pair of images, as well as some of its applications, such as obtaining epipolar lines and temporally ordering a set of images.

In a toy example in Sec. 2, we show that the 8-point algorithm can be effective for estimating the fundamental matrix. Nevertheless, this method suffers from noise in the measurements and in the procedure itself, so a more robust approach is needed. We show how to estimate the fundamental matrix in a robust way via RANSAC, building on the 8-point algorithm. It is clear that in real-world applications, those approaches that bypass the unconstrained and chaotic nature of real-scene images should work better. This includes, for example, robust methods based on RANSAC which allow filtering out noisy samples in the input data.

Additionally, we compute epipolar lines and visually and quantitively confirm that the our implementation yields a satisfactory estimation of the Fundamental matrix.

Finally, we show a real-world application in Sec 3 where despite the noisy nature of the data, we get good results. We managed to infer the ordering of a set of images based on the trajectory of a moving van and the epipolar lines of different points onto a reference image. We observe that some noise is introduced by the keypoint matching with ORB descriptors. We have seen that using more robust descriptors such as SIFT can help reduce the noise and improve the overall result.