Computer Vision

Project 3

Camera Calibration and Fundamental Matrix Estimation with RANSAC

Dristanta Das
Pratik Karmakar

Table of Contents

1. Part 1 — Estimating Camera Projection Matrix

The camera projection matrix

M= $\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$ maps homogeneous 3D coordinates to 2D image coordinates as such

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \simeq M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

Since we are given 20 ground truth correspondences in , we can solve for M by setting up a homogeneous system of 40 equations (2 for each corresponding pair of 3D -¿ 2D points):

$$\begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 \\ -u_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1Z_1 & -v_1Y_1 \\ -v_1Z_1 \\ . & . & . & . & . & . & . & . & . & . \\ . \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n \\ -u_nZ_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nZ_n & -v_nY_n \\ -v_nZ_n \end{pmatrix} * \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ u_n \\ v_n \end{pmatrix}$$

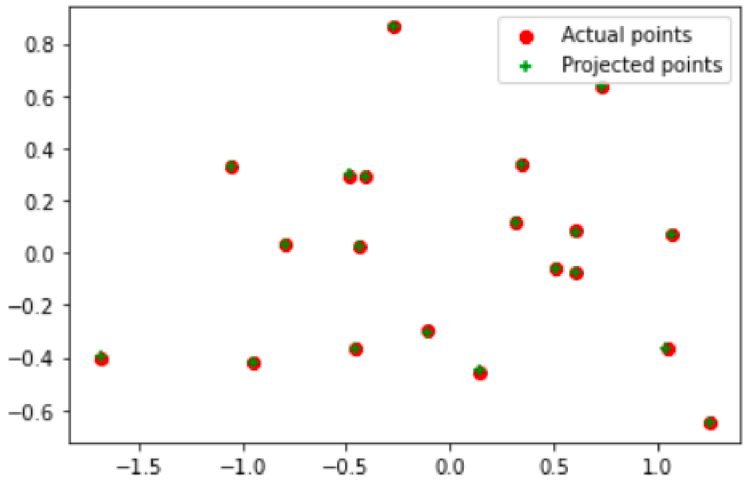
M has 12 elements, but since the scaling term m_{34} is arbitrary, we only need to solve for 11 variables. Computing the least-squares solution to the system of equations, the projection matrix for the 20 normalized points in Picture A is

$$\mathrm{M}_{NormA} = \begin{pmatrix} 0.7679 & -0.4938 & -0.0234 & 0.0067 \\ -0.0852 & -0.0915 & -0.9065 & -0.0878 \\ 0.1827 & 0.2988 & -0.0742 & 1.0000 \end{pmatrix}$$

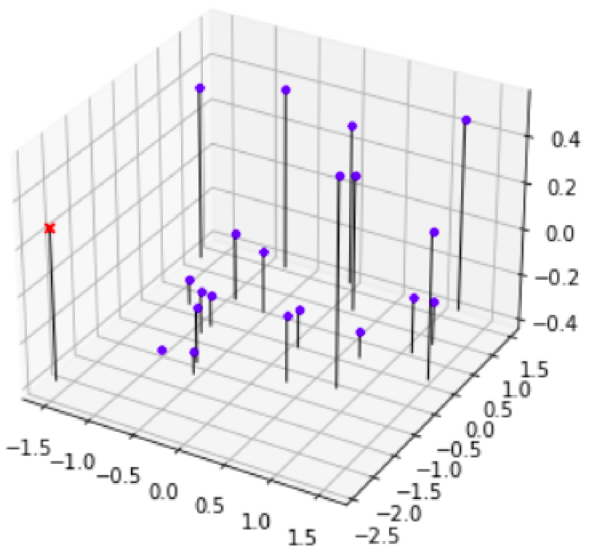
Writing M as the concatenation of a 3x3 matrix Q and a column matrix m4 as such $M = (Q|m_4)$ we can compute the camera center by solving

The total residual of this center estimate is 0.0445.

```
The projection matrix is
 [[ 0.45827554 -0.29474237 -0.01395746  0.0040258 ]
 [-0.05085589 -0.0545847  -0.54105993 -0.05237592]
 [ 0.10900958  0.17834548 -0.04426782  0.5968205 ]]
The total residual is 0.044549
```



```
The estimated location of the camera is <-1.5127, -2.3517, 0.2826>
```



Likewise, the camera projection matrix and camera center for Picture B is

$$M_B = \begin{pmatrix} -2.0466 & 1.1874 & 0.3889 & 243.7330 \\ -0.4569 & -0.3020 & 2.1472 & 165.9325 \\ -0.0022 & -0.0011 & 0.0006 & 1 \end{pmatrix}$$

$$C_B = \begin{pmatrix} 303.0967 \\ 307.1842 \\ 30.4223 \end{pmatrix}$$

$$Residual = 15.6217$$

2. Part 2 — Fundamental Matrix Estimation

The normalized eight-point algorithm is used to compute the fundamental matrix given point correspondences x = (u, v) and x' = (u', v') in the left and right images, respectively. Each point correspondence generates one constraint on the fundamental matrix F and must satisfy the epipolar constraint equation

x'$^T Fx = 0. Expanding the matrices out by multiplication, we obtain the following equation for n point corresp$

$$\begin{pmatrix} u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u'_nu_n & u'_nv_n & u'_n & v'_nu_n & v'_nv_n & v'_n & u_n & v_n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = \mathbf{0}$$

$$\Leftrightarrow \mathbf{Af} = 0.$$

where A is the n x 9 equation matrix, and f is a 9-element column vector containing the entries of the fundamental matrix F.

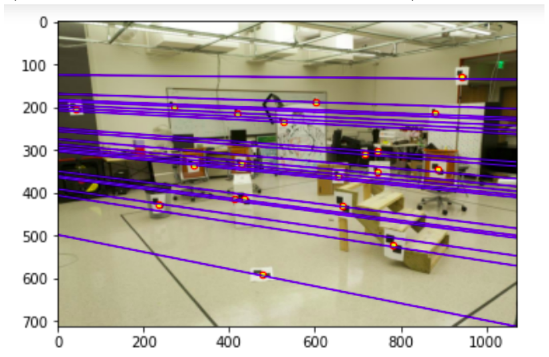
From here, the least-squares solution f is easily computed by performing singular value decomposition (SVD) on the matrix $A = U.D.V^T$. It is well-known that the vector f that minimizes $||Af||$ such that $||f|| = 1$ can be found along the column of V corresponding to the least singular value. Next, we rearrange the 9 entries of f to create the 3x3 fundamental matrix F. Then, we perform SVD on F to obtain

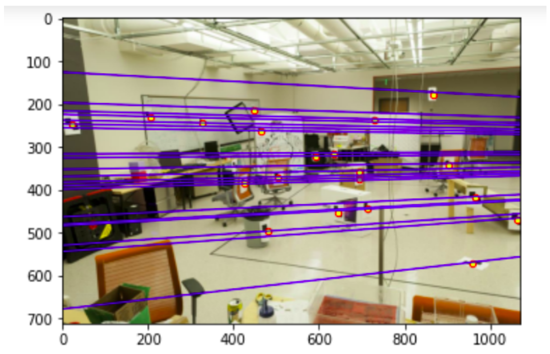
$F = U_f D_f V_f{}^T$ We set the smallest singular value of Df to 0 to create matrix D'f, thus reducing the rank of the matrix from 3 to 2, and from there we can recompute the rank-2 fundamental matrix as

$F = U_f D_f V_f{}^T$. Using the version of the eight-point algorithm without prior coordinate normalization to compute the fundamental matrix on the laboratory image pair, we obtain the following fundamental matrix F (truncated to 4 decimal places) and the epipolar lines for both images:

$$\text{F}_{unnormalized} = \begin{pmatrix} -5.3626 * 10^{-7} & 7.9036 * 10^{-6} & -0.0019 \\ 8.8354 * 10^{-6} & 1.2132 * 10^{-6} & 0.0172 \\ -9.0738 * 10^{-4} & -0.0264 & 0.9995 \end{pmatrix}$$

$$\approx \begin{pmatrix} -0.0000 & 0.0000 & -0.0019 \\ 0.0000 & 0.0000 & 0.0172 \\ -0.0009 & -0.0264 & 0.9995 \end{pmatrix}$$





However, in order to improve the accuracy of epipolar lines generated in Part III, we want to modify the 'vanilla' eight-point algorithm to recenter the point correspondences xi = (ui, vi) and x'i = (u'i, v'i) in both images to their respective centroids before proceeding to compute the least-squares solution for f. After recentering the image points, we must scale the points to be a fixed squared distance from the origin. The coordinate normalization steps can be summarized in the following steps:

1. Compute the centroid of all corresponding points in a single image:

$$\bar{\mathbf{u}} = \frac{1}{n}\sum_{i=1}^{n} u_i$$

$$\bar{\mathbf{v}} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

2. Recenter by subtracting the mean u and v coordinates from the original point correspondences to obtain

$$\tilde{\mathbf{u}} = \mathbf{u} - \bar{\mathbf{u}}$$
$$\tilde{\mathbf{v}} = \mathbf{v} - \bar{\mathbf{v}}$$

3. Define the scale term s and s' to be the average distances of the centered points from the origin in both the left and right images:

$$\mathbf{s} = \frac{\sqrt{2}}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{u}_i^2 + \tilde{v}_i^2\right)\right)^{1/2}}$$

$$\mathbf{s}' = \frac{\sqrt{2}}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{u'}_i^2 + \tilde{v'}_i^2\right)\right)^{1/2}}$$

4. Construct the transformation matrices Ta and Tb:

$$T_a = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\bar{\mathbf{u}} \\ 0 & 1 & -\bar{\mathbf{v}} \\ 0 & 0 & 1 \end{pmatrix}$$

$$T_b = \begin{pmatrix} s' & 0 & 0 \\ 0 & s' & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\bar{\mathbf{u}}' \\ 0 & 1 & -\bar{\mathbf{v}}' \\ 0 & 0 & 1 \end{pmatrix}$$

5. The normalized correspondence points can be computed from

$$\hat{\mathbf{x}}_{\mathbf{i}} = T_a \mathbf{x}$$
$$\hat{\mathbf{x}}'_{\mathbf{i}} = T_b \mathbf{x}$$

6. Solve for the fundamental matrix F by applying the eight-point algorithm on the normalized set of point correspondences computed in the previous step.

7. After obtaining the normalized fundamental matrix Fnorm, retrieve the fundamental matrix in the original coordinate frame using the following formula

$F_{orig} = T_b^T * F_{norm} * T_a$ Using the normalized eight-point algorithm on the laboratory image pair, we get

and the fundamental matrix (truncated to 4 decimal places)

$$F_{normalizedPts} = \begin{pmatrix} -0.0000 & 0.0000 & -0.0005 \\ 0.0000 & -0.0000 & 0.0037 \\ -0.0000 & -0.0051 & 0.1192 \end{pmatrix}$$

3. Part 3 — Fundamental Matrix with RANSAC

In practical situations, we will not be given ground truth correspondences between two images, so there must be a way to compute interest points and predict matches between stereo image pairs. We use the VLFeat library to generate SIFT descriptors around detected interest points and assign correspondences based on the nearest neighbor ratio test. In project 2, this method alone produced good results on some occasions, but on other image pairs (most notably the Episcopal Gaudi image set), it failed because there was no way to filter outliers from our results.

We now make use of RANSAC algorithm in tandem with the normalized eight-point algorithm to separate outliers. This addition allows us to estimate the best possible fundamental matrix with the greatest number of inliers.

RANSAC is an iterative algorithm that nondeterministically fits a model to a random sample of points taken from the dataset. As the number of permitted iterations increases, the more likely we are to find a fundamental matrix free of outliers. We can roughly estimate the number of iterations k needed to obtain a solution where some iteration of RANSAC selects only inliers with some probability p, using the following equation:

$\mathbf{k} = \frac{log(1-p)}{log(1-w^n)}$ where w is the proportion of inliers in the data, and n is the number of points required to estimate the model.

So, for example, if we assume that 50 % of the corresponding points returned in the SIFT description and matching phases are inliers, the expected number of iterations to run RANSAC in order to get a sample free of outliers with success probability 99

$k = \frac{log(1-.99)}{log(1-.5^8)} \approx 1177$

For a pair of points in a stereo image to be considered a match, it must lie on the same epipolar line as well as satisfy the epipolar constraint. Programmatically, we define an absolute error threshold of .001 such that any pair (x, x') that satisfies the equation

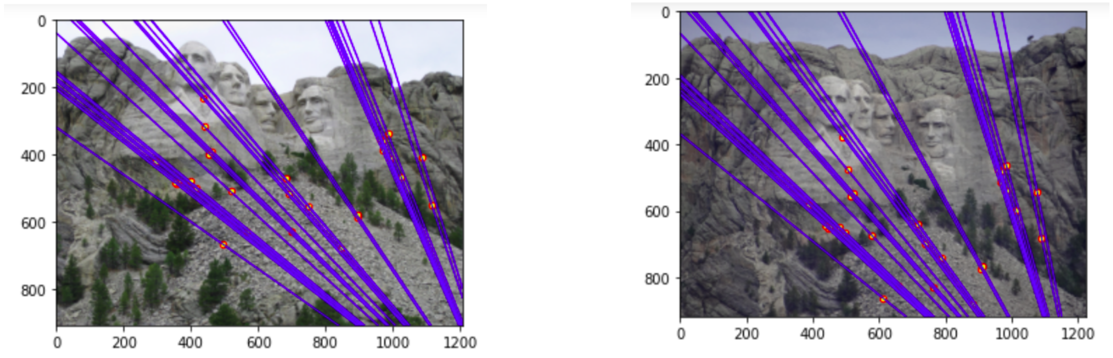
$x'^T Fx = 0$ is classified as an 'inlier' in the RANSAC algorithm.

4. Results:

We display up to 35 random correspondences obtained after running RANSAC to compute the fundamental matrix.

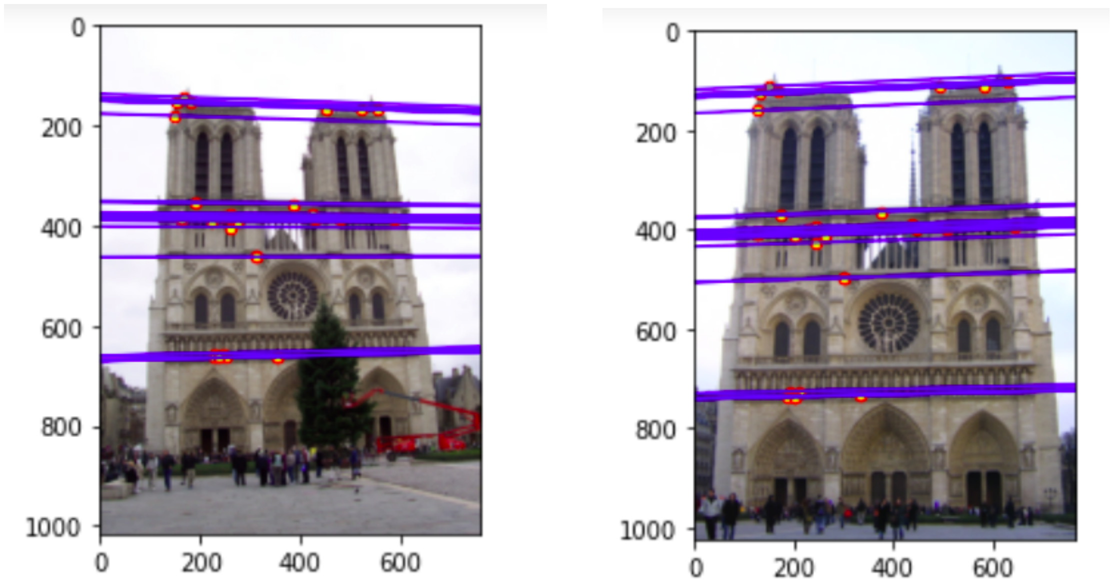
4.1 Rushmore (normalized)

Ran 1000 iterations, error threshold = .001, no coordinate normalization.



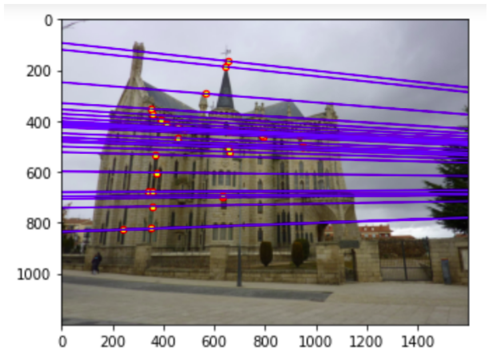
4.2 Notre Dame (normalized)

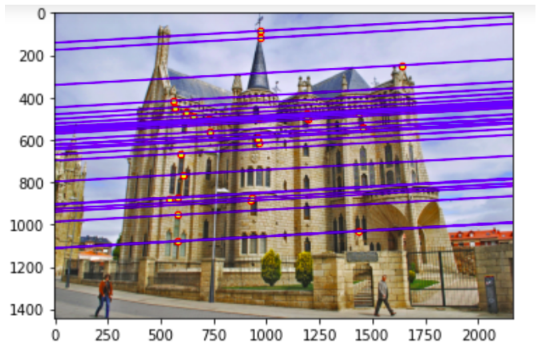
Ran 1000 iterations, error threshold = .001, no coordinate normalization.



4.3 Gaudi (normalized)

Ran 1000 iterations, error threshold = .001, no coordinate normalization.





4.4 Woodruff (normalized)

Ran 1000 iterations, error threshold = .001, no coordinate normalization.