Computer Vision : Project 3

Camera Calibration and Fundamental Matrix Estimation with RANSAC

Srijan Mallick
Alimpan Barik

Semester 2
M.Sc Big Data Analytics, RKMVERI

May 2021

Contents

1 Introduction

Camera Calibration is nothing but estimating the parameters of a camera which are required to determine an accurate relationship between a 3D point in the real world and its corresponding 2D projection (pixel) in the image captured by that calibrated camera.

We need to consider both intrinsic parameters like focal length, aspect ratio, optical center, and radial distortion coefficients of the lens etc., and extrinsic parameters like rotation and translation of the camera with respect to some real world coordinate system.

In this project, we shall try to estimate the **camera projection matrix**, which maps 3D world coordinates to image coordinates, as well as the **fundamental matrix**, which relates points in one scene to epipolar lines in another. The camera projection matrix and the fundamental matrix can each be estimated using point correspondences. To estimate the projection matrix—intrinsic and extrinsic camera calibration—the input is corresponding 3d and 2d points. To estimate the fundamental matrix the input is corresponding 2d points across two images. We shall start out by estimating the projection matrix and the fundamental matrix for a scene with ground truth correspondences. Then we'll move on to estimating the fundamental matrix using point correspondences from SIFT and **RANSAC**.

2 Some Basic Concepts

When we take an image using pin-hole camera, we lose an important information - the depth of the image. In order to get depth information, we need to use more than one camera.

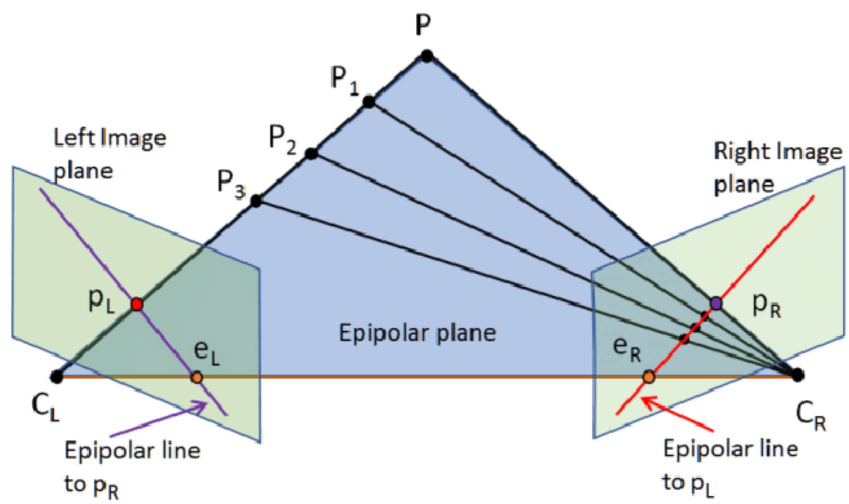


Figure 2.1

The projection of the different points on C_LP form a line on right image plane. We call it **epipolar line** corresponding to the point p_L. It means, to find the point p_L on the right image, we must restrict our search along this epipolar line. This is called **Epipolar Constraint**. Similarly all points will have its corresponding epipolar lines in the other image. The plane PC_LC_R is called **Epipolar Plane**.

C_L and C_R are the camera centers. The line joining these camera centers is called **Baseline**. The point where the Baseline interesects the Image plane is called **epipole**. To find epipoles and epipolar lines, we need two more ingredients, **Fundamental Matrix (F)** and **Essential Matrix (E)**. Essential Matrix contains the information about translation and rotation, which describe the location of the second camera relative to the first in global coordinates.

3 Projection Matrix

The camera projection matrix maps homogeneous 3D coordinates to 2D image coordinates such that -

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

Where M is camera projection matrix.

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$$

We can solve for M by setting up a homogeneous system -

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

M has 12 elements, but since we are setting $m_{34} = 1$, we only need to solve for 11 variables. We solved this using least squares with $np.linalg.lstsq()$ function.

$$M = \begin{pmatrix} 0.76785834 & -0.49384797 & -0.02339781 & 0.00674445 \\ -0.0852134 & -0.09146818 & -0.90652332 & -0.08775678 \\ 0.18265016 & 0.29882917 & -0.07419242 & 1. \end{pmatrix}$$

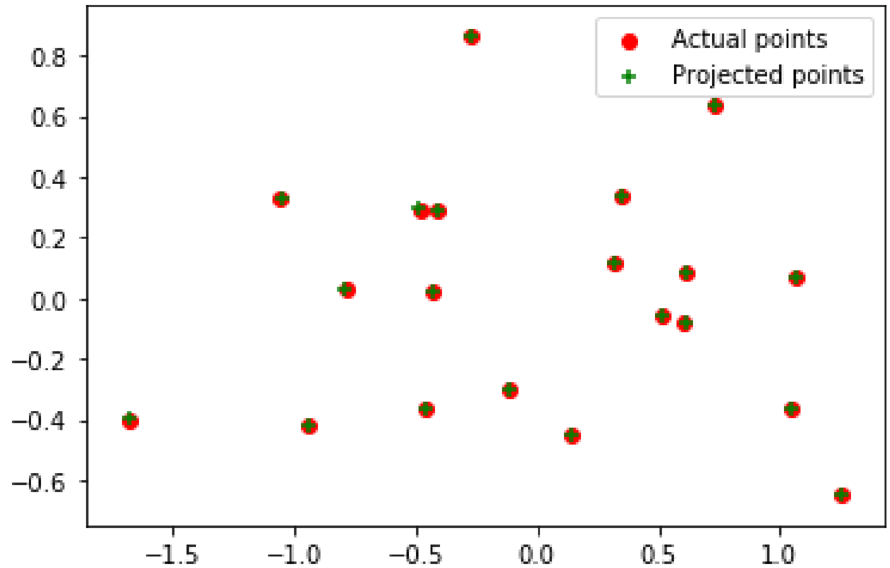


Figure 3.1

4 Calculating the Camera Centre

Now we shall see how to calculate the center of the camera from the camera projection matrix, derived in the previous step. Camera projection matrix M is given by

$$M = \begin{pmatrix} 0.76785834 & -0.49384797 & -0.02339781 & | & 0.00674445 \\ -0.0852134 & -0.09146818 & -0.90652332 & | & -0.08775678 \\ 0.18265016 & 0.29882917 & -0.07419242 & | & 1. \end{pmatrix}$$

As shown in the matrix above we have partitioned the matrix M into two parts. The left part i.e. the 3*3 submatrix is Q(say) and the 3*1 column matrix in the right of the partition is m(say). Then we can compute the center of the matrix by the following formula-

$$cc = -Q^{-1}m$$

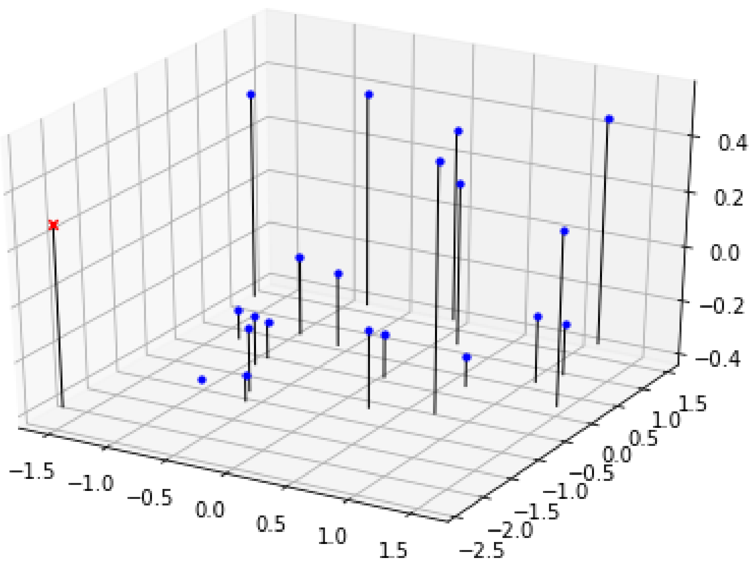


Figure 4.1: Camera center

5 Estimating the Fundamental Matrix

The fundamental matrix is a matrix that finds the duality between line and points between a pair of images. The fundamental matrix F satisfies-

$$x^T F x^{'} = 0$$

where x and $x^{'}$ are two points on the two images respectively. In matrix form we can write-

$$\begin{pmatrix} \mathrm{u}^{'} & v^{'} & 1 \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0$$

or,

$$\begin{pmatrix} \mathrm{u}^{'} & v^{'} & 1 \end{pmatrix} \begin{pmatrix} f_{11}u & f_{12}v & f_{13} \\ f_{21}u & f_{22}v & f_{23} \\ f_{31}u & f_{32}v & f_{33} \end{pmatrix} = 0$$

or, $$f_{11}\mathrm{uu}^{'} + \mathrm{f}_{12}\mathrm{vu}^{'} + \mathrm{f}_{13}\mathrm{u}^{'} + \mathrm{f}_{21}\mathrm{uv}^{'} + \mathrm{f}_{22}\mathrm{vv}^{'} + \mathrm{f}_{23}\mathrm{v}^{'} + \mathrm{f}_{31}\mathrm{u} + f_{32}\mathrm{v} + f_{33} = 0$$

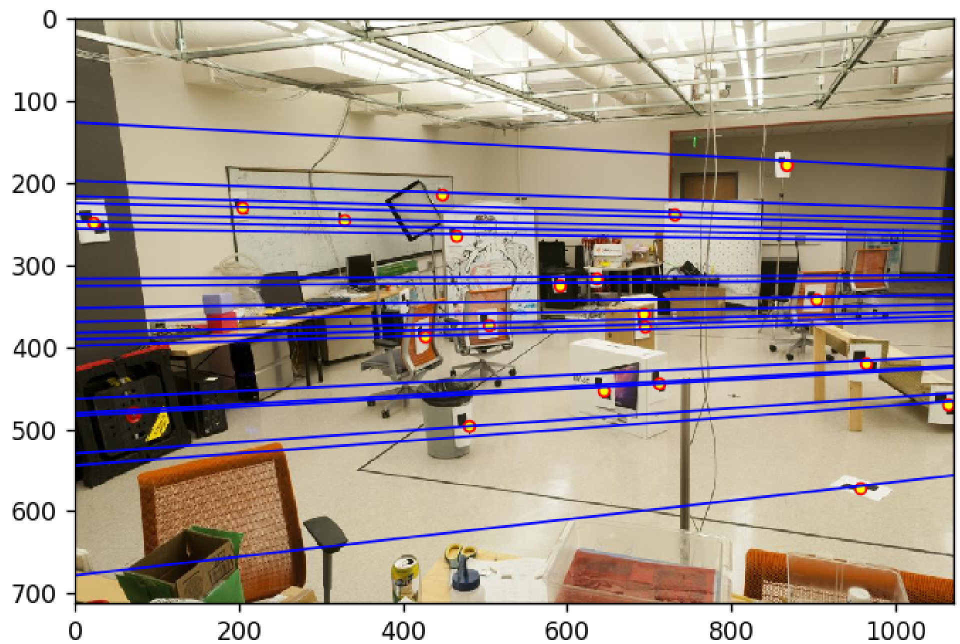
If there are n such pair of points then we will get n such equations.

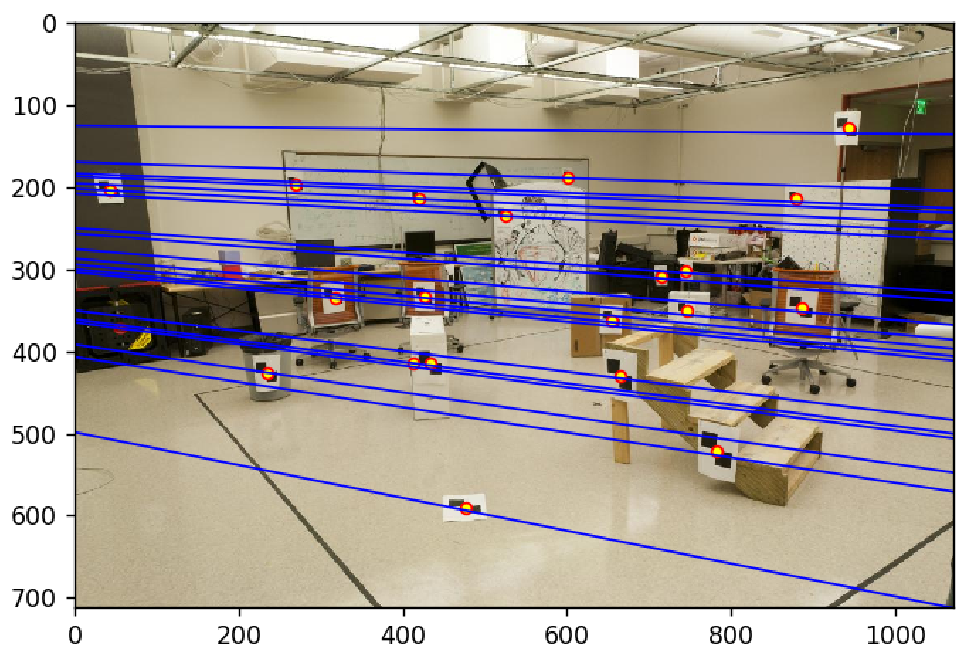
$$\begin{pmatrix} u_1'u_1 & u_1'v_1 & u_1' & v_1'u_1 & v_1'v_1 & v_1' & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n'u_n & u_n'v_n & u_n' & v_n'u_n & v_n'v_n & v_n' & u_n & v_n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = \mathbf{0}$$

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

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. We solve f in the same way of that of the Camera projection matrix. However, the fundamental matrix is a rank 2 matrix. As such we must reduce its rank. In order to do this we can decompose F using singular value decomposition into the matrices $UDV^{'} = F$. We can then estimate a rank 2 matrix by setting the smallest singular value in D to zero thus generating D_2 .The fundamental matrix is then easily calculated as $F = UD_2V^{'}$

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:


0
100
200
300
400
500
600
700
0
200
400
600
800
1000


0
100
200
300
400
500
600
700
0
200
400
600
800
1000

6 Fundamental Matrix with RANSAC

Now we shall find the best fundamental matrix using RANSAC on potentially matching points. We shall find best eight keypoints using RANSAC. We ran our algorithm on pictures of Mount Rushmore, Notre Dame, Episcopal Gaudi and Woodruff and got these results :

6.1 Mount Rushmore

This pair is easy, and most of the initial matches are correct. The base fundamental matrix estimation without coordinate normalization will work fine with RANSAC.

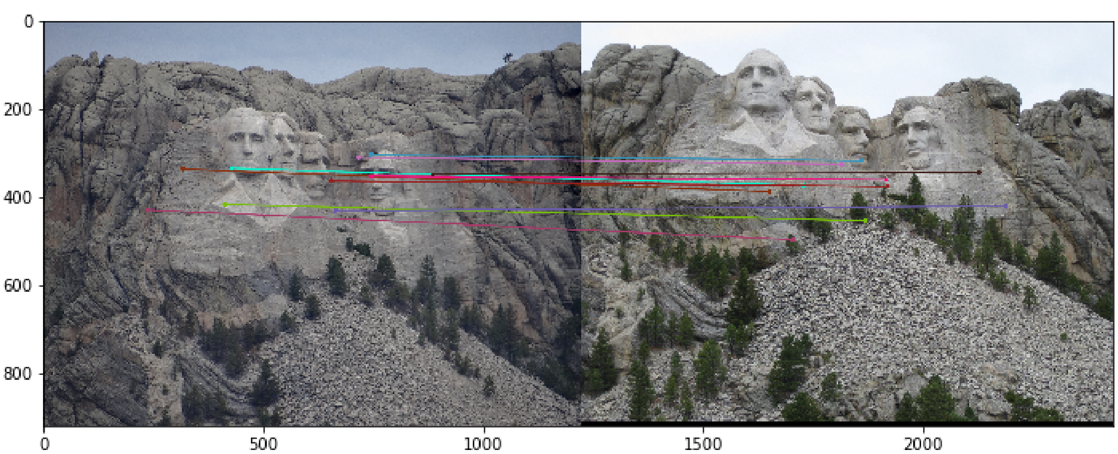


Figure 6.1: Matched Images

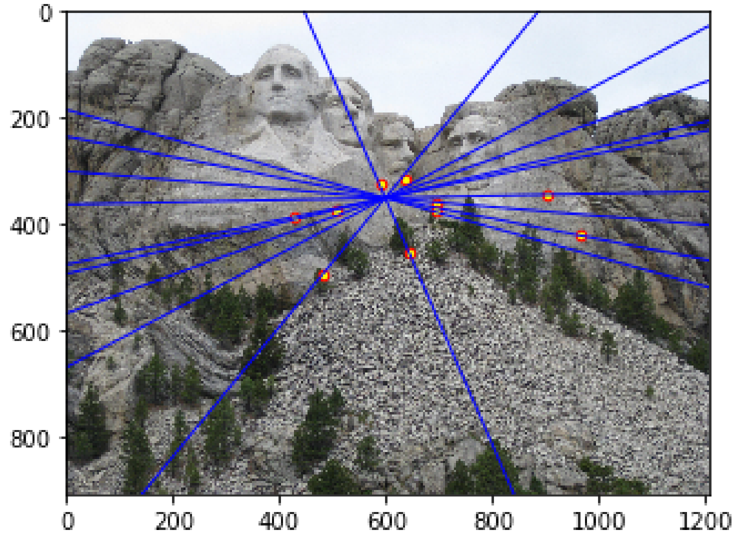


Figure 6.2: Epipolar Lines in first image

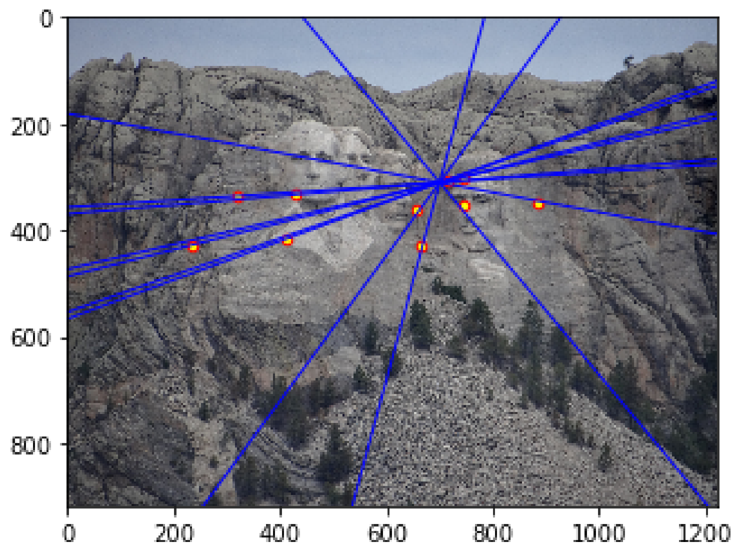


Figure 6.3: Epipolar Lines in second image

6.2 Notre Dame

This pair is difficult because the keypoints are largely on the same plane. Still, even an inaccurate fundamental matrix can do a pretty good job of filtering spurious matches.

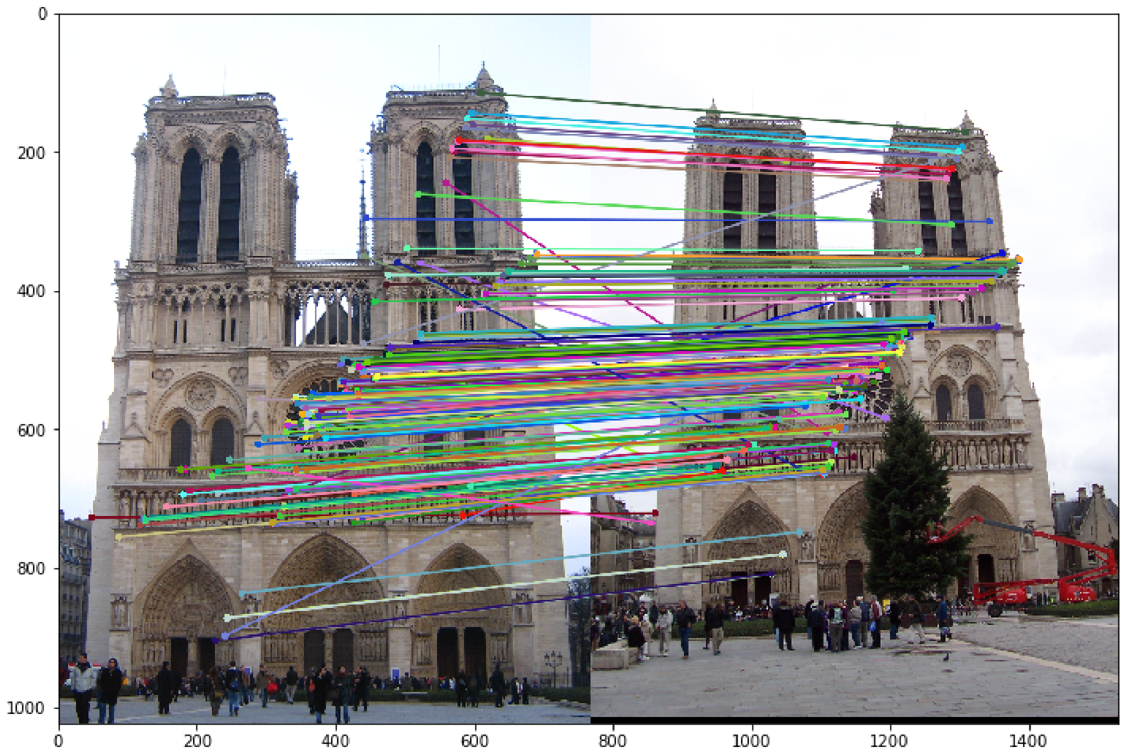


Figure 6.4: Matched Images

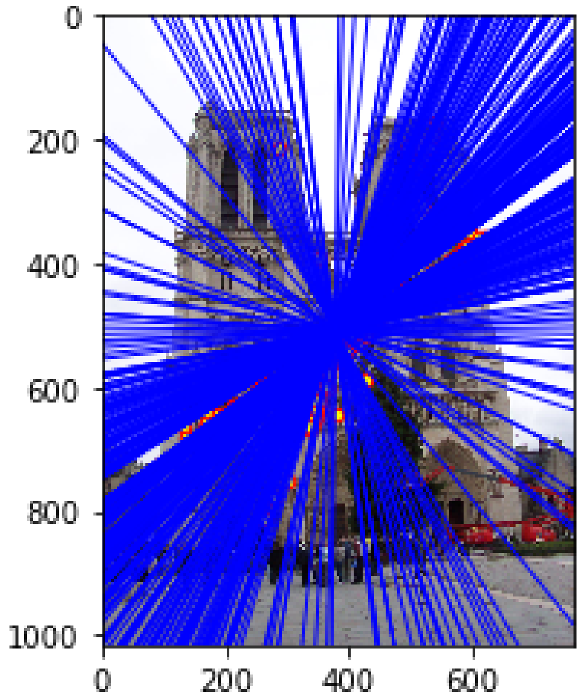


Figure 6.5: Epipolar Lines in first image

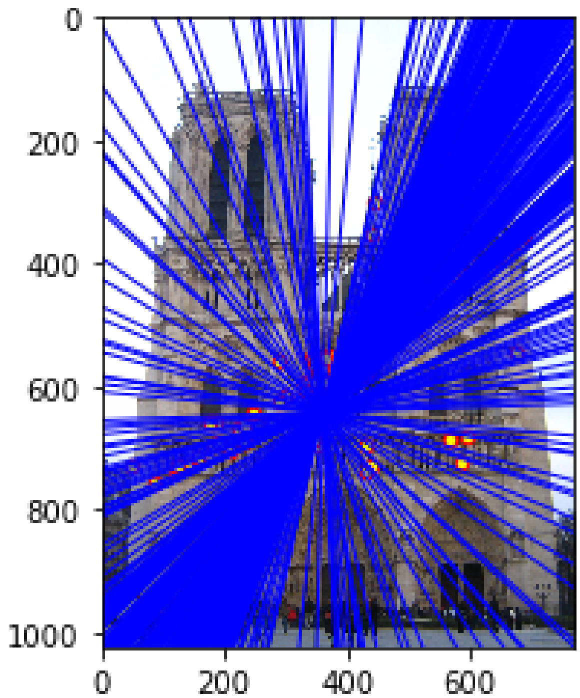


Figure 6.6: Epipolar Lines in second image

6.3 Episcopal Gaudi

This pair is difficult and doesn't find many correct matches unless you run at high resolution, but that will lead to tens of thousands of ORB features, which will be somewhat slow to process. Normalizing the coordinates seems to make this pair work much better.

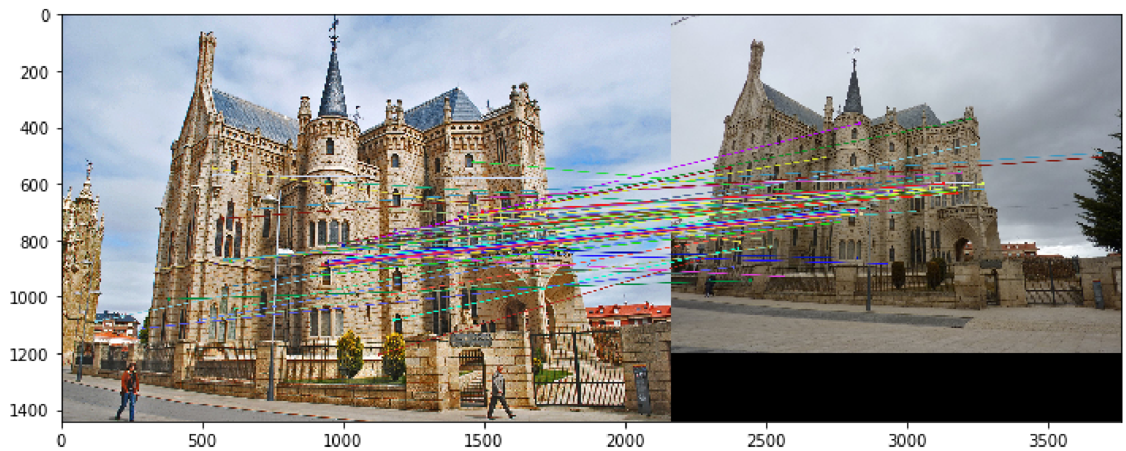
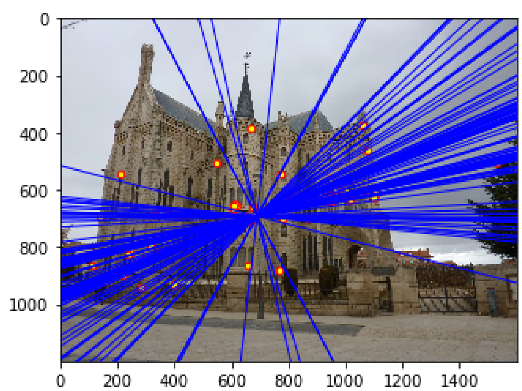
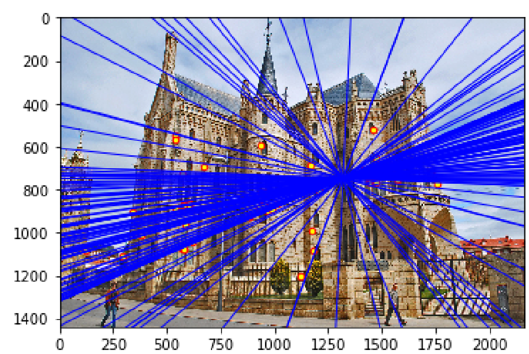


Figure 6.7: Matched Images



(a) Epipolar Lines in the first image



(b) Epipolar Lines in the second image

6.4 Woodruff

This pair has a clearer relationship between the cameras (they are converging and have a wide baseline between them). The estimated fundamental matrix is less ambiguous and you should get epipolar lines qualitatively similar to part 2 of the project.

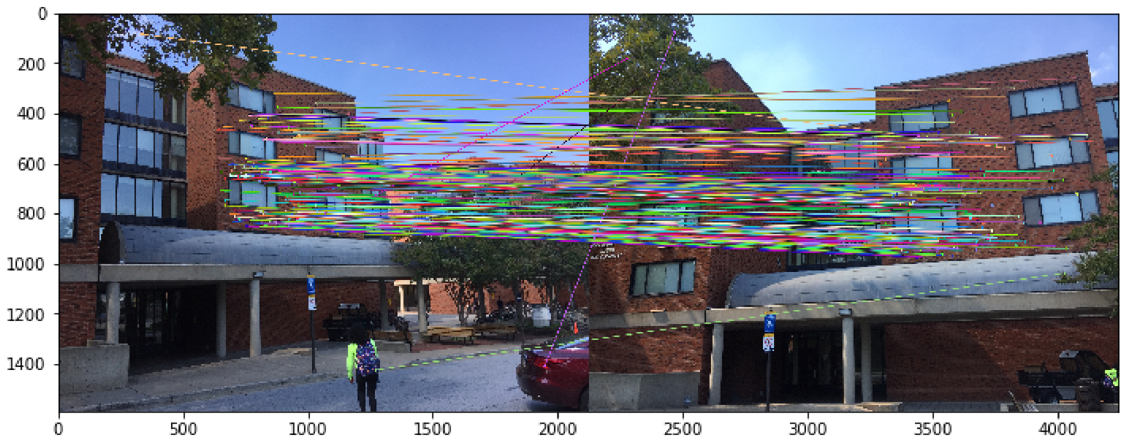
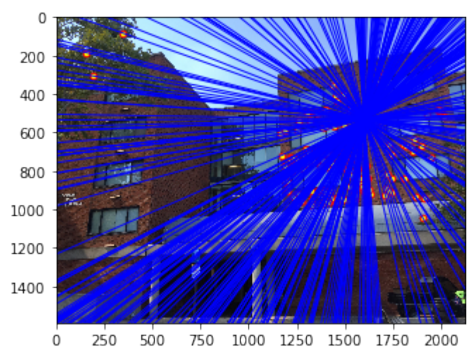
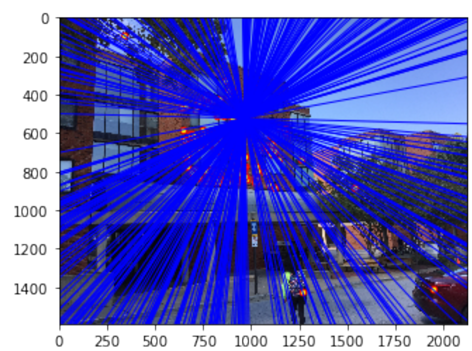


Figure 6.9: Matched Images



(a) Epipolar Lines in the first image



(b) Epipolar Lines in the second image

7 Conclusion

In this project we estimated the camera projection matrix or calibration matrix, which maps 3D world coordinates to image coordinates, as well as the fundamental matrix, which relates points in one scene to epipolar lines in another. You started out by estimating the projection matrix and the fundamental matrix for a scene with ground truth correspondences. Then we moved on to estimating the fundamental matrix using point correspondences from ORB, which is an alternative to SIFT. By using RANSAC to find the fundamental matrix with the most inliers, we could filter away spurious matches and achieve near perfect point to point matching.

8 References

- https://docs.opencv.org/master/da/de9/tutorial_py_epipolar_geometry.html
- Hata, K., and Savarese, S., *"CS231A Course Notes 3: Epipolar Geometry"*