COMP90086 Computer Vision Assignment 3

Einon McGrory-Perich 992697

October 2022

1 Introduction

The fundamental matrix is a construct that links the geometry of two related views, and can allow for the transformation between two scenes to be identified. This can be useful in a number of fields, such as simultaneous localisation and mapping (SLAM) which can be used for autonomous systems, and digital 3d reconstructions. This matrix can be constructed using the 8 point algorithm. The fundamental matrix can then be used to estimate the change of the camera between the two views, without a prior understanding of the camera parameters. If these are identified, then the essential matrix can be computed, and decomposed into the transformations that had occurred.

2 Feature Identification and Correspondences

To construct this fundamental matrix at least 8 key features in both images are required. These key features can be derived by utilising a feature extraction method, such as the scale invariant feature transform (SIFT) which identifies points of interest. An example of these features are shown in Figure 1.

It can be seen that a number of noticeable key points are identified, demonstrated by the circles of various size. Using an appropriate matching library, such as the Fast Library for Approximate Nearest Neighbours (FLANN), these key points can be related between images in a fast and high dimensional space. The FLANN used a KD Tree with 5 trees, performed 50 checks, and used the 2 nearest neighbours. A visual representation of the matched features is shown in Figure 2.

Unfortunately, a limitation of the nearest neighbour calculation included in FLANN is that it does not automatically compute one to one feature matching. This can be problematic as the duplicate values in one image can cause problems within the 8 point algorithm as it makes the model rank deficient. A manual enforcement of this can be made which only allows for one instance of each key point in either image.

When identifying key point correspondences, Lowe's ratio test is used to filter out poor matches. This test required that the ratio of the distances between two key points was less than 0.7. If this was increased it was found that poor matches were included, and too low might have too few correspondences for some images. Features that satisfied this ratio were separated and used for the 8 point algorithm.

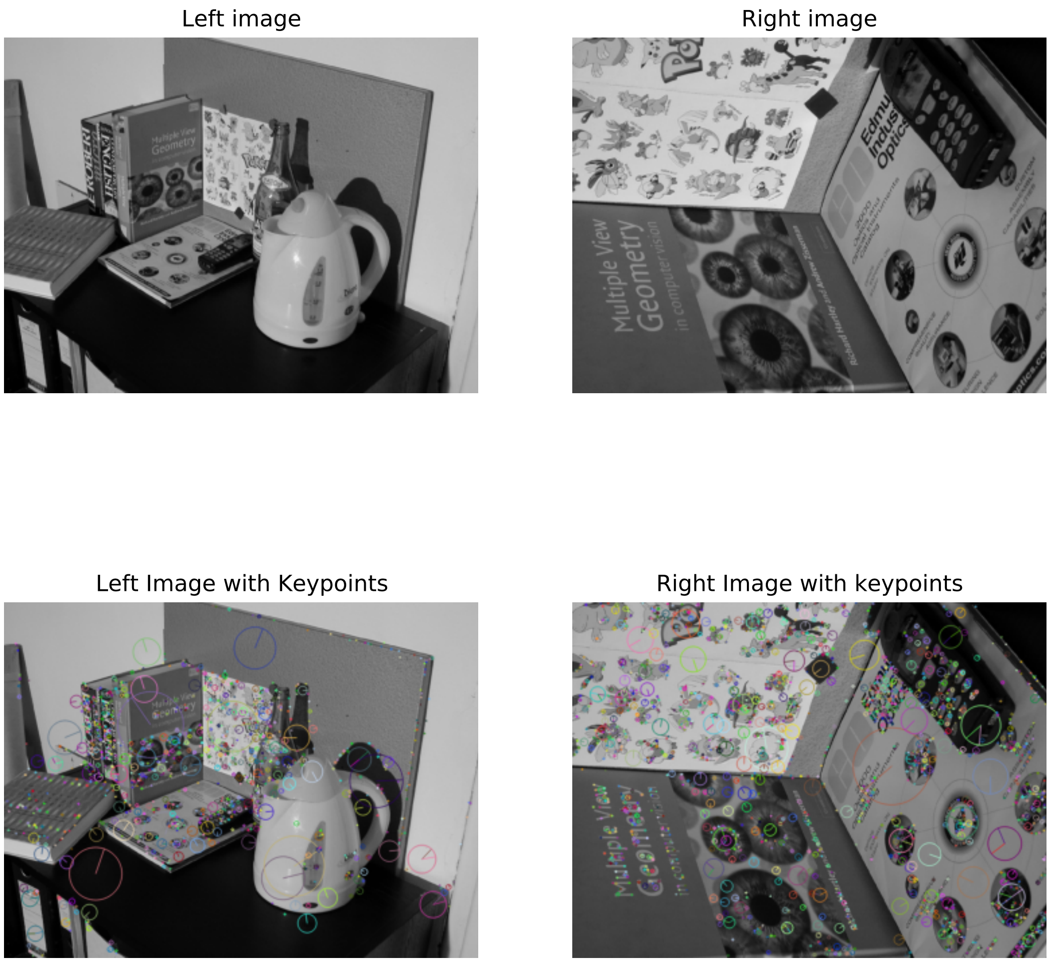

Figure 1: SIFT Key Points on Sample Transformed Images

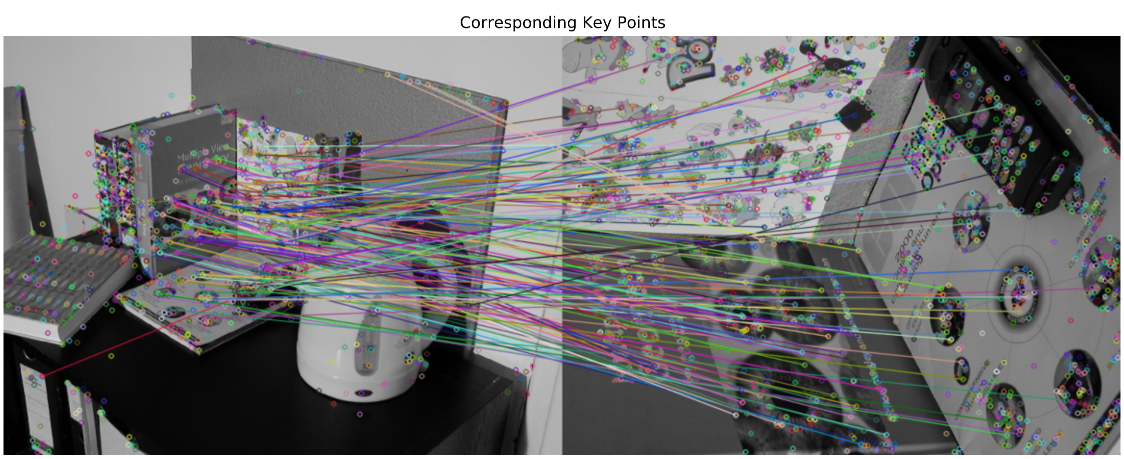

Figure 2: Matched Features

3 8 Point Algorithm

The 8 point algorithm is a process that can be used to construct the fundamental matrix, F, from 8 matched features in two images portraying a transformed scene.

3.1 Shift and Scale of Coordinate Frame

The image coordinate frame starts in the top left corner, and increments by pixels. This suggests that pixels in the bottom right corner would have the largest magnitude, and would therefore have greater impact in the calculations than those closer to the origin. As such any pixels need to be shifted, and scaled down to a similar magnitude. To center the n matched points in the image, the mean $\bar{\mu}$ in each direction and image can be calculated and subtracted from every point to form the centered points $(\hat{x}, \hat{y})$. This value can then be scaled down by using (1) for both images, as done in Hartley's normalised 8 point algorithm [1].

$$s_{l,r} = (\frac{1}{2n}\sum_{1}^{n}(\hat{x}^2 + \hat{y}^2))^{-1/2} \tag{1}$$

This total transformation from image coordinates to normalised coordinates is shown in the homogeneous transformation matrices, T_l for the left image and T_r for the right image in equation (2).

$$T_l = \begin{bmatrix} s_l & 0 & 0 \\ 0 & s_l & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\bar{\mu_{xl}} \\ 0 & 1 & -\bar{\mu_{yl}} \\ 0 & 0 & 1 \end{bmatrix} \qquad T_r = \begin{bmatrix} s_r & 0 & 0 \\ 0 & s_r & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\bar{\mu_{xr}} \\ 0 & 1 & -\bar{\mu_{yr}} \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

With all the feature points normalised, the Fundamental Matrix can be constructed.

3.2 Design Matrix

The design matrix can be used to represent the coefficients of F for the system of linear equations produced when the correspondence between the two views, as shown in (3), is expanded. By sampling 8 random matching points from the two images, an 8 by 9 matrix is produced, which is ensured to have a rank of 8 when no duplicate points are included. The left image points are represented by $\hat{p}$, the right image represented by $\hat{q}$, and the Fundamental Matrix by $\hat{F}$, all in normalised coordinate form.

$$\hat{q}\hat{F}\hat{p} = \begin{bmatrix} q1 & q2 & 1 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} p1 \\ p2 \\ 1 \end{bmatrix} = 0 \tag{3}$$

The null space of this design matrix can then be computed by taking the singular value decomposition (SVD) and taking the eigenvector corresponding to the smallest eigenvalue of the system. This results in the parameter values for a draft F matrix shown in (3).

As this vector attempts to make one parameter of the fundamental matrix 0, a SVD of F can be taken and the smallest singular value forced to 0, before being reconstructed. This will ensure that any epipolar lines constructed converge to a single point representing the camera location, instead of having a very small spread. This is shown in Figure 3, where the image computed with the F not rank deficient has small spread at the point of convergence.

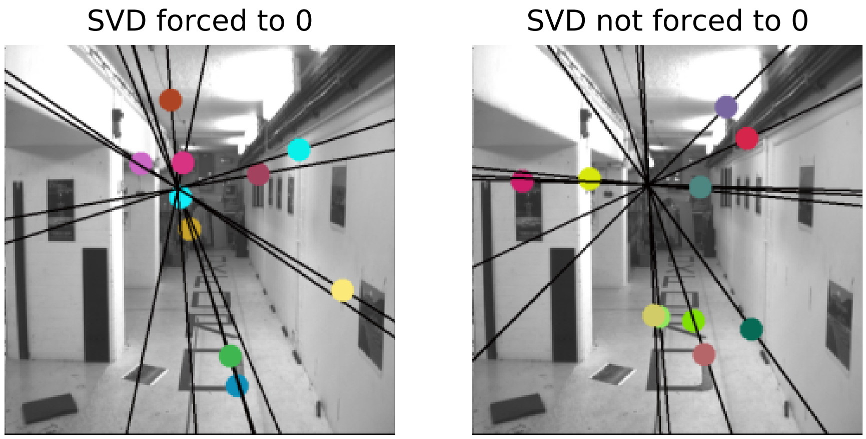

Figure 3: Comparison of Epipolar Lines when F Forced to be Singular

The Fundamental Matrix has now been constructed, but only within the normalised coordinate system. As it is easiest and intuitive to design for thresholds within the image coordinate frame, this matrix can be transformed back with (4).

$$F = T_r^T \hat{F} T_l \tag{4}$$

This Fundamental Matrix can now be used to construct epipolar lines. An epipolar line can be constructed by taking a point from one image and taking the dot product with the fundamental matrix, as shown in (5) which demonstrates the construction of a right epipolar line from a left image point.

$$l = Fp = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} p1 \\ p2 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \implies 0 = ax + by + c \tag{5}$$

The distance between this constructed epipolar line and the corresponding point can then be calculated, to determine if it is an inlier or outlier. The distance calculation is shown in (6).

$$d = \frac{|aq_1 + bq_2 + c|}{\sqrt{a^2 + b^2}} \tag{6}$$

The threshold that determined whether a point was an inlier or outlier was initially chosen arbitrarily at 2 pixels. This was problematic as certain images such as the rotunda and graffiti had very few numbers of matched features. To accommodate for this a dynamic threshold approach was taken depending on the number of corresponding key points. If there were more than 100 points, a distance less than 1 was required to be an inlier, between 50 and 100 a distance of 2 was required, and anything less than 50 had a distance of 3.

3.3 RANSAC Loop

This algorithm was then implemented in a RANSAC loop and the fundamental matrix that produced the highest number of inliers was chosen. To determine the number of trials N the loop should run, the probability of at least one sample was estimated using (7). The probability of an outlier or inlier, e, was chosen at 0.5, the required probability p was set to 0.99 and the number of required points was 8. This resulted in at least 1176 loops.

$$N = \frac{ln(1-p)}{ln(1-(1-e)^s)} \tag{7}$$

F was finally recalculated using all of the normalised inliers from the best performing result of the RANSAC loop.

4 Results

To evaluate the performance of the calculated fundamental matrix using the 8 point matrix, a comparison was made to OpenCV's *findFundamentalMat()* function. It was found that in most situations the epipolar lines were similar, suggesting that the 8 point algorithm performs to a satisfactory standard.

Figure 4 shows how the epipolar lines produced from OpenCV's Fundamental matrix are comparable to the calculated epipolar lines. As these epipolar lines converge towards the center of the image, it is expected that only a forwards transformation has been applied, which intuition would agree with. It can also be seen that for the corridor image, a high number of inliers were found from the matched points.

Similarly, the bookshelf image in figure 5 also generates converging epipolar lines in the first image. Epipolar geometry suggests that this convergent point should be the location of the camera in the second image. The two images demonstrate this as the second image would be expected to have taken the image from the perspective of the epipole in the first image.

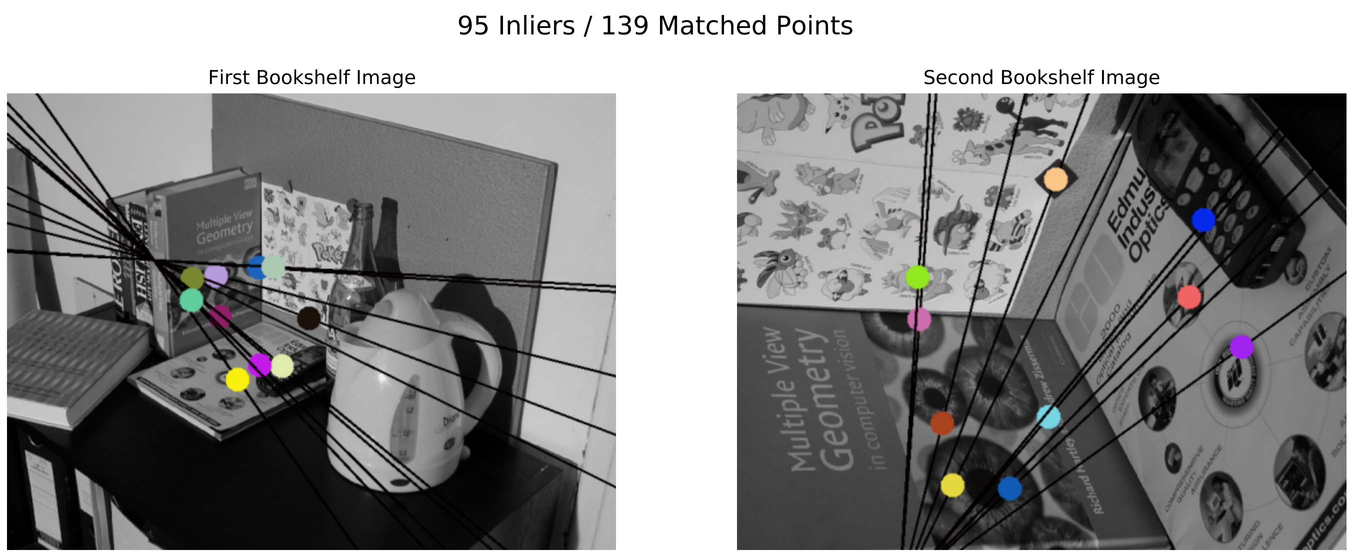

Figure 5: Performance on Bookshelf Image

An image that performed quite poorly was Graffiti, as shown in figure 6. As seen there were only 39 matched points, with only 9 of these as inliers. This demonstrates that the small number of matching points were insufficient in calculating a good fundamental matrix. On further inspection of the images, the identified feature points do not even match, with 5 features on the character in

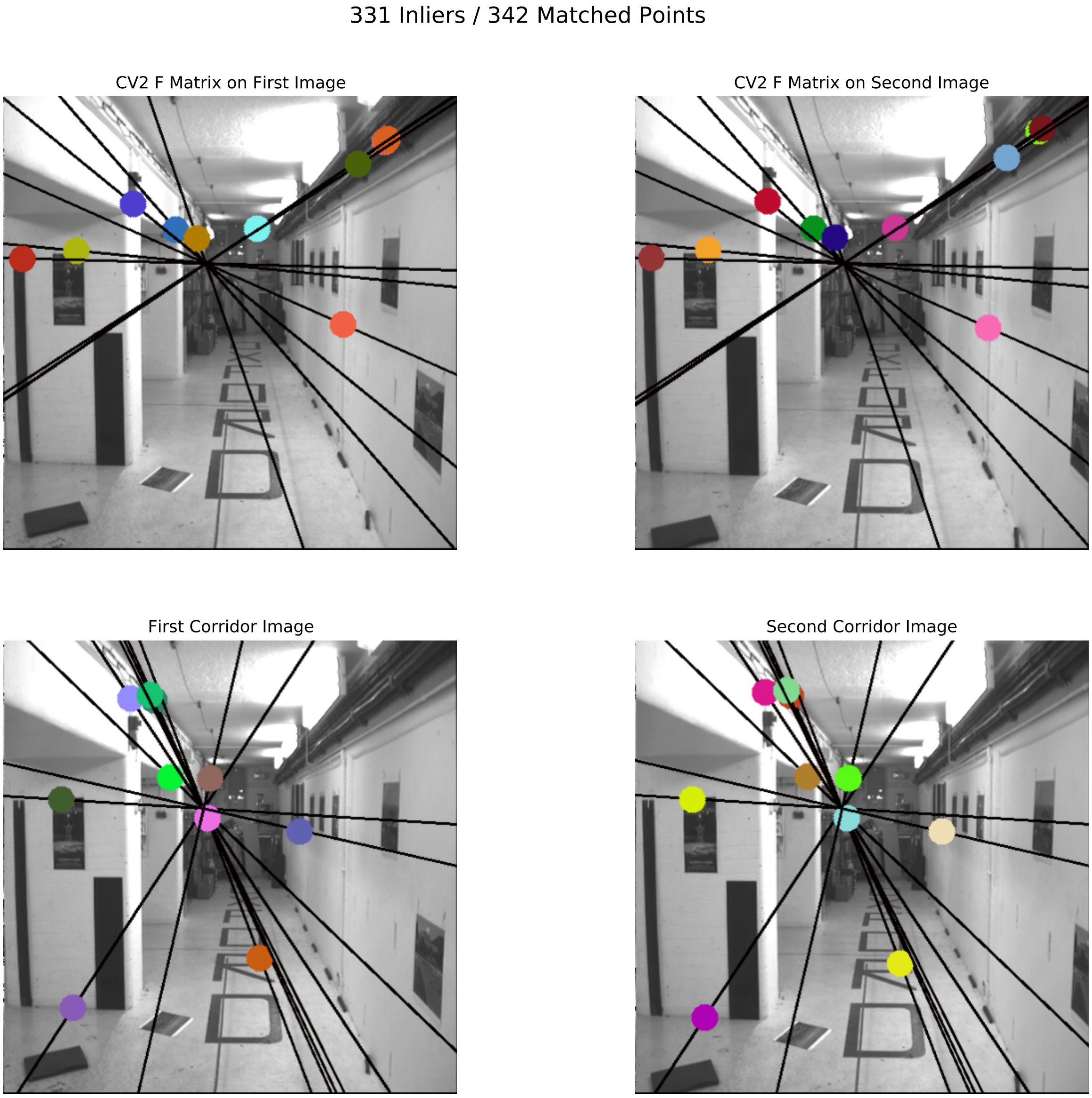

Figure 4: Comparison of OpenCV's (top) and the Calculated (bottom) Fundamental Matrix when Evaluated on Corridor Images

the right image, and only one in the left image. Potential improvements might be to adjust the Lowe's ratio to increase the number of points, and attempt a more rigorous matching algorithm.

The Kyoto image shown in figure 7, is very large with over 1000 matched points. Of these, 677 were inliers when the calculated F Matrix was computed. When compared to the OpenCV implementation, it produces very similar lines suggesting that the algorithm works appropriately. The transformation appears to mainly be a horizontal translation with some rotation.

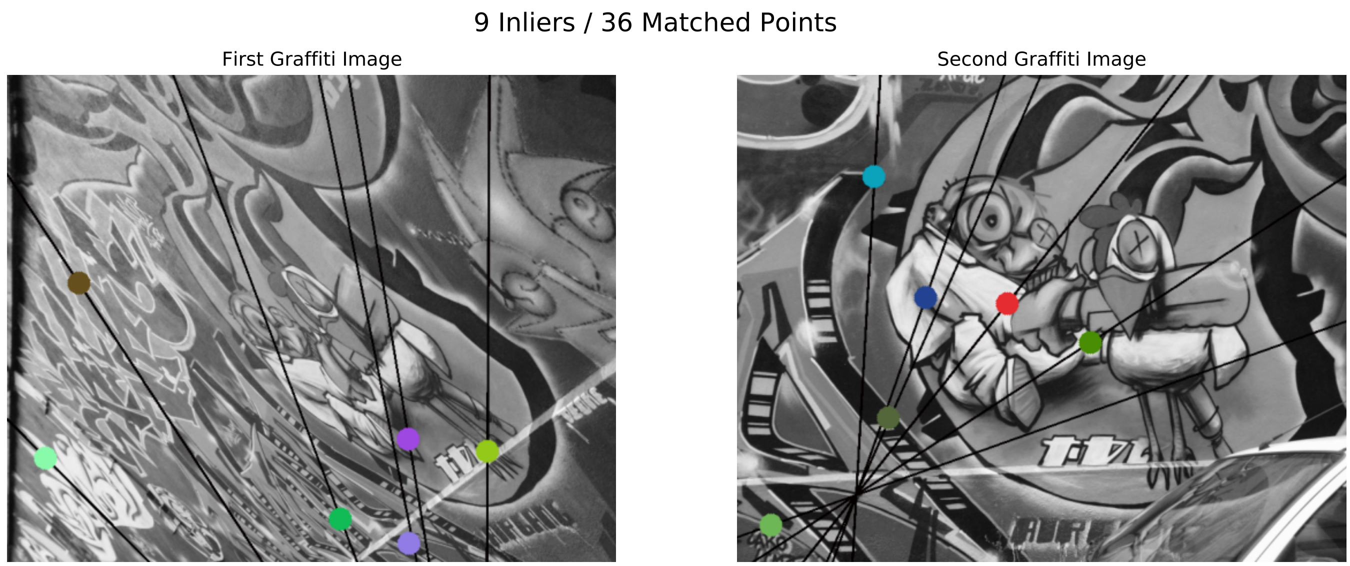

Figure 6: Performance on Graffiti Image

Figure 7: Comparison of OpenCV's (top) and the Calculated (bottom) Fundamental Matrix when Evaluated on Kyoto Images

5 References

[1] W. Chojnacki and M. J. Brooks, "Revisiting Hartley's normalized eight-point algorithm," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1172-1177, Sept. 2003, doi: 10.1109/TPAMI.2003.1227992.