COMP90086 Computer Vision, 2022 Semester 2
Assignment 3: Fundamental Matrix Calculation

1.
Keypoints for images A and images B were detected using Scale invariant feature transform (SIFT). SIFT features are found by calculating the gradient at each pixel in a window around a detected keypoint. Furthermore, a FLANN based matcher as a part of Open CV was used to identify good feature matches between image A and image B. This library performed a K-nearest neighbour algorithm to find the nearest neighbours and uses a threshold ratio (Lowe's Ratio) to find good matches by comparing the ratio between the closest and second closest match.

2.
The fundamental matrix is useful for finding correspondences between images when camera parameters are unknown (uncalibrated cameras). Having the fundamental matrix along with a point in an image enforces a constraint epipolar line which enables us to find a corresponding point in the other image.

Shifting and transforming of correspondences (normalisation) is essential for ensuring stability while estimating the Fundamental matrix. The images may have their origin in the top left or in the center, the translation step brings the centroid of the set of images to the origin. The pixels were further scaled to a common factor to ensure all of them have equal magnitude. The average distance from the origin and left images was reduced to $\sqrt{2}$, this was picked based on research from the normalised 8 point algorithm (Hartley, n.d.).
The normalisation was imposed by using a transformation matrix which later is used to renormalise the scaled F matrix. The transformation matrix follows the following structure:

$$T = \begin{bmatrix} s & 0 & -st_x \\ 0 & s & -st_y \\ 0 & 0 & 1 \end{bmatrix}$$

Where:

- s is the scaling factor
  - $s = \frac{n\sqrt{2}}{\sum(t_x^2 + t_y^2)}$
- t_x is the centroid of the x coordinates
- t_y is the centroid of the y coordinates

After normalisation an equation of homogenous linear system with 9 unknows were set up to satisfy the following equation:

$$\begin{bmatrix} x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = 0$$

A design matric function was used to form a pair of x' , x for each correspondence as the same F must apply to all correspondences.
We set this equation in form of Af = 0 where the null space of matrix is the solution of f, we recover this value using singular value decomposition and the set the smallest singular value to 0.

After recovering a F, this matrix along with matches from left images are used to recover epilines on the left image using their dot product. The distance between the line and corresponding points from the right image were used to classify if a point is an inlier or outlier. A 2-pixel error was used as the threshold distance for classifying a point as an inlier or outlier.

Finally, these steps were repeated in a RANSAC loop to find the best possible fundamental matrix with the greatest number of inliers. RANSAC is an iterative algorithm that 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 (Dai, n.d.)
The iterations required for running this loop was found using the following equation:

$$k = \frac{\log(1-p)}{\log(1-w^n)} = \frac{\log(1-0.991)}{\log(1-0.6^8)} = 1203.53 \approx 1204$$

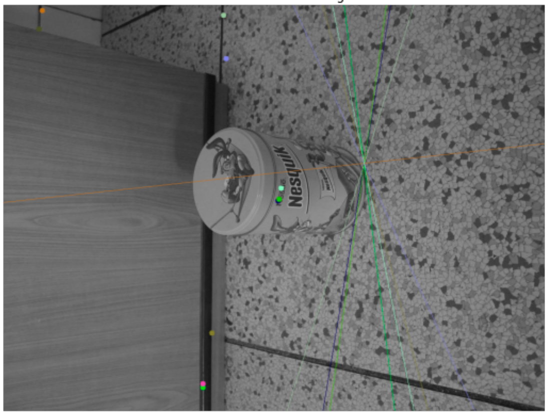
Where p is the probability of finding n- 8 inliers successfully and w is the proportion of inliers of the corresponding points returned by SIFT.

As it is difficult to estimate the proportion of inliers in our data estimating the perfect number of iterations required for the RANSAC loop can be error prone or more than required. An adaptive RANSAC may be more appropriate in our case where there are multiple images to find the best Fundamental matrix.

3.
Images with more inliers are the image have a better chance of having the key points match up correctly. Images with small translations and rotations work very well compared to images with more extreme translations or rotations. For example, from the figures below from table 1 we observe images with small parallel or horizontal movements or orthogonal movements work well. But images with drastic rotations or translations do not work well.

Images with good epiline estimation

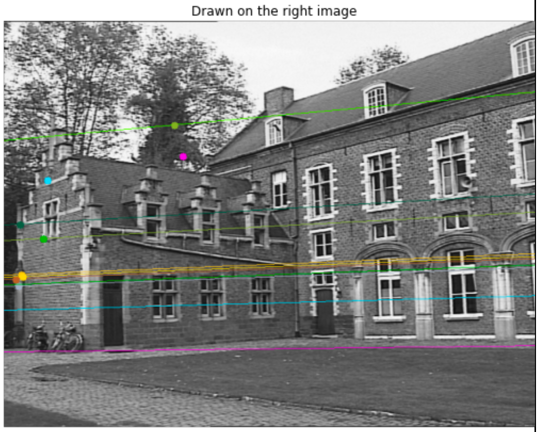
Drawn on the left image



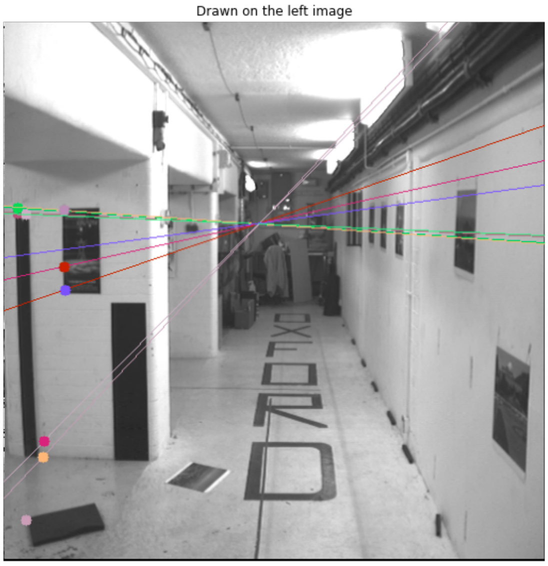
Drawn on the right image

Drawn on the left image

Drawn on the right image



Drawn on the left image



Drawn on the right image

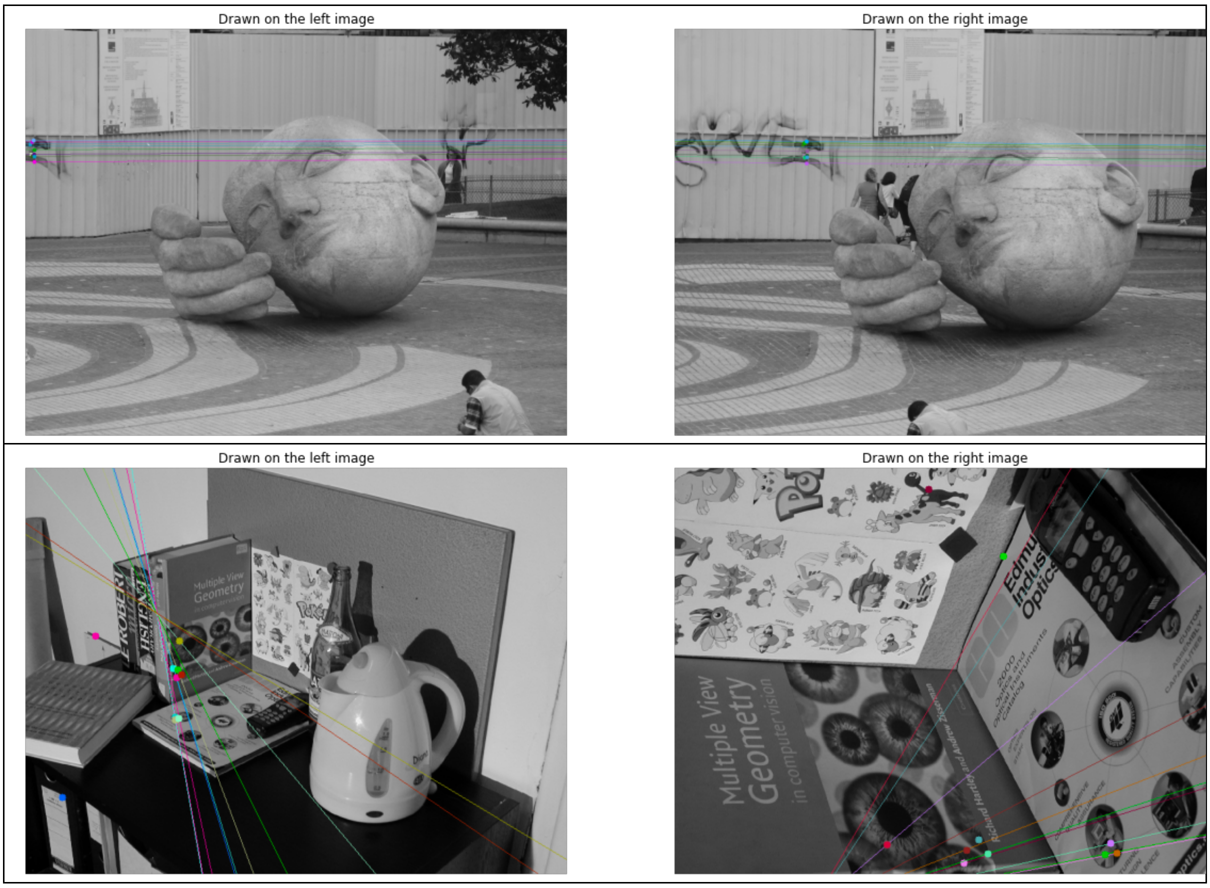


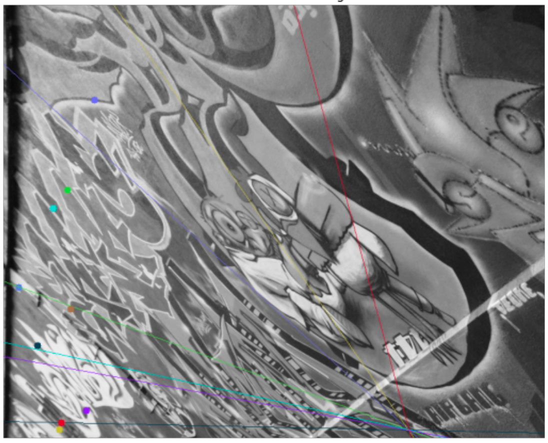
Table 1 Images with good inliers and epipolar estimation

Images with poor inlier and epiliers estimation

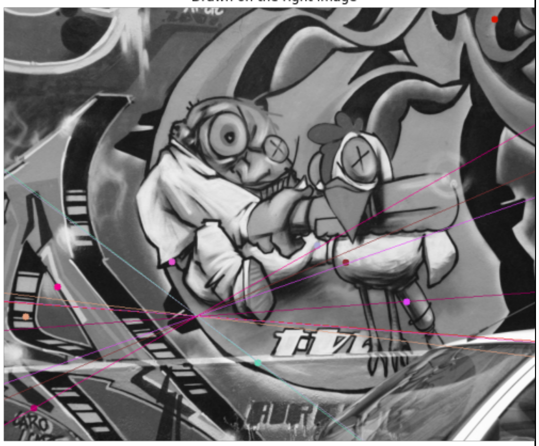
Drawn on the left image

Drawn on the right image

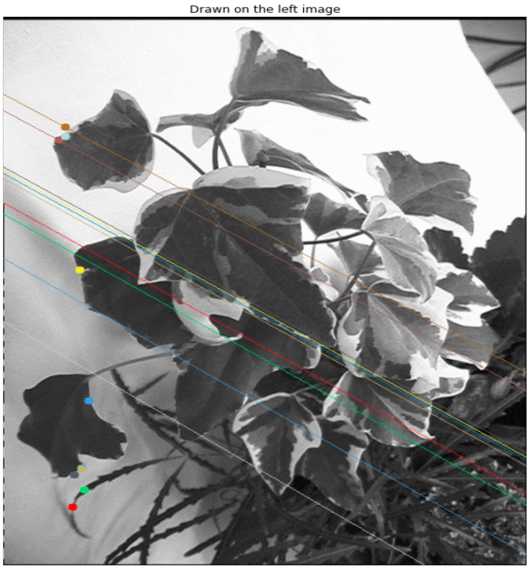
Drawn on the left image



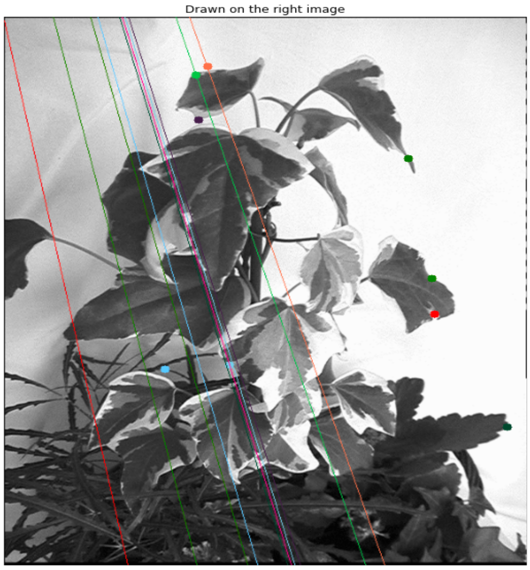
Drawn on the right image



Drawn on the left image



Drawn on the right image



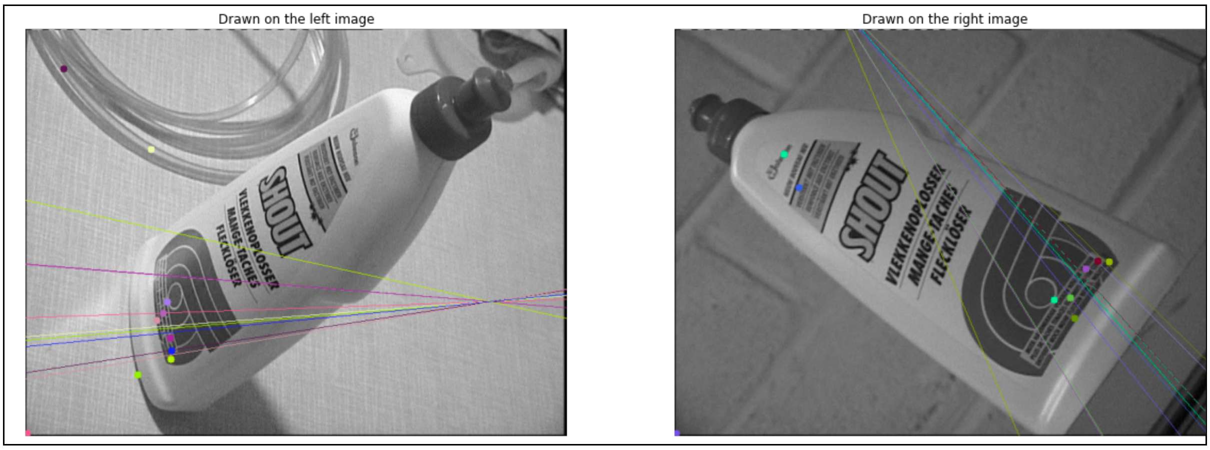


Table 2 Images with poor epiline and inliers construction