Imperial College London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Computer Vision

Author: Jia Ji
(CID: 01734749)

Date: November 20, 2022

1 Question1: Detect Salient Feature

I will choose SIFT to detect salient features. Edges(e.g.hand rail), corners(e.g.the corner of a window), blobs(e.g.the ball), and ridgese.g.touch line of the pool) can all be good features. SIFT includes scale space extreme detection, keypoint localization, orientation optimization, and keypoint descriptor. Finally, SIFT descriptors are characterized by location, scale, and orientation. Both frames' features have distinct scales, and SIFT specializes in automated scale selection by using DoG . Furthermore, some edges may be recognized as key points but lack sufficient contrast, and SIFT sets a D0G threshold that might remove those less valuable features. Additionally, the usage of gradient makes the description resistant to fluctuations in intensity. Aside from that, the viewpoint shifts between two frames, which is managed by employing a dominant orientation.

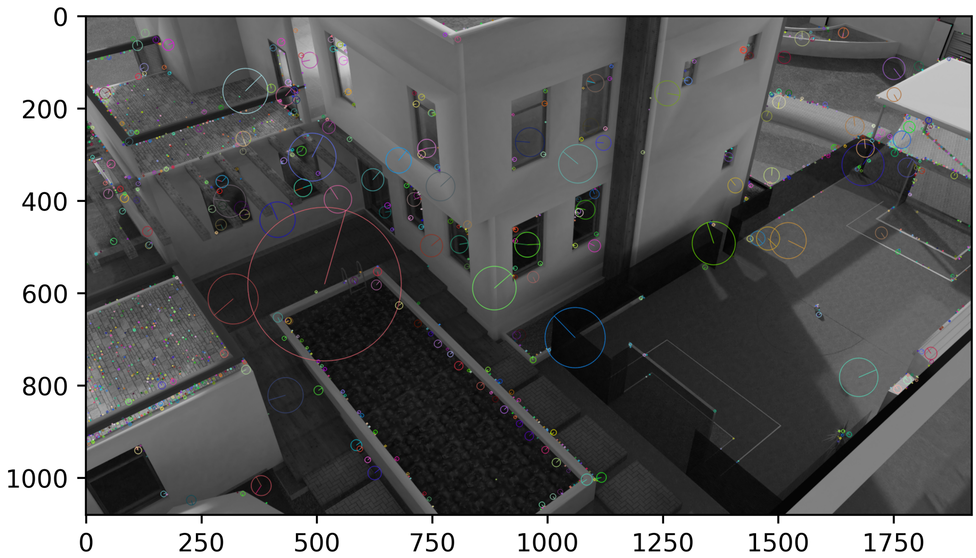
2 Question 2: Match Salient Feature

Brute-Force KNN is chosen to match salient features. Since we use SIFT to detect salient features, we use Norm_L2 here to measure distance. First, we calculate the distance between a feature vector and its nearest and second-nearest neighbours. The ratio of the two distances is then computed. If the ratio approaches one, the matching is confusing and should be refused. A ratio close to 0 indicates a favourable match. We collect matched features which pass the ratio test.

3 Question 3: Implementation

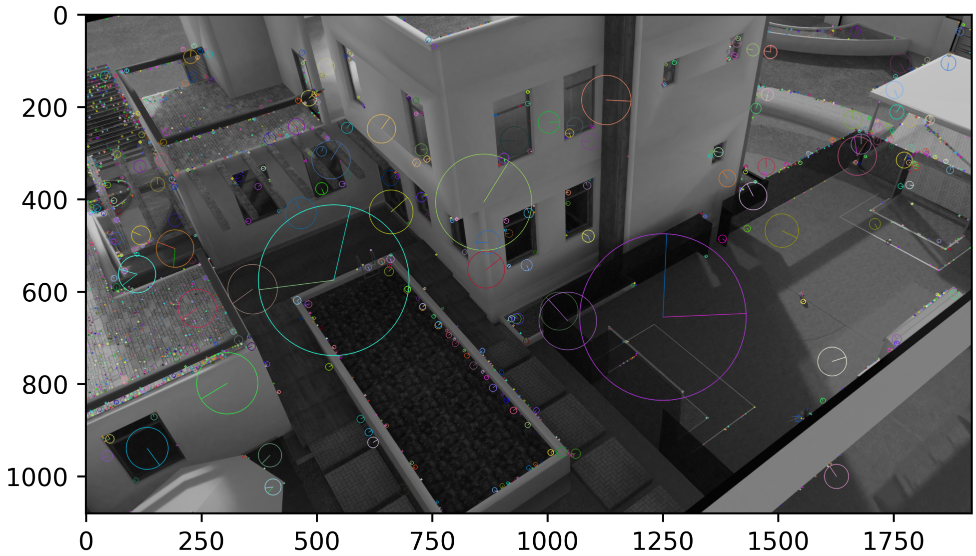
3.1 a

Frame 1 with keypoints detected by SIFT



lp

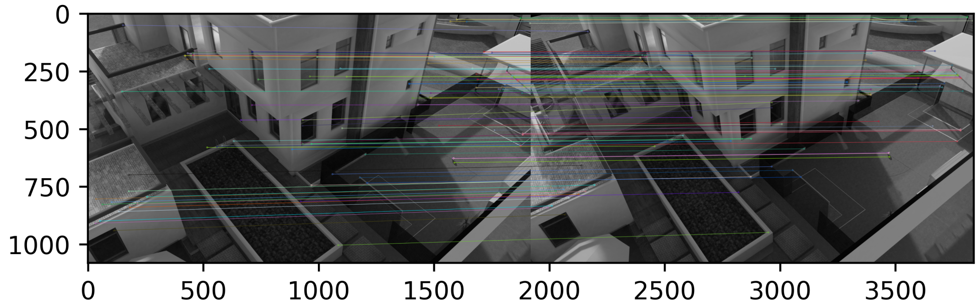
Frame 2 with keypoints detected by SIFT



The circles in the frames are salient points found by SIFT. The size of the circle shows the descriptor's scale. The orientation of the description is defined as the direction of the circle's radius.

3.2 b

Plot Matches between Frame1 and Frame2 by using BFKnn



The horizontal lines define the matches between two frames.

3.3 c

The Fundamental Matrix Estimated by Matched Features between the Two Frames

```
[[ 6.38824288e-08 -2.04308287e-06 -5.30828380e-04]
 [ 3.25967154e-06 -6.66512233e-07 -2.93759870e-02]
 [-1.76545297e-04  2.67710638e-02  1.00000000e+00]]
```

The Fundamental Matrix computed by Intrinsic and Extrinsic Matrix

```
[[-8.94716961e-04 -1.95802212e-01  1.10989238e+02]
 [ 5.60156947e-02  3.04586525e-03  2.93130697e+03]
 [-4.19999293e+01 -2.58338898e+03 -1.17061260e+05]]
```

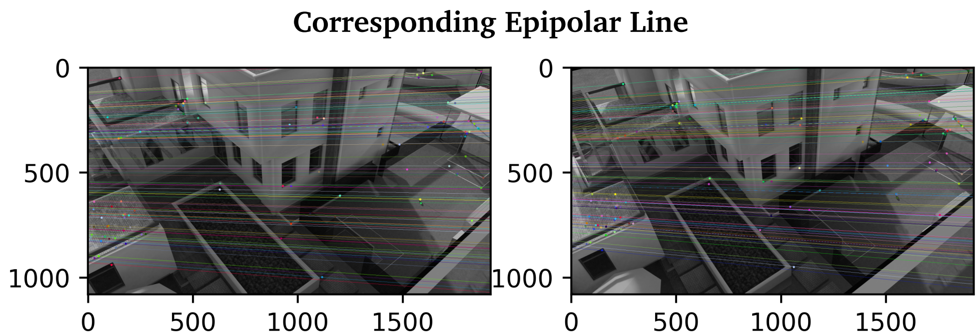
The Fundamental Matrix computed by Intrinsic and Extrinsic Matrix after Normalisation

```
[[ 7.64315164e-09  1.67264740e-06 -9.48129532e-04]
 [-4.78516075e-07 -2.60194129e-08 -2.50407946e-02]
 [ 3.58785898e-04  2.20686928e-02  1.00000000e+00]]
```

We can see that the normalised computed F and normalised estimated F are similar, but there is some small difference.

The equation $x'^T Fx = 0$ is used in the approach that estimates the fundamental matrix by matching characteristics between the two frames (estimated F). We calculate the estimation of F based on pairs of features again and again, removing outliers until the number of outliers is fewer than a certain threshold. Based on all correct matches, we improve F. The equation $F = e' \times (P'P+)$ is used in the approach that computes the fundamental matrix using intrinsic and extrinsic matrix (computed F). When I re-project feature points using computed F, I believe it is more accurate because there are less points with high errors.We only apply BFKnn when estimating F with matched features, which may result in mismatched features. We could attempt other optimization techniques, such as the symmetric test.

3.4 d



Using SIFT, we discovered several probable matches between Frame1 and Frame2. We may construct the estimated fundamental matrix using those matches. When calculating the fundamental matrix, we eliminate any outliers. Outliers are points that do not obey the uniqueness constraint, the ordering constraint, the smoothness constraint, or have a substantial inaccuracy while re-projecting. Remember that $I' = Fx$. As a consequence, we may use the fundamental matrix and the features in one of the frames to draw the matching epipolar lines of those features on the other frame. In the two graphs above, the line in the left frame corresponds to the epipolar line of features in the right frame, and vice versa. Remember that the appropriately matched points x' should be located on the epipolar line $I' = Fx$ (i.e. $x'^T Fx = 0$). As a result, correctly matched spots should have a line of the same colour run through them.

3.5 e

```
The length of swimming pool is 9.559358682303817.
The width of swimming pool is 3.073993857852149.
The area of swimming pool is 29.385409874407546.
The length of football field is 14.990683602513355.
```

We may compute the world coordinates in homogeneous form using the projection matrix of the two cameras (the dot product of K and [R|t]) and the image coordinates of the corners of the swimming pool and football field. The homogeneous coordinates are then transformed into euclidean coordinates. Finally, the euclidean distance between two points may be calculated.