CW for CV

Yupei Li, CID:02265650

1

I will choose SIFT because it can effectively find the key point and match the points between two figures, and it is invariant to rotation, scaling, and brightness changes, and also stable to view angle changes, affine transformations, and noise. And it has no requirements on the number of feature points. I would choose corner points as primary key points since they can effectively reduce the amount of information data while retaining the important features of the image, thus improving the calculation speed, and processing the reliable matching of images in real time. It keeps the features of the whole image because we could use corner to resemble them to detect objects and edges and get the descriptor of them as additional features to utilize. But I could add more other key points such as edge points, bright spots in dark areas and dark spots in bright areas using SIFT.

2

I choose SIFT to match the points in figures, and I follow 3 steps:

- Find the key point I build the Guassians pyramid and input required images to get the DOG pyramid by calculating difference between two adjacent Gaussian scale spaces. Then I compare each points with 26 points around to find the extreme point, which is key point.
- Decide the descriptor I find points around the key points (in one circle with radius equals 3 sigma). Then I calculate the gradient direction of all points to build histogram graph. The biggest value in the graph is the direction, which compose the descriptor.
- Match points For each key points in the first graph, I calculate the (Euclidean) distance between the descriptor and the descriptor of all the key points in the second graph. Then I use KNN to find the most similar two points, which are matched points.

3

3.1

Figure 1: Key points in image1

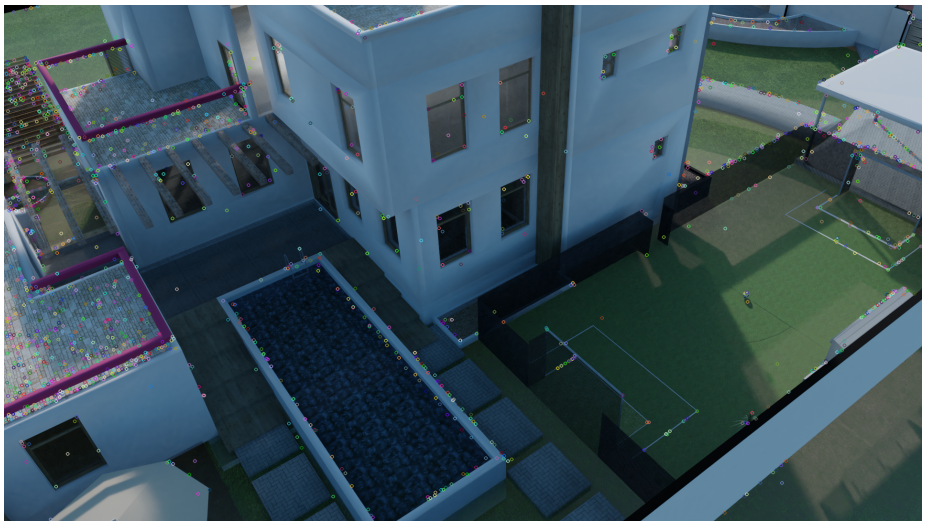

Figure 2: Key points in image2

3.2

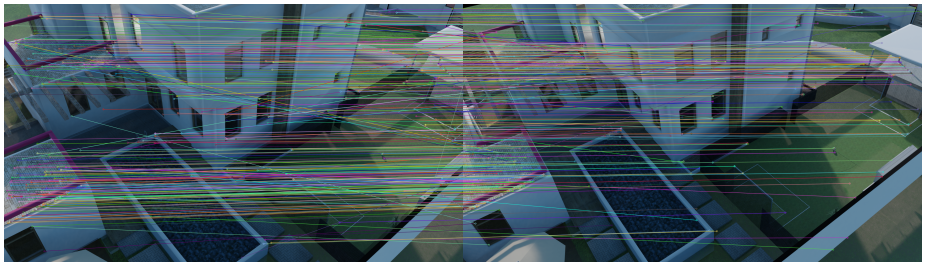

Figure 3: Matched points

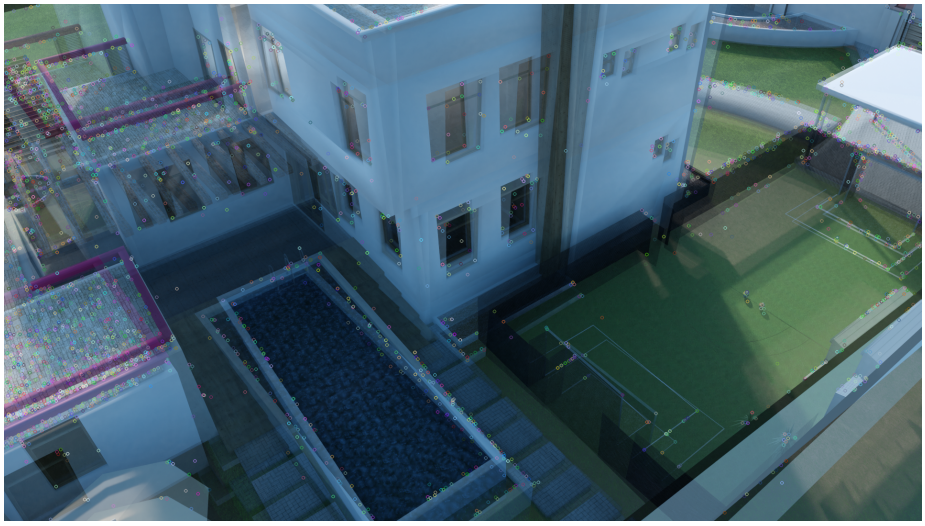

Figure 4: Composite images

3.3

Fundamental matrix use 8 points:

$$\begin{pmatrix} -2.81245986e-08 & 3.12032784e-06 & -1.58813139e-03 \\ -2.68837233e-06 & 2.75187229e-06 & -9.89971718e-03 \\ 1.22348090e-03 & 6.04739666e-03 & 1.00000000e+00 \end{pmatrix} \quad (1)$$

Fundamental matrix use parameter in camera:

$$\begin{pmatrix} 7.64315161e-09 & -4.78516067e-07 & 3.58785894e-04 \\ 1.67264739e-06 & -2.60194129e-08 & 2.20686928e-02 \\ -9.48129529e-04 & -2.50407946e-02 & 1.00000000e+00 \end{pmatrix} \quad (2)$$

Two matrix are different. The disagreements may lie in the fact we sample points in practice for the first method, which can select wrong matches (SIFT caused) while we use parameters to directly get the matrix in theory for the second. Also it may because the first method select points ramdonly in good matches set and cannot get the optimal value of the fundamental matrix.

The second method behaves well theoretically since it gives accurate parameters. But for the first one, more practical, it relies on the data to a large extent. There may be error on value of pixel, noise etc. One method to improve is we can do normalization on the input (selected 8 points) to eliminate the possible noise. And we can also improve the way they match the key points and refine how we terminate the 8 points procedure such as set smaller threshold.

3.4

For a threshold 1e-2, I got 1 pair of points:[1094, 238, 1], [1059, 240, 1]. And for a threshold 1e-3. I got 34 pairs of points that fit for the epipolar constraint. To identify these points, I loop over the good matches set to find points $x_1 \in$image$_1$ and $x_2 \in$image$_2$ that fits for the equation:

$$x_2^T F x_1 = 0 \tag{3}$$

where F represents the fundamental matrix. But due to some noise, we set the threshold. If the left side is smaller than threshold, we deem this point fits for the epipolar constraint.

3.5

Following the steps:

- Detect the key point (corner of the swimming pool and football field)
- Do rectification for two images
- Find required disparity (map)
- Do 3D reconstruction to get X, Y, Z coordinate in world coordinate system to calculate the distance

Finally we get the area with swimming pool 27.32 (meter square) (all approximately, including the rest data) with length 9.21 (meter) and width 2.96 (meter). And the football length 15.54 (meter)

4

I choose numDisparities = 256 and blockSize = 5

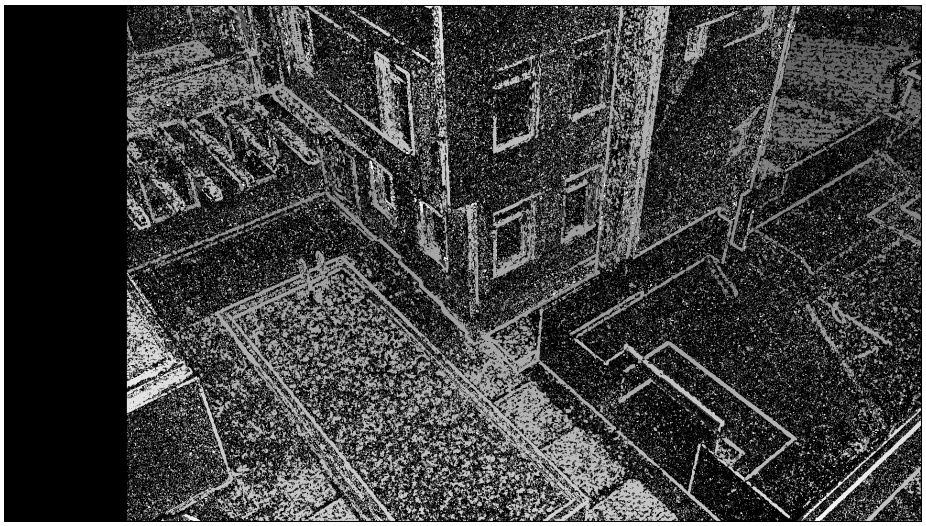

Figure 5: disparity map

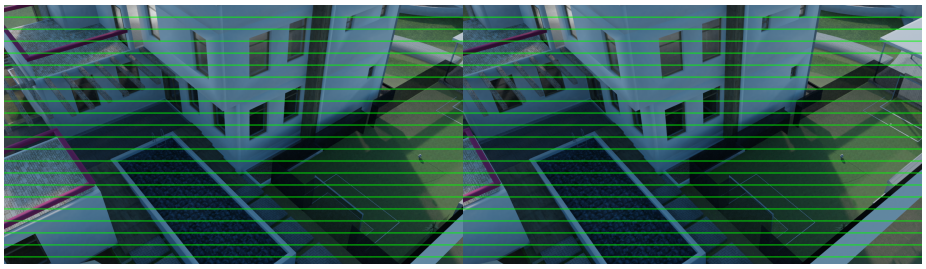

Figure 6: rectification