CS 6476: Computer Vision
Fall 2019 PS4
HUILI HUANG

Short answer questions [25 points]

1. (a): If we take any position that are local maxima in scale-space, we would obtain more repetitive and less distinctive interesting points. Because one point could be the local maxima across various scales.
   (b) If we take any position whose filter response exceed threshold, the result would depends on the value of threshold. If the threshold is high, local maxima may not pass and there will be no interest point. If the threshold is too low, additional interest points would appear. Thus, using threshold will be more distinctive while less repetitive.

2. When we use RANSAC to compute the epipolar lines, we compute x2.T * F *X1=0. However, since the uncalibrated view, not all epipolar line meet the same point. The inlier line will be on the minimize of the Epipolar lines constraint.

3. (a): Textureless surfaces: If one of the frame's surface is texture less, it is hard to create a correspondence of scanline.
   (b):occlusion and repetition: If there is occlusion and repetition, there may be few differences between two frames. It will be hard to decide which window to choose.

4. The dimension of a SIFT keypoint descriptor is 128. Each dimension represents one of the 8 gradient direction in a 4 by 4 spatial grid.

5. The dimensionality of Hough parameter space would be 4. For each descriptor, there are: position information (coordinate x and y), scale and its orientation of model.

2 Programming problem [75 points]

1. Raw descriptor matching:

Figure 1 Choose Region

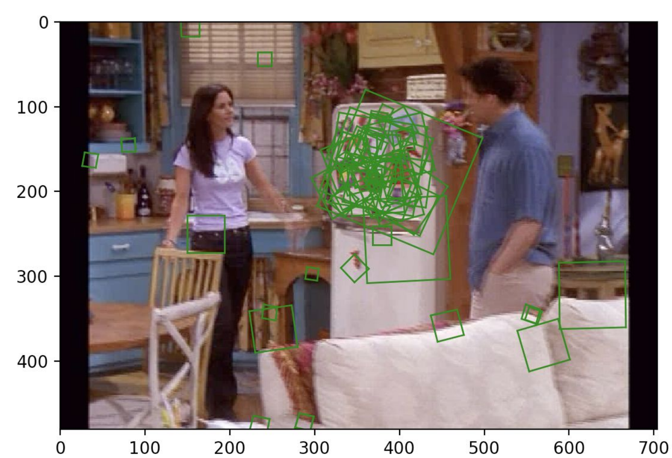

Figure 2 Matching patechs

2. Visualizing the vocabulary:

In this part, we train 40 images' descriptor by kmeans (k=1500). After that, we use getPatchFromSIFTParameters() to find 25 patches per word. As you can see, Figure 3 is tie and Figure 4 is light. Note that there are some outliers, this can be due to the number of clusters is not enough or the similarity between different descriptor is not obvious.

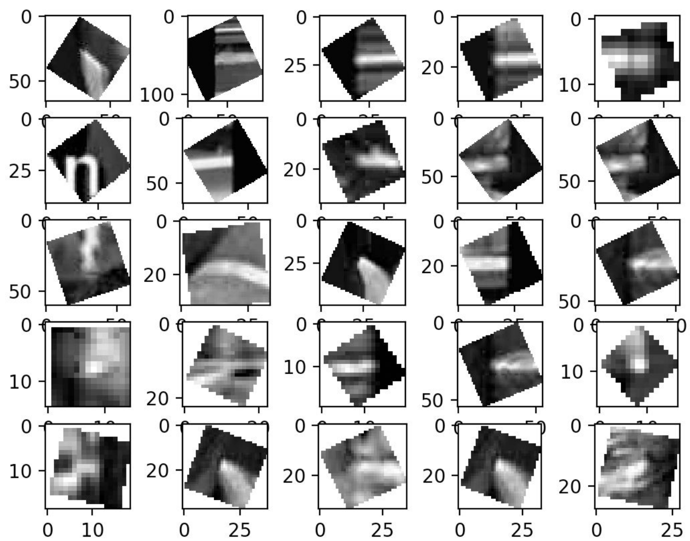

Figure 3 Words when ID=131

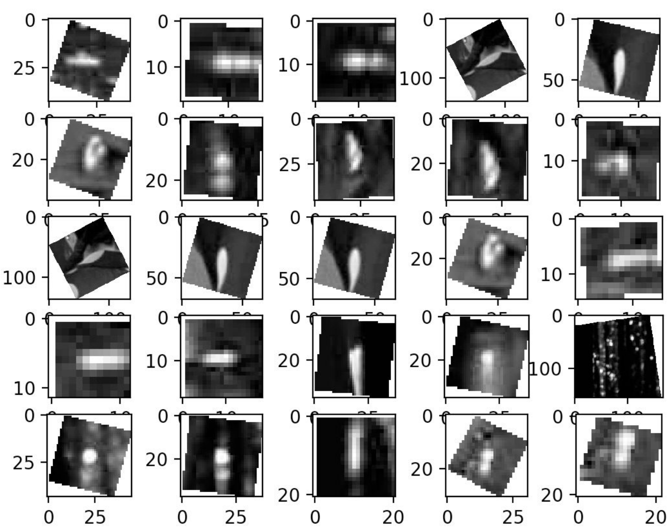

Figure 4 Words when ID=102

3. Full frame queries:

In this part,we train 50 images' descriptor by kmeans (k=1500). Then, we build a histogram.pkl file which contains all histogram of the frames and a histogram_mapping.pkl contains correspond image information. Then, we choose images whose ID is 85, 55 and 70 to execute full frame queries. The code perform well and the result is as follows:

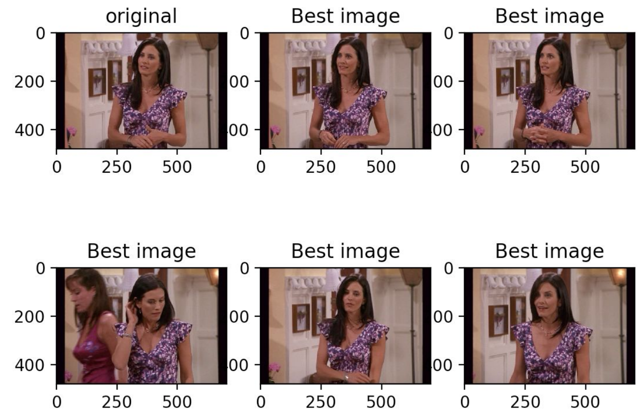

Figure 5 Frames we get when ID=85

Figure 6 Frames we get when ID=55

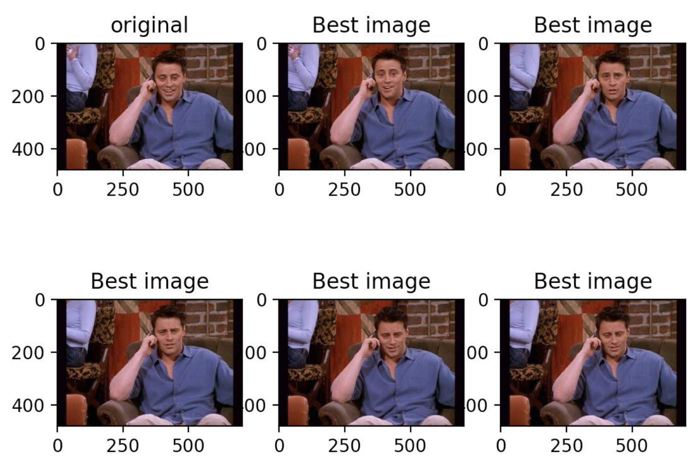

Figure 7 Frames we get when ID=70

4. **Region queries:**

In this part, the idea is as same as the full frame queries part. The only difference is that we choose the region we query instead of choosing a frame. We choose feature like dress, window and tie. As you can see, the first two feature perform well. However, SIFT cannot recognize tie. This may because :

1. we do not have enough data to train the model
2. the texture of the tie is likely to the wall (has stripe on it) so that it is hard to predict.

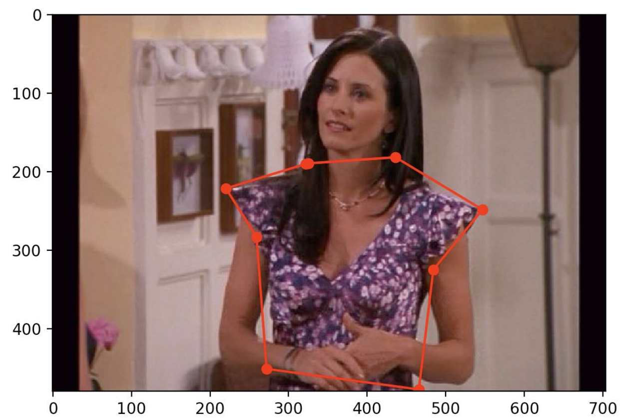

Figure 8 Region we get to search dress

As you can see in Figure 9, we find different dress in other frame.

Figure 9 Image we get for dress

Figure 10 Region we get to search window

Figure 11 Image we get for window

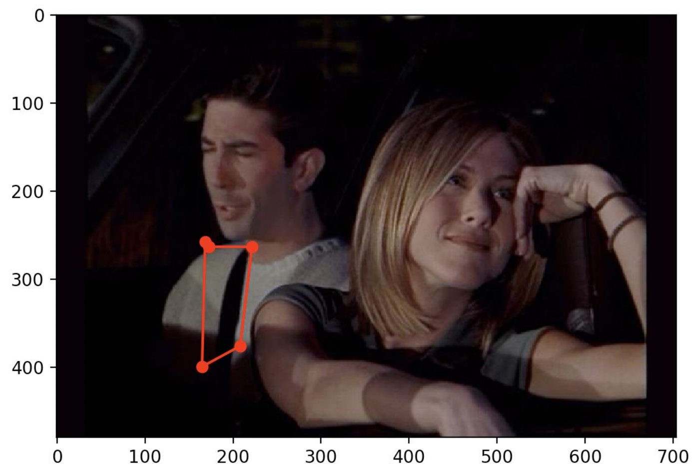

Figure 12 Region we get to search tie

Figure 13 Image we get for window