Taylor Yang

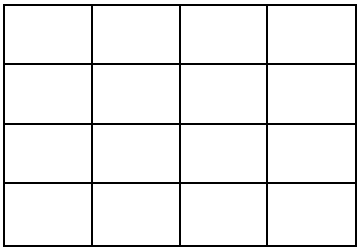
998055387

1. When performing interest point detection with the Laplacian of Gaussian, how would results differ if we were to (a) take any positions that are local maxima in scale-space, or (b) take any positions whose filter response exceeds a threshold? Specifically, what is the impact on repeatability or distinctiveness of the resulting interest points?

a) This one will do well if we have a lot of repeat figures in one image. But what it will different from b) is if we change the rotation of the figure, or what if the surrounding is very similar to the figure we choose (day and nights difference) a) will have error of finding the interest point

2. What exactly does the value recorded in a single dimension of a SIFT key point descriptor signify?

A SIFT descriptor has 128 dimensions. It was build as the square, where inside each block it has an 8 dimensional histogram. The histogram was the gradient of the orientations of that select region. So one single dimension of a SIFT is one orientation of one of the box.

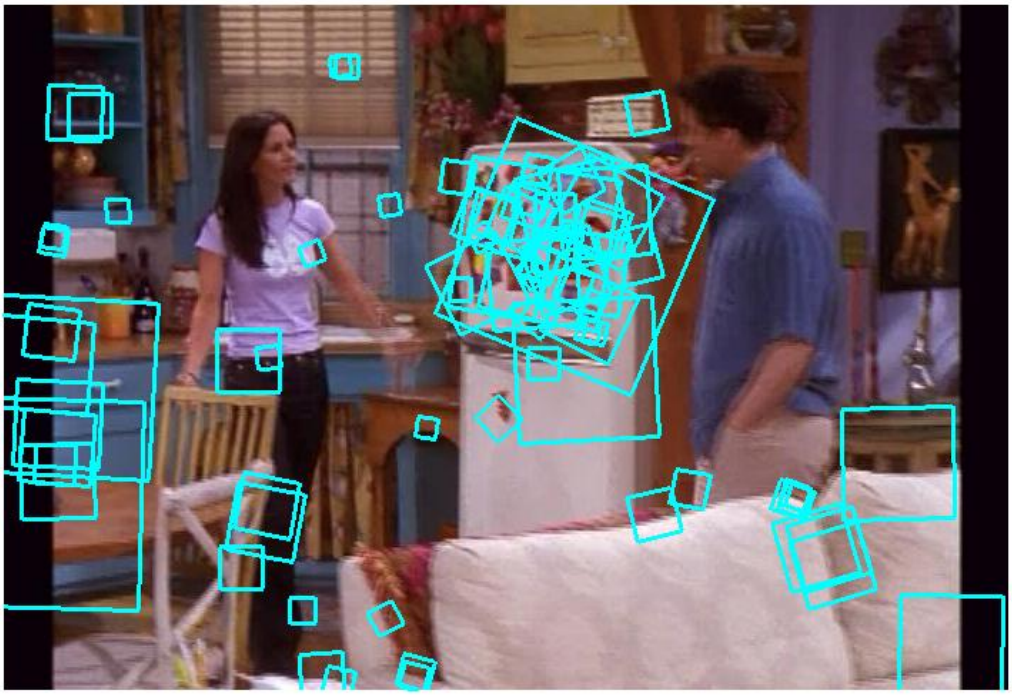


3. If using SIFT with the Generalized Hough Transform to perform recognition of an object instance, what is the dimensionality of the Hough parameter space? Explain your answer.

For each model feature, we will record 2D location, scale, and orientation of the model.

Each match will vote in a 4D Hough space

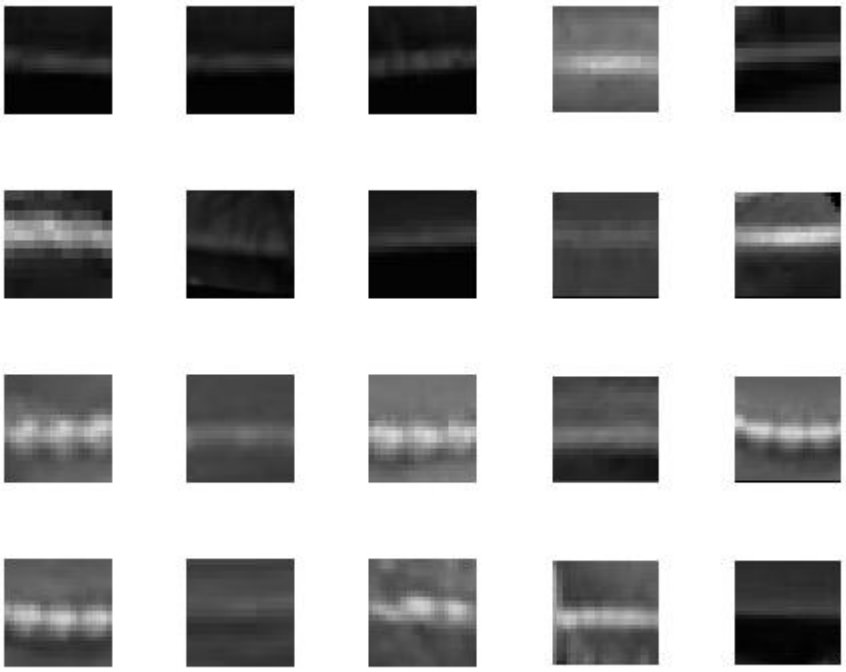
1. Raw descriptor matching
   I used threshold 0.19



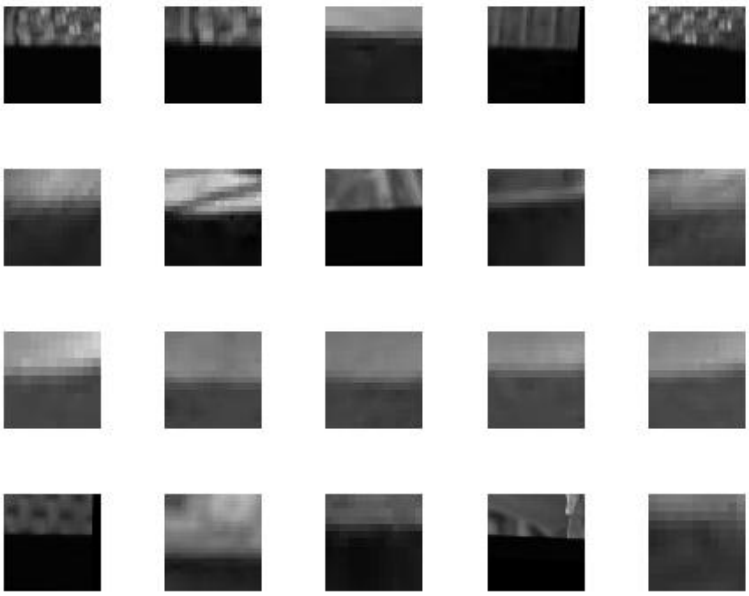
2. Visualizing the vocabulary

I used 30 figures from each frame. I used 5973 frames in total. Because there are some frames that have less than 30 descriptors so I just ignore those frames. I then used 1500 as how many visual words I will build.

Word 200, threshold 0.5. I chose 25 patches from word 200. It looks like a white line in the middle.



Word 888 threshold 0.5, it looks like some thing lighter on the top, and a dark thing on the bottom.



3. Full frame queries
   I chose frame 2 as my query frame

Here are the top 5 frames that I found similar to frame 2

reading frame 3 of 6612

reading frame 4 of 6612

reading frame 5 of 6612

reading frame 6 of 6612

The second one I choose, as my query frame is frame 37

Here are the top 5 frames that I found similar to frame 37


213

.

The third one I choose is frame 200

reading frame 199 of 6612

reading frame 202 of 661

reading frame 203 of 6612

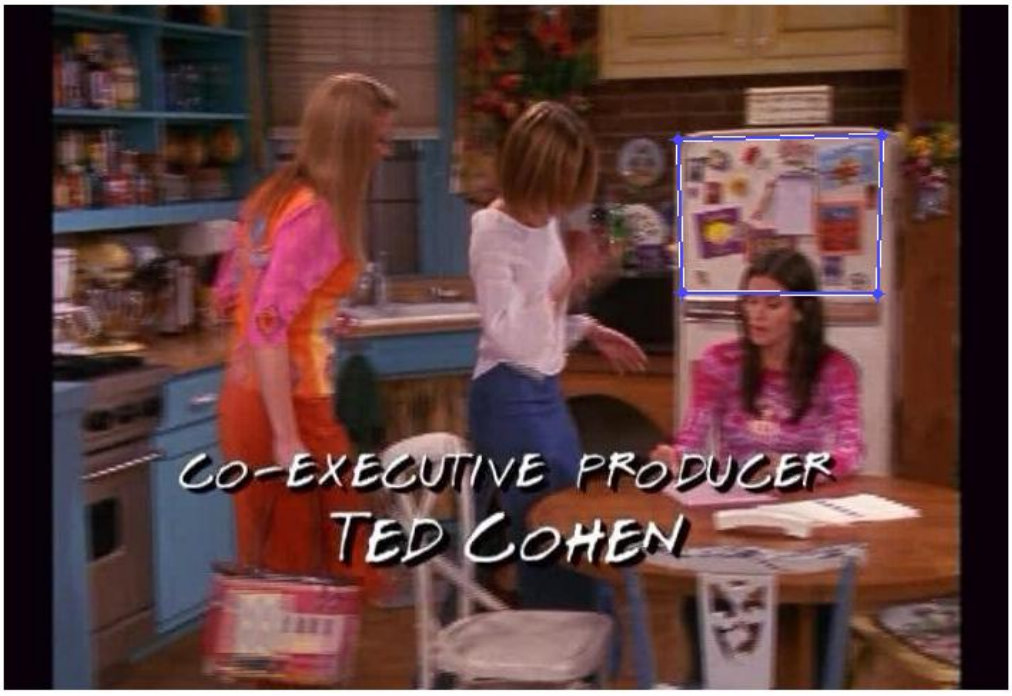
reading frame 204 of 6612

For part3, I build a histogram matrix for all the frames. Each row is one frame histogram. In order to find top 5 similar frames, I build a similarity vector, which has the similarity of the frame I chose to all the other frames. I sort my vector from descend order and ignore the NAN ones, then I chose top 2-6 because top 1 will be the same image.

4. Region queries
   I use Frame 25 as my resource frame. I chose the man's shirt as my region

I then chose Frame 22, the blue lamp.

For Frame 2209, I chose the stickers on the fridge. This one we can see that even the people and scene changed we can still find it. However, it also happens some failure part where we can see it misunderstand the picture frames from the second and last image as its similar ones.



Frame 863, We can see this one correctly find the woman’s shirt.

part 4

For part 4 I follow the idea of applying weights for my bag of words based on the frequency they appear. Also I build a stoplist that will put the word in if that word appears 4000 times more. However, the result looks very similar for my trying. The professor said its okay because sometimes we can not always see the big difference of our selected image. But