# 1 - Short Answer Questions

## Q1

When performing interest point detection with the Laplacian of Gaussian, how would the results differ if we were to

(a) take any positions that are local maxima in scale-space, or
Ans. When we take only points which are local maxima in scale space, it results in less number of interest points with more repeatability and distinctiveness.

(b) take any positions whose filter response exceeds a threshold?
This approach results in more interest points subject but are subjected to chosen threshold value. Its used in harris corner detector and can have good balance between repeatability and distinctiveness based on chosen threshold value.

## Q2

What is an “inlier” when using RANSAC to solve for the epipolar lines for stereo with uncalibrated views, and how do we compute those inliers?
Ans. Inliers are points which satisfy the below equation using fundamental matrix:

$$x^T F x' = 0 \quad \text{with} \quad F = K^{-T} E K'^{-1}$$

Fundamental Matrix can be estimated by either 7/8 point algorithm.

## Q3

Name and briefly explain two possible failure modes for dense stereo matching, where points are matched using local appearance and correlation search within a window.
Ans. Two possible scenarios where stereo matching with correspondence search along scanlines using normalized correlation or SSD will fail are-

(a) Textureless Search

(b) Occlusion, repetion

## Q4

What exactly does the value recorded in a single dimension of a SIFT keypoint descriptor signify?
Ans. Value recorded in single dimention of SIFT vector signify the contribution of SIFT patch to sift vector.

## Q5

If using SIFT with the Generalized Hough Transform to perform recognition of an object instance, what is the dimensionality of the Hough parameter space?
Ans. Hough paremeter space should be 4Dimentional with dimensions being position, scale, orientation and votes.
This is because in SIFT we try to capure interest points at different scales, orientations and positions and therefore accordingly we need more dimentions for voting using Hough Transform.

# 2


Figure 1:

Steps are as follows:

1. Calculate distance between feature points to nearest feature points in other image using dist2() function.

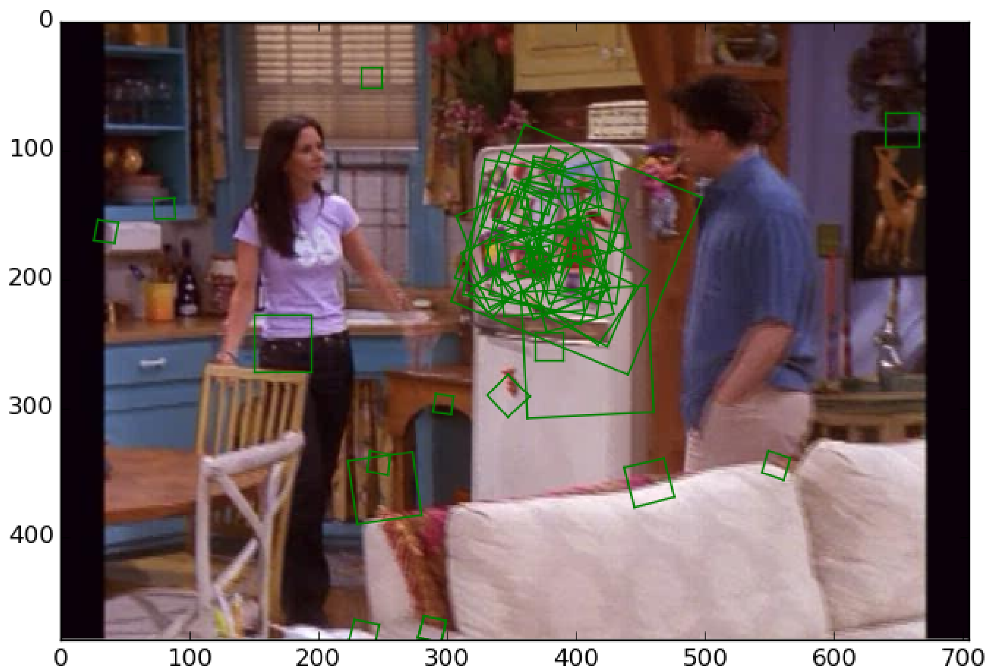



Figure 2:

2. Calculate minimum distance for each row and append features row wise to matrix matchFeatures.
3. I tried to filter best results by thresholding matching far away from mean value. I used 80% as threshold value.

## 2

Visualizing Vocabulary

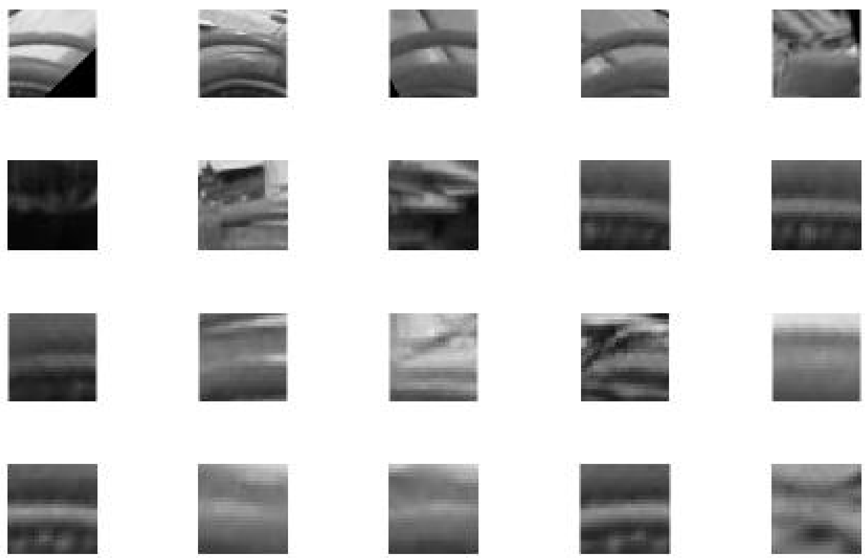
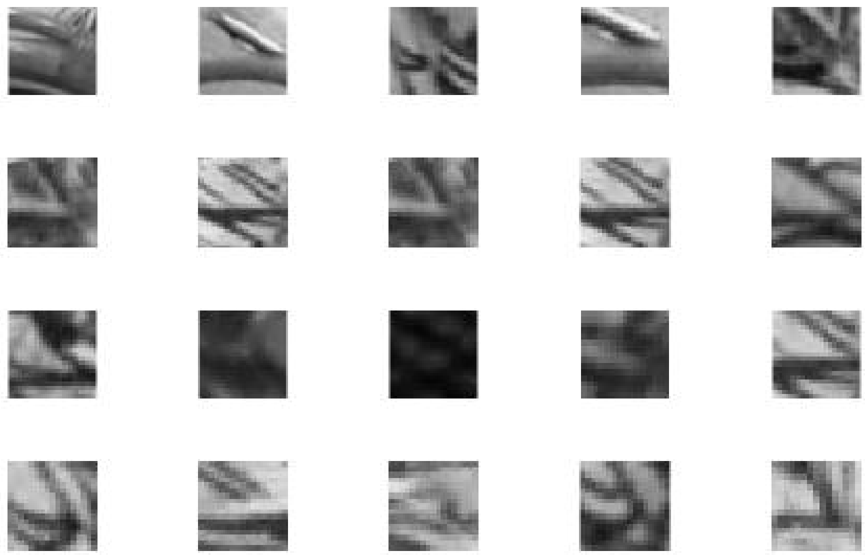


Figure 3:

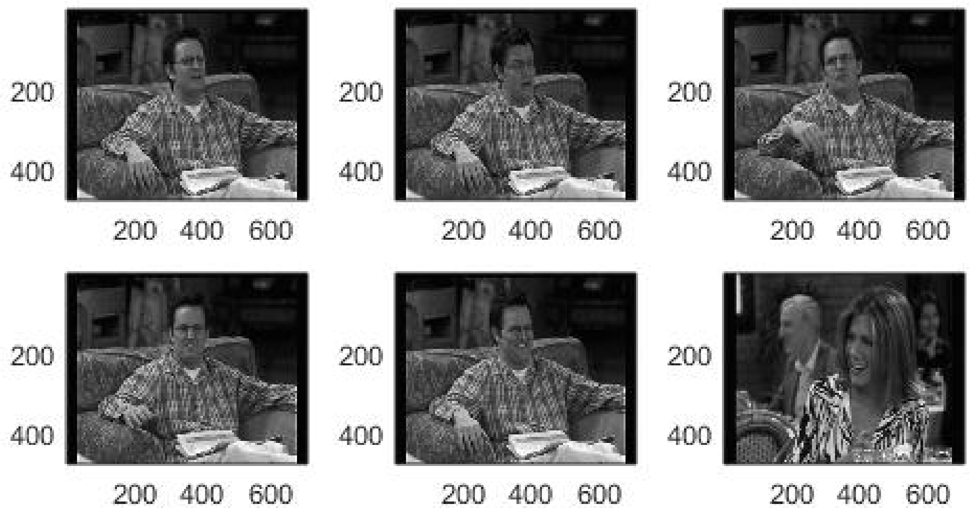


Note: I created vocabulary only for 40 images from friends_0000000060.jpeg to friends_0000000100.jpeg. Some observation about the images above

1. Patches belonging to same word are similar. This is desired as paches making up the a word shouldn't be much different.
2. I read all *.mat files and store feature descriptors in memory.
3. I used kmeans2 algorithm from scipy to cluster images with k=600. Clustering algorithm returned centroid of clusters and codebook which contains mapping of each feature vector with derived clusters.
4. I chose randomly two clusters and took 6 patch images which made chosen cluster/WORD.

# 3

Full Frame Queries

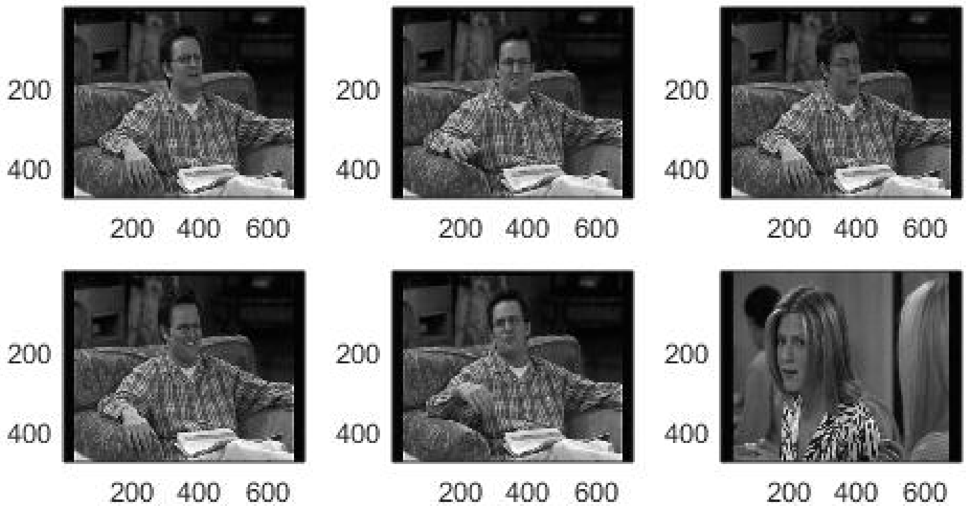

200
400
200
400
600


Here query image was matched with similar feature set of 40 images.
Here are few important points of my algorithm:

1. I used histogram to find similarity between frames.
2. I calculated similarity of histograms using I1*I2/(I1+I2).
3. Results we quite accurate using this approach.

# 4

Region Queries


200
400
200 400 600


I queried for 2 different images. The results were quite accurate for this case as well.
Approach is similar to full frame queries with only difference being that this time used region histogram to match instead of full frame histogram.

There was a bad case also observed but I think its due to smaller image set.