1 Short answer questions

1. If we use any local maxima positions, interest points at all the scales will be detected. This will increase the repeatibility of interest points across different images since same interest point can occur at two different scales in two different images. However, it will decrease the distinctiveness of the interest point. This is because now the poorly localized interest point will not be invariant to photometric and geometric differences.

   If we use a threshold to filter positions, the repeatibility will decrease as the threshold increases. A valid interest point may be filtered out because it occurs at a lower scale than threshold, thus descreasing repeatibility across images. Contrary to the first case, the distinctive increases here because by thresholding we remove all the poorly localized interest points.

2. We need to solve the equation $X_2^T E X_1 = 0$ where $e_1 = X_2^T E$ to solve for epipolar lines. To compute inliers we will compute the distance between a point and it's corresponding line. If for a correspondence, $d(x_1, e_2) < t$ and $d(x_2, e_1) < t$ where t is the threshold, it will classified as inlier.

3. The two possible failure modes for dense stereo matching are:

   - Textureless surfaces - If the scanlines are in textureless areas it is difficult to find correspondences. This is because the error function will not achieve a minima in textureless areas.
   - Repititive patterns - Similarly, in the case of repititive patterns, many points on the scanline will give similar results, hence it will be difficult to find correspondences.

4. In SIFT, 16×16 window is divided into 4×4 blocks ie 16 blocks and gradient is computed in 8 directions for each block. A single dimension of SIFT corresponds to one of the 8 gradient directions in one of the 16 blocks.

5. The dimensionality will be 4 corresponding position (x, y), scale and orientation.

2 Programming problem

Figure 1: Image 1 with marked ROI

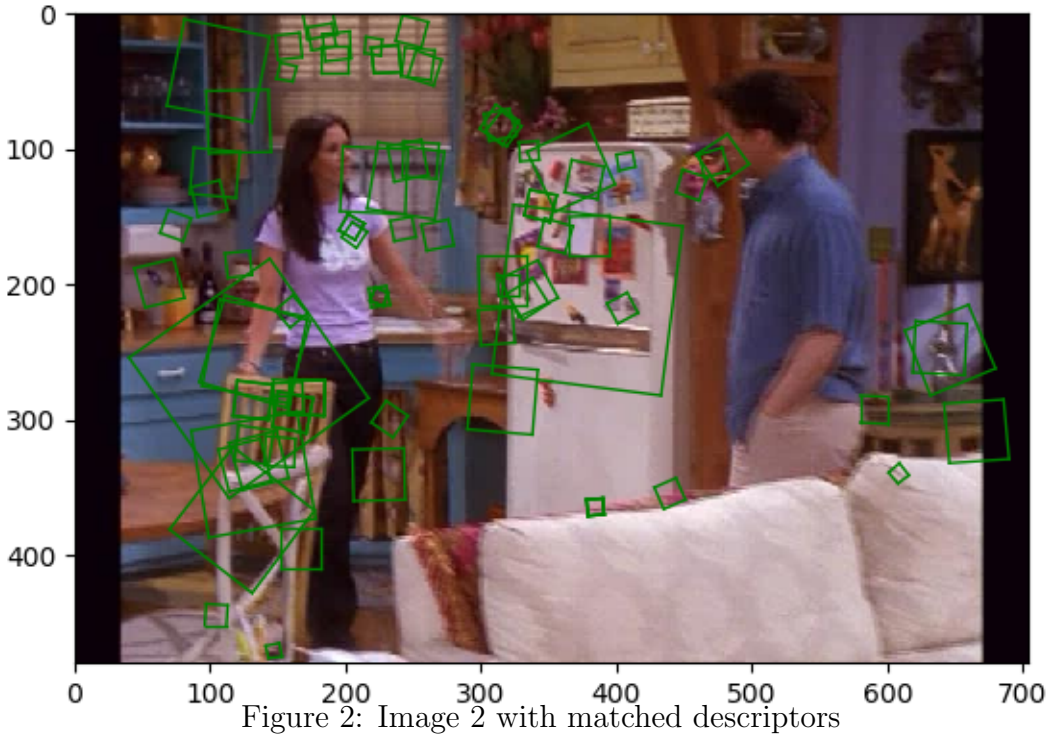

Figure 2: Image 2 with matched descriptors

1. The chair pattern is captured across the whole scene, in the chair, windows, furnishings.

2. Vocabulary construction Number of clusters/words : 500
   Number of frames used for clustering: 100
   Sampling frequency of frames: 5 ie every 5th frame is used for clustering
   Number of descriptors used for each frame: random descriptors upto 2000

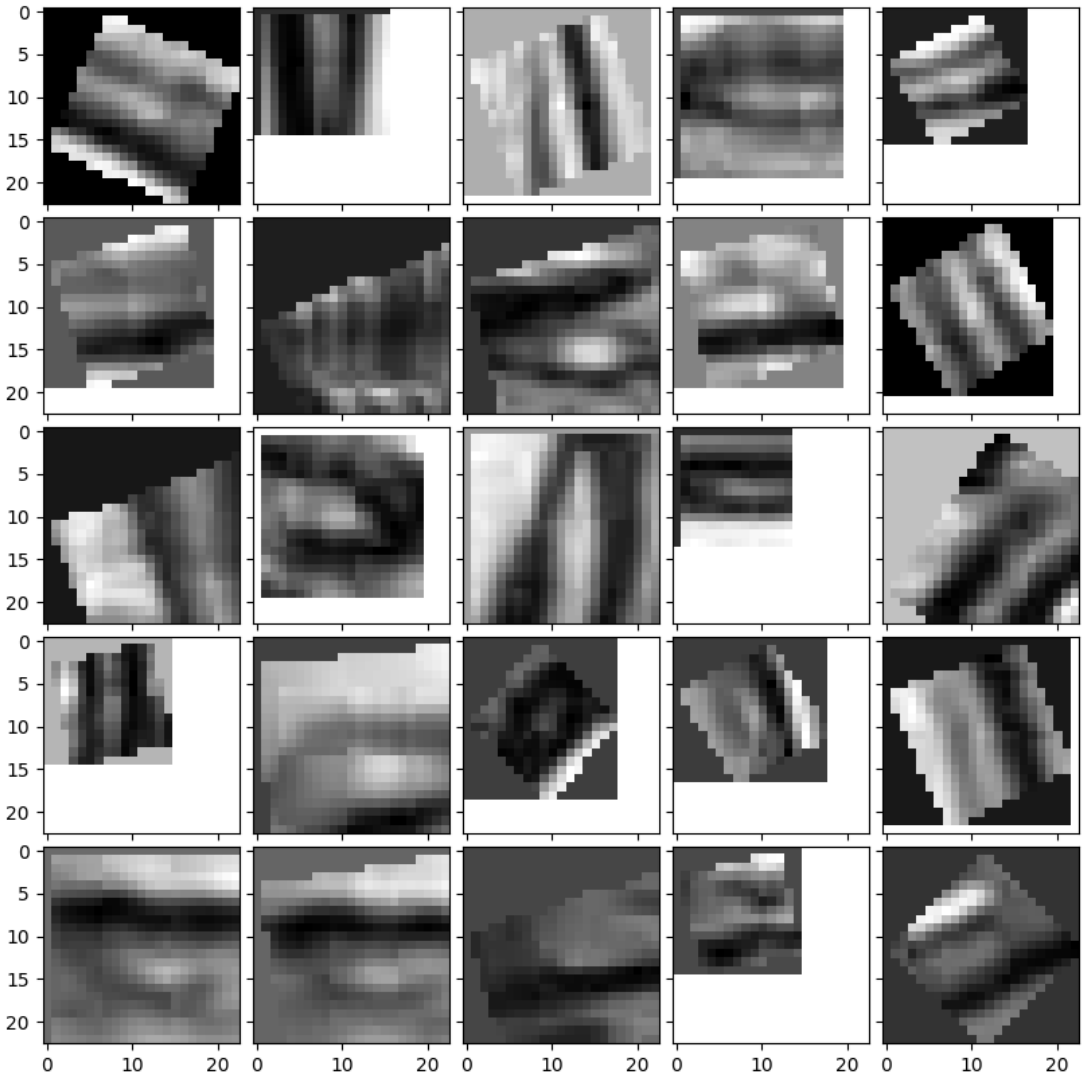

Figure 3: Patches of word 95

We can observe that a texture pattern of lines is captured in the patches in various scales and orientations.

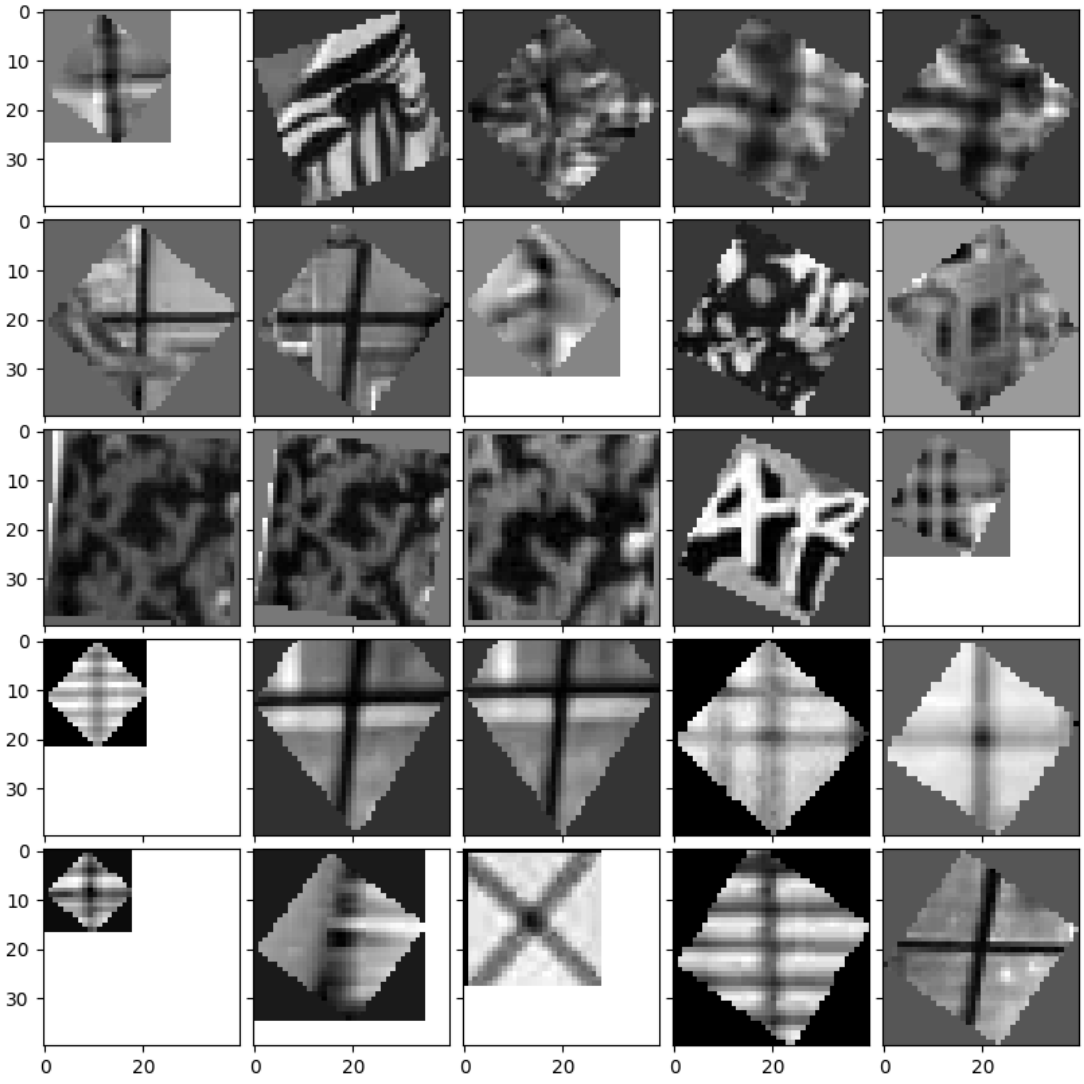

Figure 4: Patches of word 401

This word captures the cross texture pattern both detailed and zoomed across various objects like windows, opening credits.

3. In the following examples, top left image is the query image and rest are the top 5 results in the following order $[[Q, 1, 2], [3, 4, 5]]$

Figure 5: Full Query 1

This query gave very good results, as there is less clutter, sharp edges and the actor occupies most of the frame, therefore leaving little room for varying descriptors.

Figure 6: Full Query 2

1,2,5 are relevant images in the results, whereas the rest are false positives. These two false positives are also present in the third example. This can be attributed to these two images having lots of interest points compared to other images and since I have used upto 2000 interest points, it might have led to mis-classification of features in different clusters.

Figure 7: Full Query 3

This example retrieves two relevant images. In the third image, it captures the scene despite the actor not being present. This can be attributed to the shimmering background on the stage.

Figure 8: Region Query 1

4. I captured the pattern of shirt as interest region. The top 3 results retrieve the same scene while the other two tries to captures the pattern in dress.

Figure 9: Region Query 2

I selected opening credits in this query. It gives very good results capturing opening credits in different scenes.

Figure 10: Region Query 3

This is a failure case where the interest region is refrigerator frame. It gives images with occluded or no refrigerators at all, just aligning to the noise.

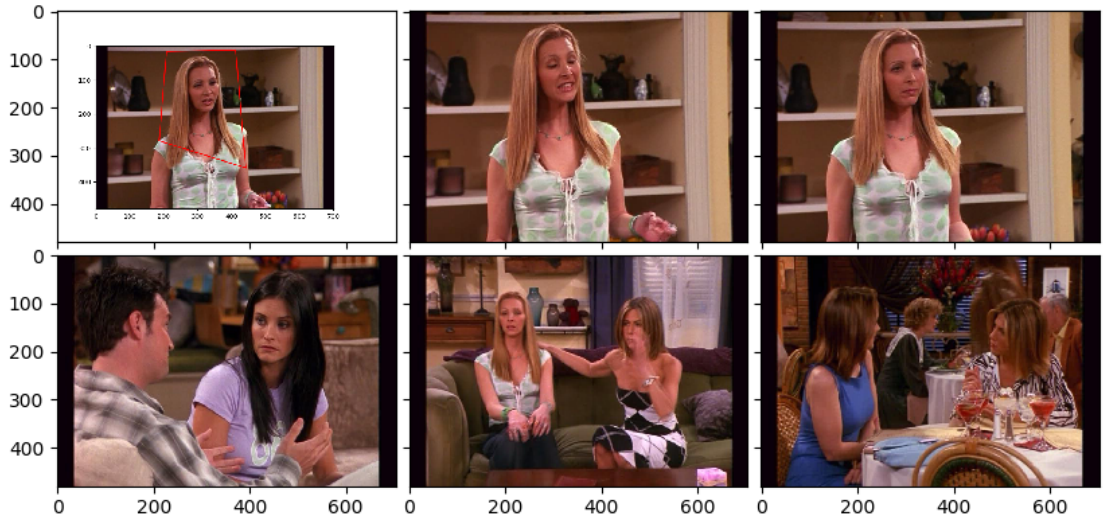

Figure 11: Region Query 4

In this query I tried to capture the character. It gives good results, but since color is not considered while matching features, it gives an actor's image having similar hair and face also as a result(3rd and 5th result).