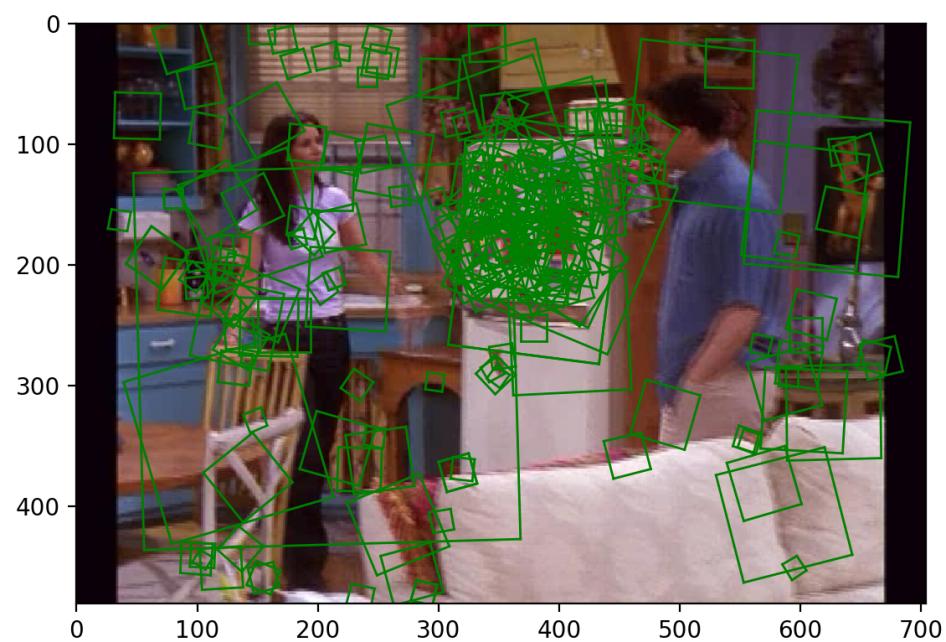


PS4

Shaurye Aggarwal

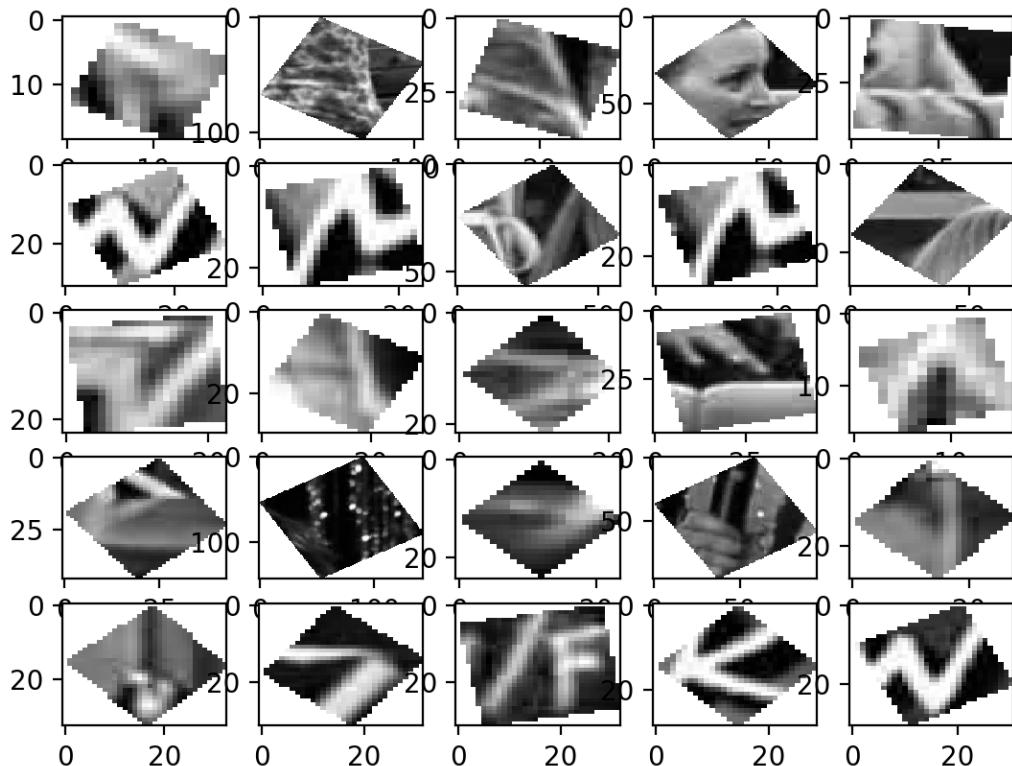
Q2.

Q2.1

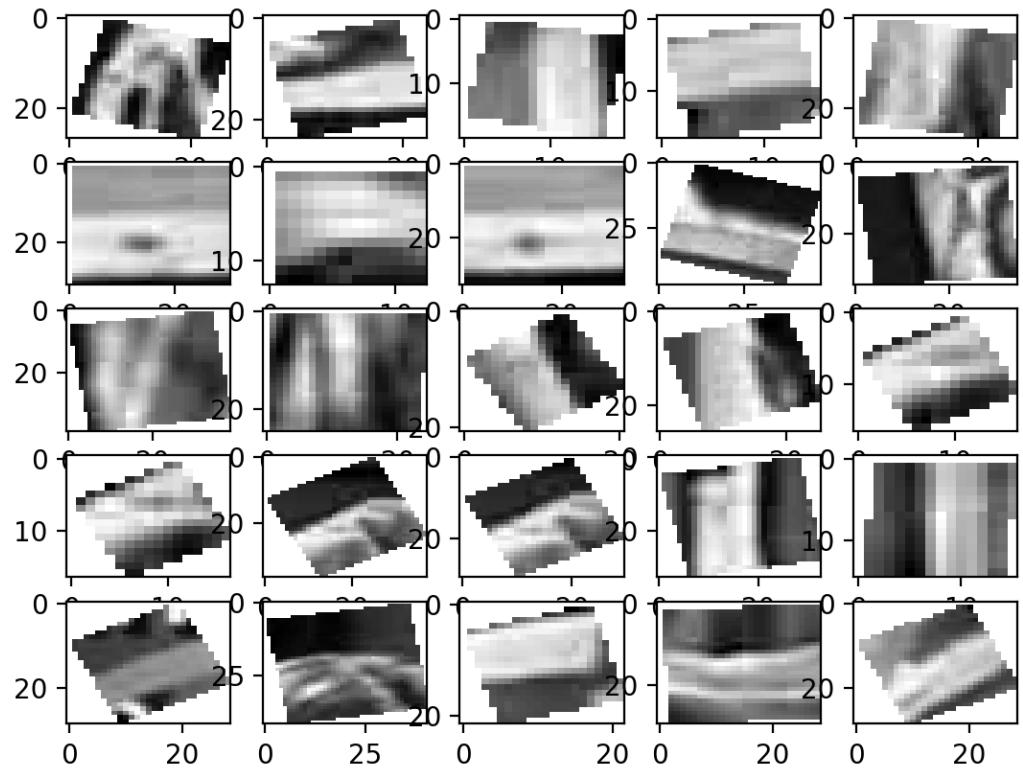


Q2.2

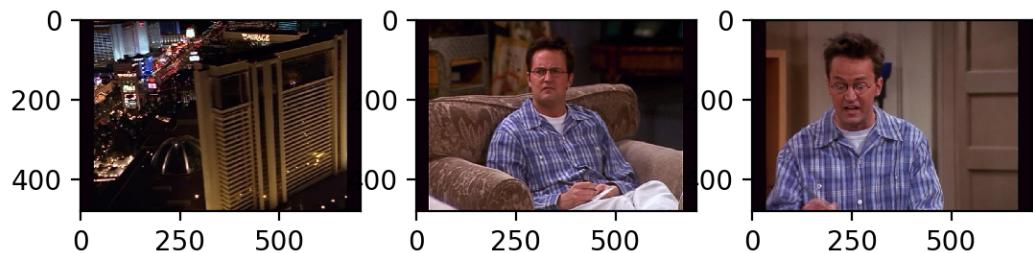
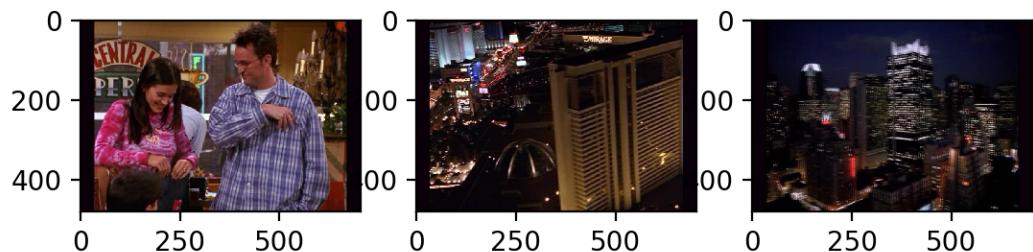
The following figures show first 25 patches of 2 random labels out of 1500 labels. Labels to different descriptors were assigned on the basis of kmeans that was trained iteratively using a training dataset containing 3000 random images from the original dataset. The first picture shows a pattern where the lighter pixels(closer to white) form a zig-zag like structure with sharp acute angles over darker surfaces(closer to black). The structures kind-of look triangular. However, some patches, for example, patch in the fourth row and 2nd column, don't exhibit similar structure. This might be due to poor visibility or due to less number of features being explored by kmeans (less number of kmeans center).

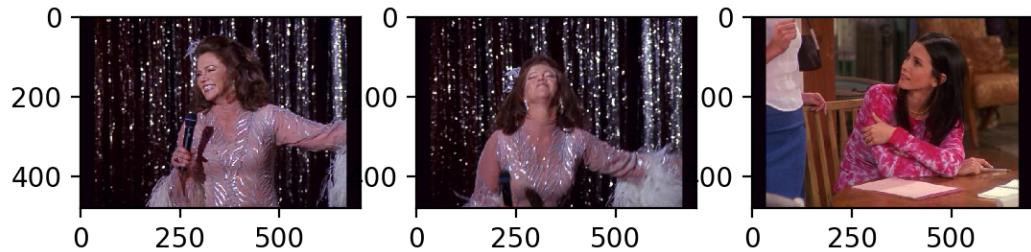
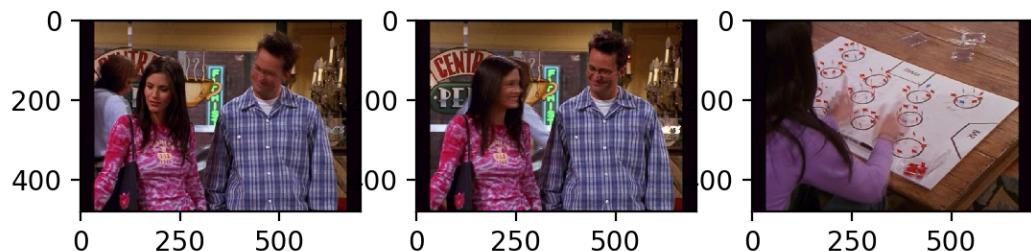
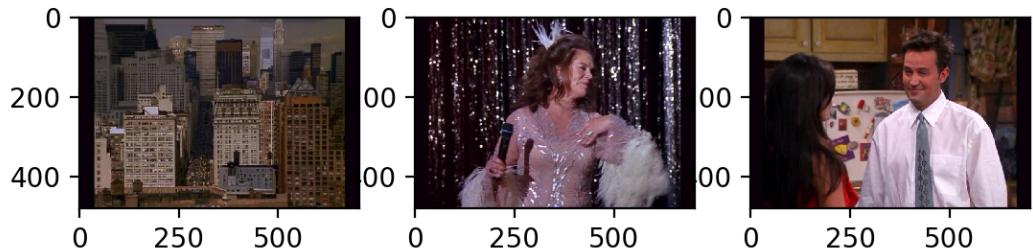
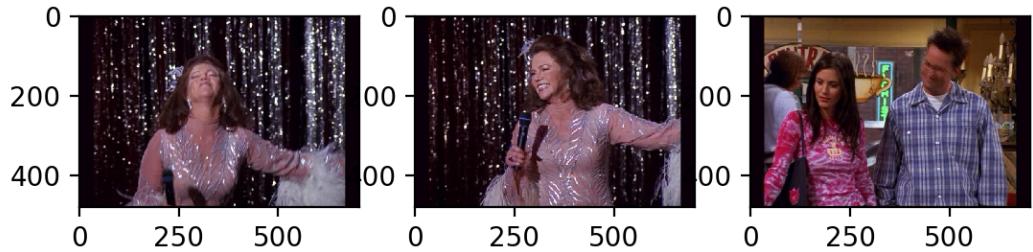


The second picture displays the sample patches for the second randomly chosen word from the vocabulary. The patches in this case are characterized by stripe of light pixels (pixels closer to white) surrounded by darker patches formed by pixels closer to black. While some of the surrounding dark patches are darker than the other as some are closer to gray while the others are closer to black. Some patches in this case are less similar than the other patches. This can again be attributed to the limited numbers of centers, not enough to generalize the different descriptors.



Q 2.3





The final image in each of these picture is the query image. These figures show that a sample of 3000 with only 1500 feature isn't quite accurate in describing a frame. This might be partially due to the clustering algorithm used as mini batch k means was used with maximum batch size of 500 and the model was trained iteratively. It can be seen while there are a few resemblances as in the first query and the last one, a lot of the retrievals get high score with query even when they correspond poorly due the poorly trained model having more bias towards such pictures.

Q1.1 If we take any positions that are local maxima in scale-space, the features we choose would be less distinctive as scale might have difference values that may produce many correspondences to many other interest points and more repeatable as they have a higher probability of having similar scale. If we take any positions whose filter response exceeds a threshold, the results of the detection would basically depend on the threshold and each image might have different threshold thus giving less repeatable features. It will be however more distinctive.

Q1.2 Once the fundamental matrix, F , is calculated using the 8-point/7-point algorithm, an inlier when using RANSAC to solve for the epipolar lines for stereo with uncalibrated views is a pair of points which minimizes $x^T * F * x$, which would be zero for perfect correspondence. A threshold can be set for $x^T * F * x$ and all points under that threshold would be considered inliers.

Q1.3 Two possible failure modes for dense stereo matching, where points are matched using local appearance and correlation search within a window are:

- a) Occlusions: Occlusions may cause a few objects to appear/disappear when viewed from different angles. Moreover, due to occlusion a correspondence point in one view might be occluded by something else in the other image.
- b) Texture-less surfaces: If texture-less surfaces are characterized by scanlines, it will be difficult to correspond the images as the correspondence depends on local appearance.

Q1.4 Each dimension in the SIFT descriptor corresponds to one of the 8 gradient directions of the 4x4 arrays

Q1.5 Four Dimensional: each descriptor votes on-

- i) X location

- ii) Y location
- iii) Scale
- iv) Orientation