Problem Set 4

Problem 1

Question 1

If we take any positions that are local maxima in scale-space, our results will be more repeatable and less distinctive. This is because using local maxima will mitigate differences in size across two different images. This process is also less distinctive because interest points won't differentiate between different objects that have similar features. Example: a sparkly dress and the NYC skyline.

If we take any positions whose filter response exceeds a threshold, we will have less repeatability and more distinctiveness. Using a threshold makes the interest point detection not recognize changes in scale, thus being more selective and increasing distinctiveness while decreasing repeatability.

Question 2

In RANSAC, inliers are points that fit the hypothesized model. With epipolar lines, we are trying to estimate the fundamental matrix E, a mapping between 2 images. The x's for which $x^T E x' < \varepsilon$, a threshold, are the inliers.

Question 3

1. Repeated patterns: multiple points give similar results, so finding correspondence is difficult
2. Textureless surfaces: likewise, textureless surface is also is a possible failure mode because the constraint for uniqueness will not be met and the error function may not find a minima, making finding correspondences hard

Question 4

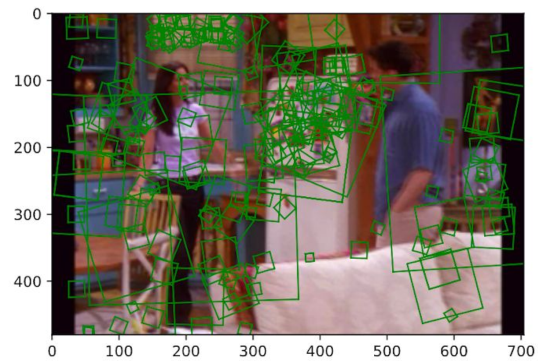
A single dimension in the SIFT descriptor represents a gradient direction with the frequency of pixels in about the same direction.

Question 5

Hough space will be 4 dimension corresponding to x, y, scale, and position of the SIFT descriptor.

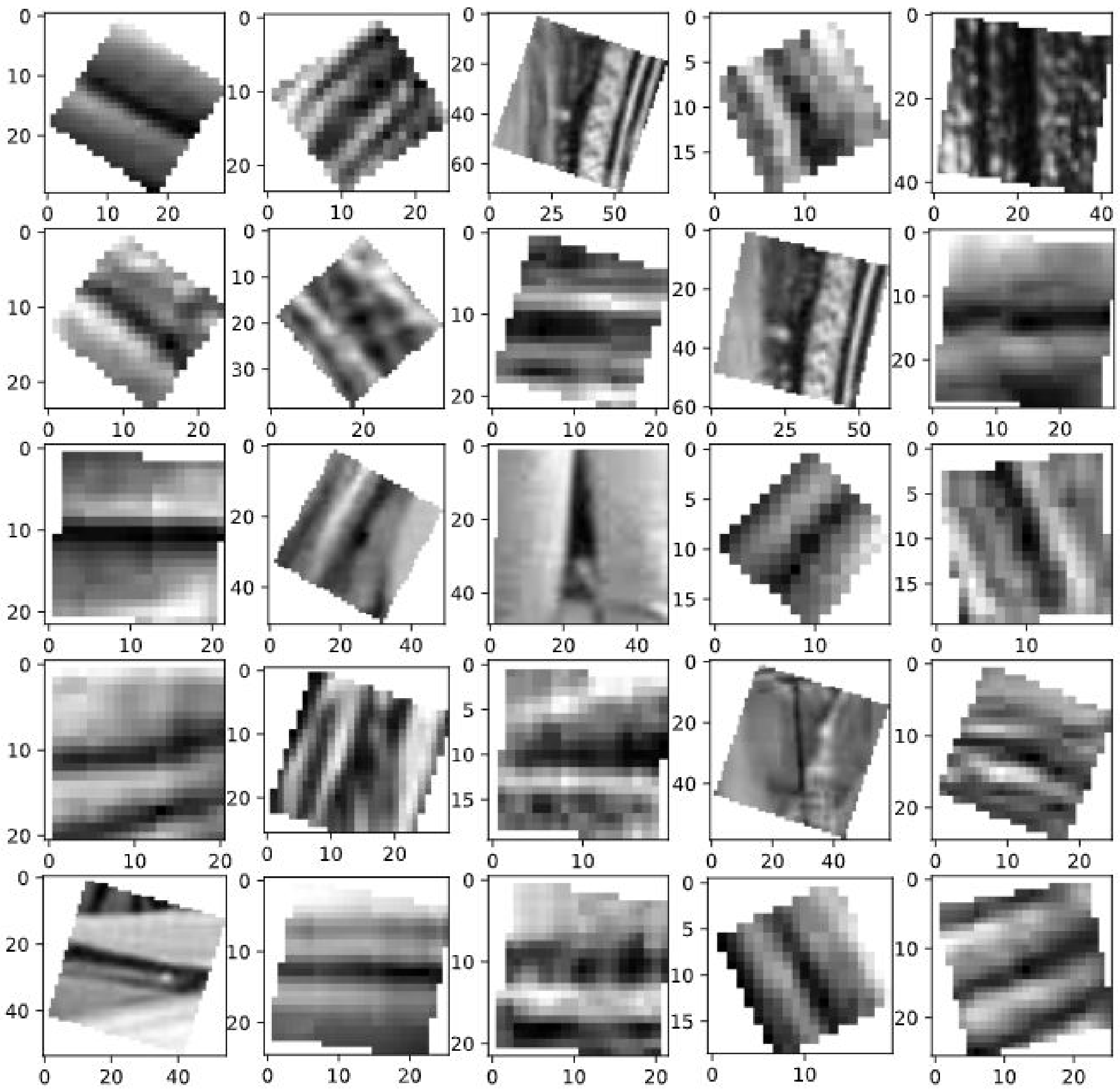
Problem 2

Question 1

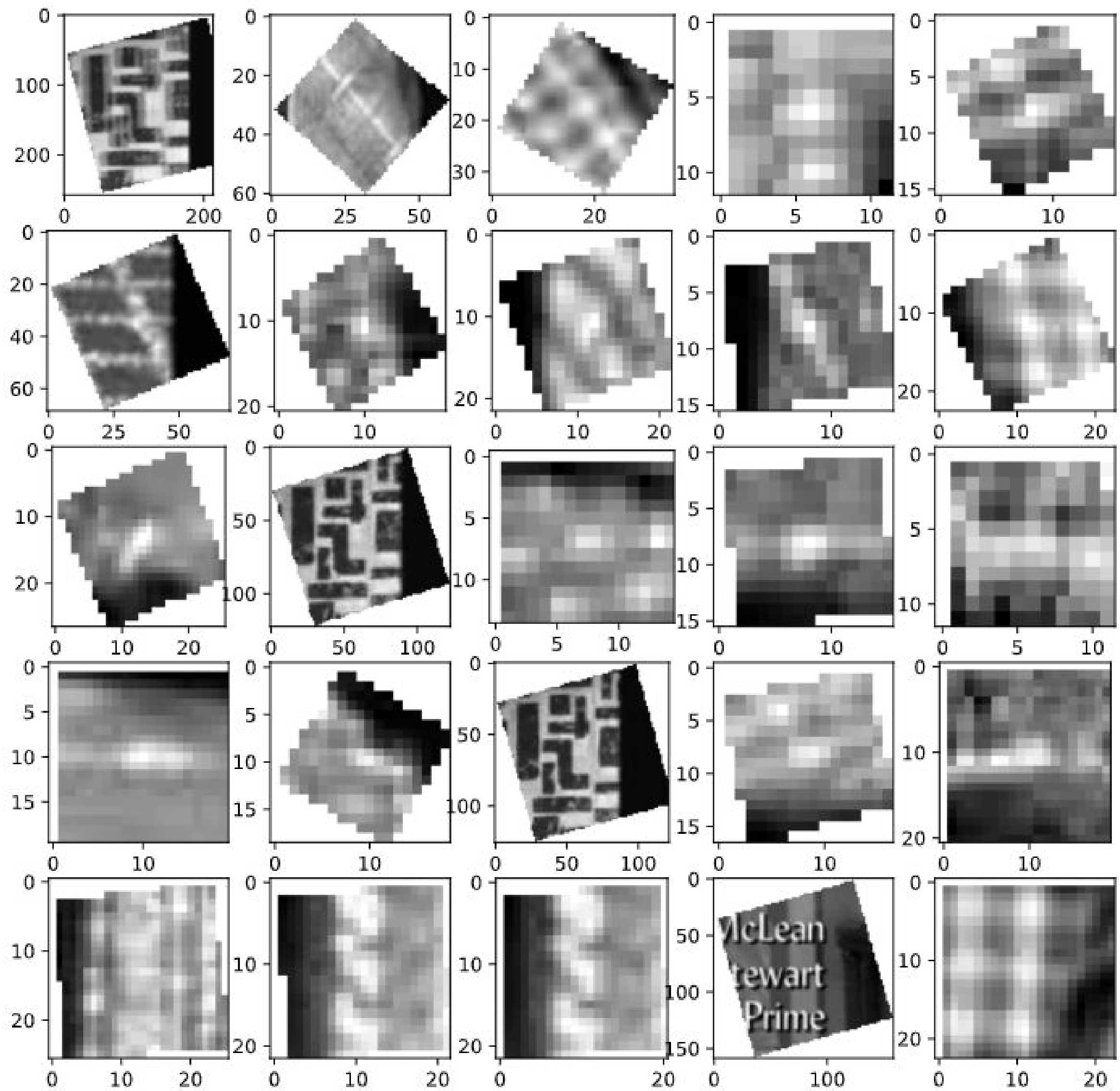


Since there are a lot of corners on the fridge area, you can see that most of the interesting points in the frame are concentrated on the fridge, with some similar features scattered around the frame.

Question 2

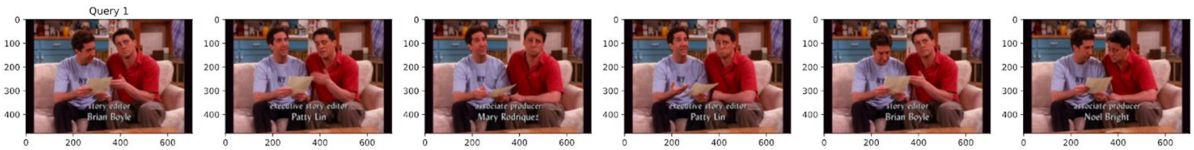


Here we see the straight line pattern common across some frames. The patterns aren’t an exact match because the corresponding object may have been shifted, rotated, and scaled from frame to frame.

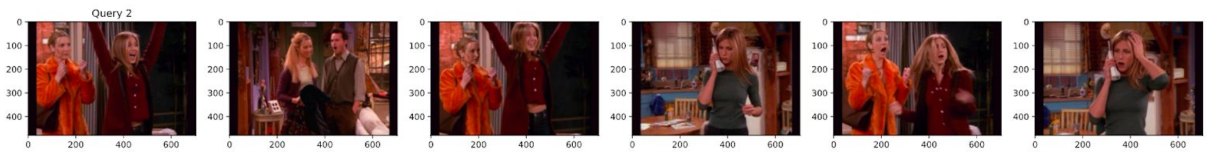


This pattern is a little more varied. We have a few instances where the matches are very close. In general, this patch is a checkered pattern with many corners. One of the reasons the text patch may have been included is because of the sharp contrast with its background in that frame, giving it a checker like pattern.

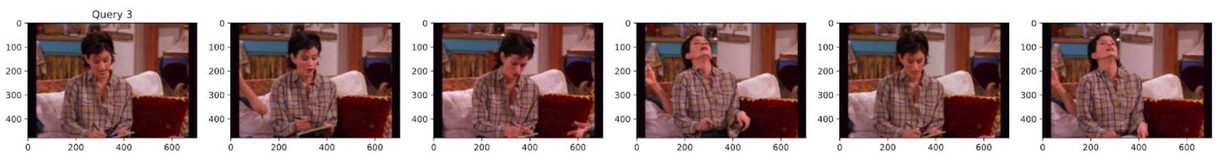
Question 3



Query 1: Similar frames were recognized very easily. The same 2 actors in different positions but the same room.

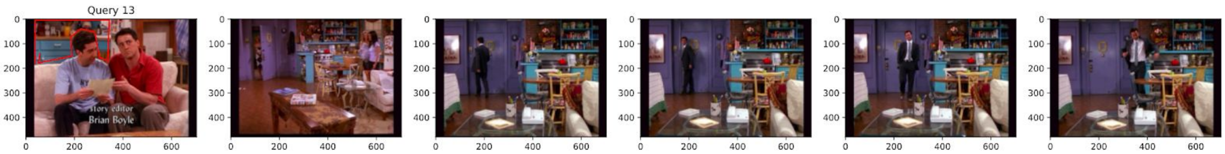


Query 2: The results were a little more wonky here. The same actress appears in all the frames, but the frames vary a lot more. The background changes and clothes change as well.

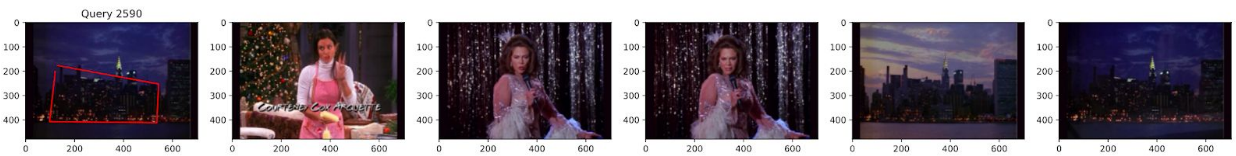


Query 3: Like the first, this result was very good. The same actress is simply shown in different positions but the 2 images were similar enough to be clustered together.

Question 4

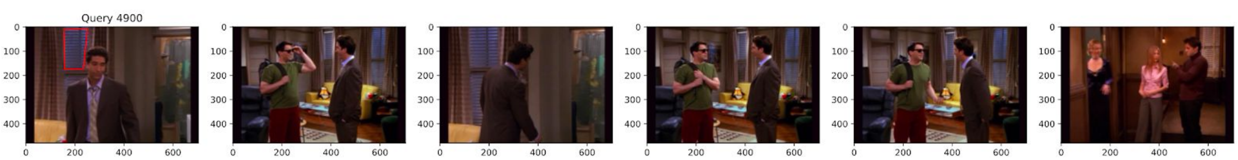


Query 1: Here we see the same cabinet object in all the pictures. Our program was successfully able to identify the selected descriptors in completely different scenes.



Query 2: One of the shortcomings of this algorithm is seen here. The lights from the skyline are very similar to the pinpointed interesting points in the Christmas tree and the sparkly backdrop for the dancer. Thus we have some inappropriate matches.

Query 3: I believe this is an edge case because the flowers on the table that were highlighted probably don't show up at any other time in the video, causing only the same scene to be shown.



Query 4: This is a great example of the algorithm working. The blinds are shown across different scenes, even with different people. Furthermore, the blinds have been shifted, slightly rotated, and scaled in the different frames but they are still recognized, which is amazing.