Jorge Betancourt
PS4

Short Answers

1) Scale-space is used for determining the scale of the of the interest points.

   In the case of using any local maxima in scale-space may not define the interest point for each run, the scale of the interest point may end up having different values. This will lead to less distinctive but more repeatable features.

   If the threshold value is too high, local maxima may not pass and the result will be an empty set for the interest point. For that reason, using thresholding will be more distinctive (since all the candidates on different frames will be similar.) but less repetitive.

2) An inlier in this case is a line that minimizes the epipolar constraint. However, with uncalibrated view, not all the epipolar line will meet the same point. This means the distance between the line and the corresponding point will be the determining factor for the outliers for each point.

3) One point of failure can occur if the scanlines end up being in the texture-less areas or an otherwise smooth surface, it is difficult to associate image to each other. Another point that should be considered is the material types of the objects in the scene. If the objects are non-lambertian (relating to the way light reflects off surfaces), it may become difficult to match the windows.

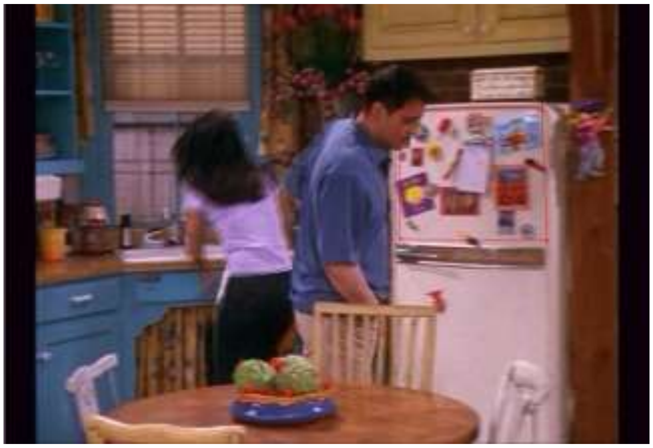
4) SIFT descriptors are 128 dimensional vectors, where each dimension represents one of the 8 gradient directions in a 4 by 4 neighboring region.

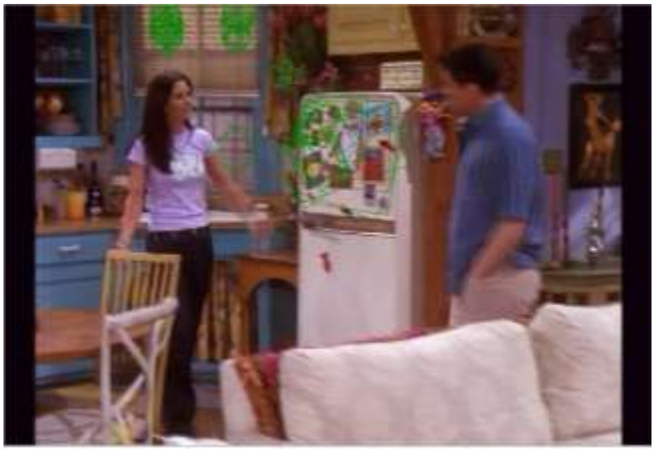
5) The dimensions of the Hough space would be 4. The position of the descriptor in the x and y, scale and orientation of the model will be searched in order to find the best model.

Jorge Betancourt
PS4

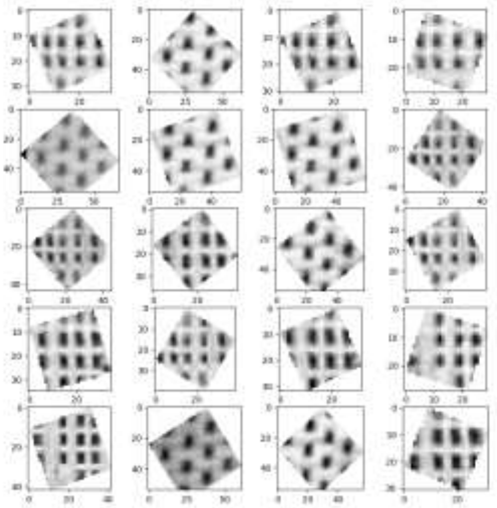
Programming

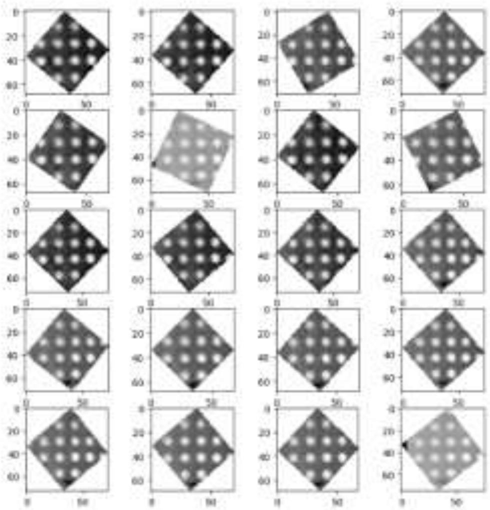
Raw Matching





Visual Vocab





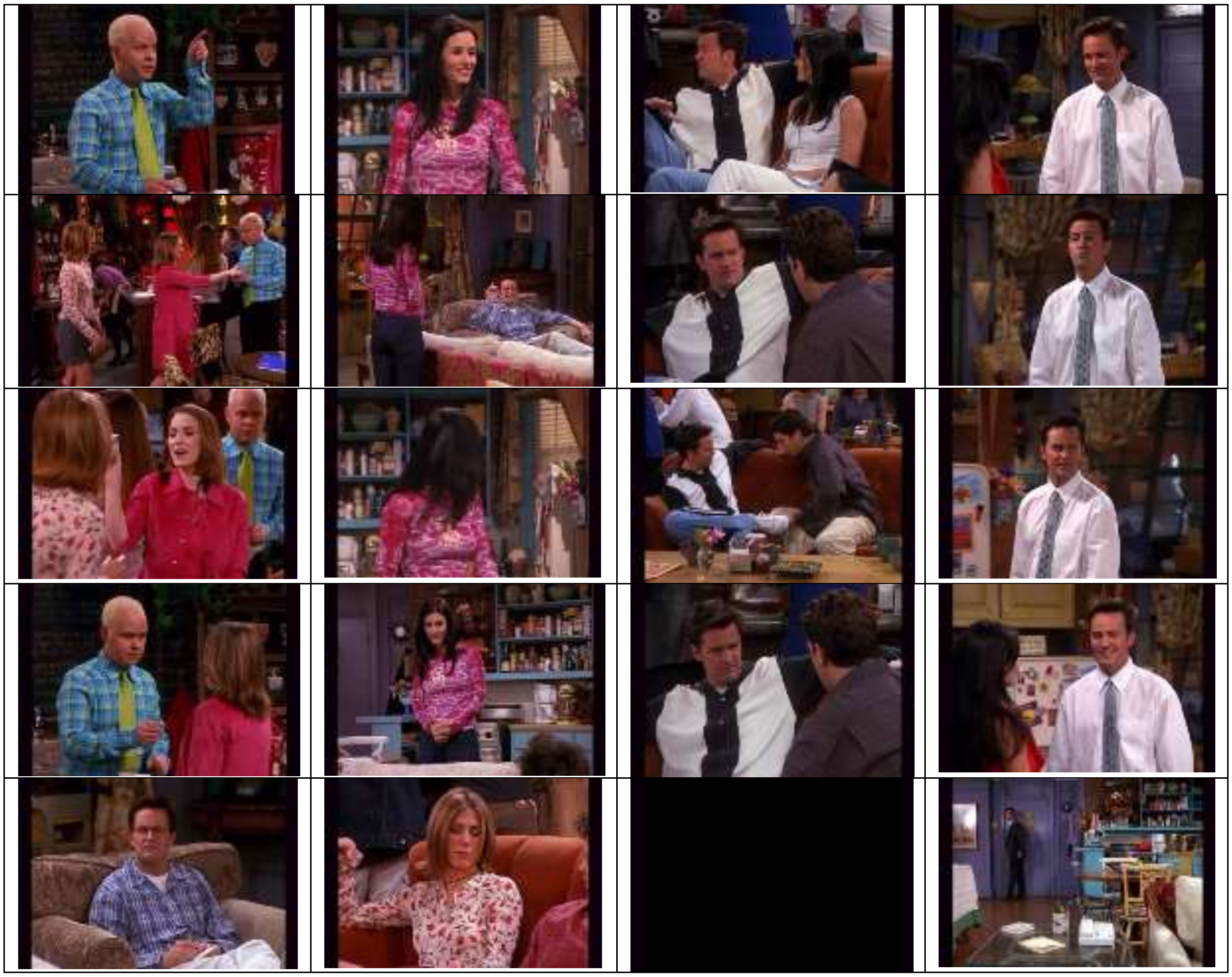
The word on the left is a squarish mesh with black centers. Probably a close-up of the wicker chairs in the restaurant. The word on the right is white polka-dots on a black background, probably from a dress or a curtain. These words are grouped because they all have the same texture and therefore similar gradients.

Jorge Betancourt
PS4

Full Frame Query

In the above figure, you can see the results from when I tried to find frames similar to the dinner scene, the pink top scene, and the Joey’s blue shirt scene. I took all the descriptors and placed them in the bag of words. Then I compared the histograms between that image and the other frames. The images that are underneath the top most image are the frames with the highest similarity scores. There similarity is based on their descriptors, not what the human eye sees as similar. This means images with similar gradients seem to have a higher similarity.

Region Query



Jorge Betancourt
PS4

The region query results are displayed above with the last row being failure cases. I chose to track clothing because the gradient stays consistent while giving many different perspectives and amid different backgrounds. I took descriptors from within selected regions and built a histogram from the patch. Then I compared it to the full histogram of the other frames. The similarity scores were overall lower, but the ones with the patches in common still outranked other frames due to higher similarity still outranked the other frames. The occasional error can usually be explained by gradients that look similar to the region. Similar clothing or furniture that looks like a rectangular tie explain most of the errors I encountered.