Computer Vision Assignment 4

L. Carter Price

11/2/2019

Short Answer Questions

1. 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(b)take any positions whose filter response exceeds a threshold? Specifically, what is the impact on repeatability or distinctiveness of the resulting interest points?

    If we were to take positions based on a threshold, then there is a possibility that you would get multiple interest points at the same location because results at different scales spaces may still exceed the threshold. This negatively impacts distinctiveness where with the local maxima method, we will only have a single distinct position taken by definition.

    Additionally, if we take position based on threshold and the lightness of the image changes, then the points detected will change. The threshold value may detect more or less interest points. This is also true if the threshold is changed. As a result, the thresholding method results in poor repeatability across images. Alternatively, the local maxima method will return the same results with changes in lightness and will not be dependent upon any external parameter like a threshold value making the local maxima method much more repeatable.

2. 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?

    An “inlier” is a correspondence pair that agrees with the calculated fundamental matrix within a certain threshold. The inliers are computed by taking one (x,y) point and then transforming it through the matrix and comparing this result with the original matched correspondence point. If the error between the result and the original matched correspondence is less than a certain threshold then we say this is an inlier. For further clarification, given a correspondence (x1,y1) in image 1 and (x2, y2) in image 2. Multiply (x1,y1) by the Fundamental Matrix to calculate (x2’,y2’). Compute the error between (x_2’,y_2’) and (x2,y2). If the resulting error is below a certain threshold, we say it agrees with the Fundamental matrix and is an “inlier”. We repeat this process until a fundamental matrix is found with the most inliers. From the fundamental matrix, the epipolar lines can be found for each correspondence.

3. 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.

    1. Textureless surfaces – In textureless surfaces, you may not be able to find a unique match for each correspondence.
    2. Repetition – Similarly with repetition there may be multiple matching correspondences within the search window.

    In both cases, these points are no longer useful or give bad/incorrect data that causes the stereo estimation to fail.

4. What exactly does the value recorded in a single dimension of a SIFT KeyPoint descriptor signify?

    The value signifies the number of gradient values (a histogram bin value) in a region for a particular gradient orientation. There are 8 different gradient orientations considered. For a 4x4 grid, each cell counts, normalizes and weights the gradients in that region for a particular gradient direction. Then this is done for all 8 gradient directions. For example, the scalar in cell 1x1 might be the histogram bin value for gradient number 2 in the region occupied by cell 1x1. (4x4x8 = 128)

5. If using SIFT with the Generalized Hough Transform to perform recognition of an object instance, what is the dimensionality of the Hough parameter space? Explain your answer.

    The Hough parameter space is 4-dimensional -- x-translation, y-translation, orientation and scale. Each feature match gives an alignment hypothesis for these 4 dimensions. We then vote in this space to verify the correct object recognition with this special verification.

Programming Problem

1.

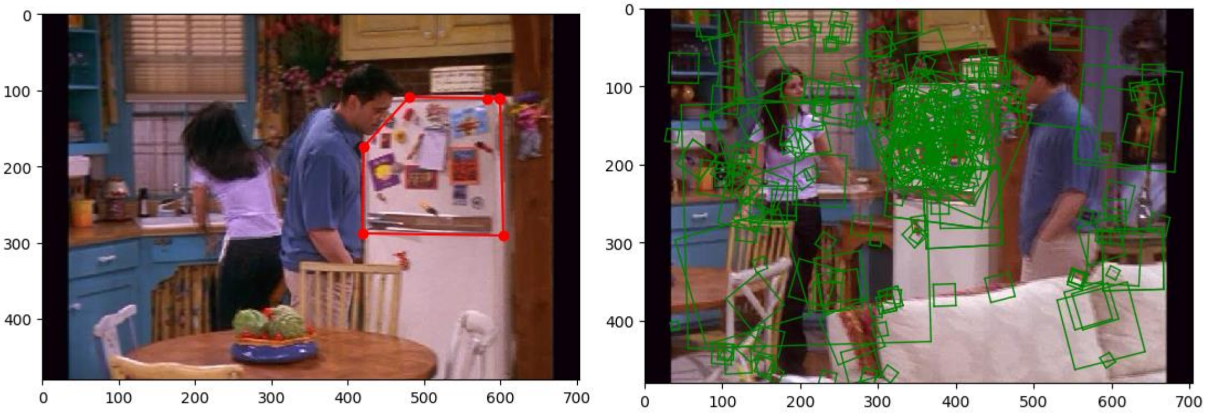

Figure 1: The left shows the region of interest outlined on the fridge. On the right, the image shows when each descriptor from the ROI is mapped to a descriptor in the entire image. It clearly shows many incorrect matches and noise throughout the image.

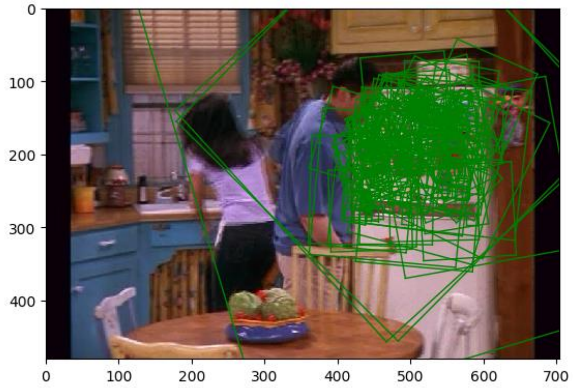

Figure 2: This shows the features displayed in the Region of interest for reference to compare to Figure 1.

2. Visualizing the Vocabulary

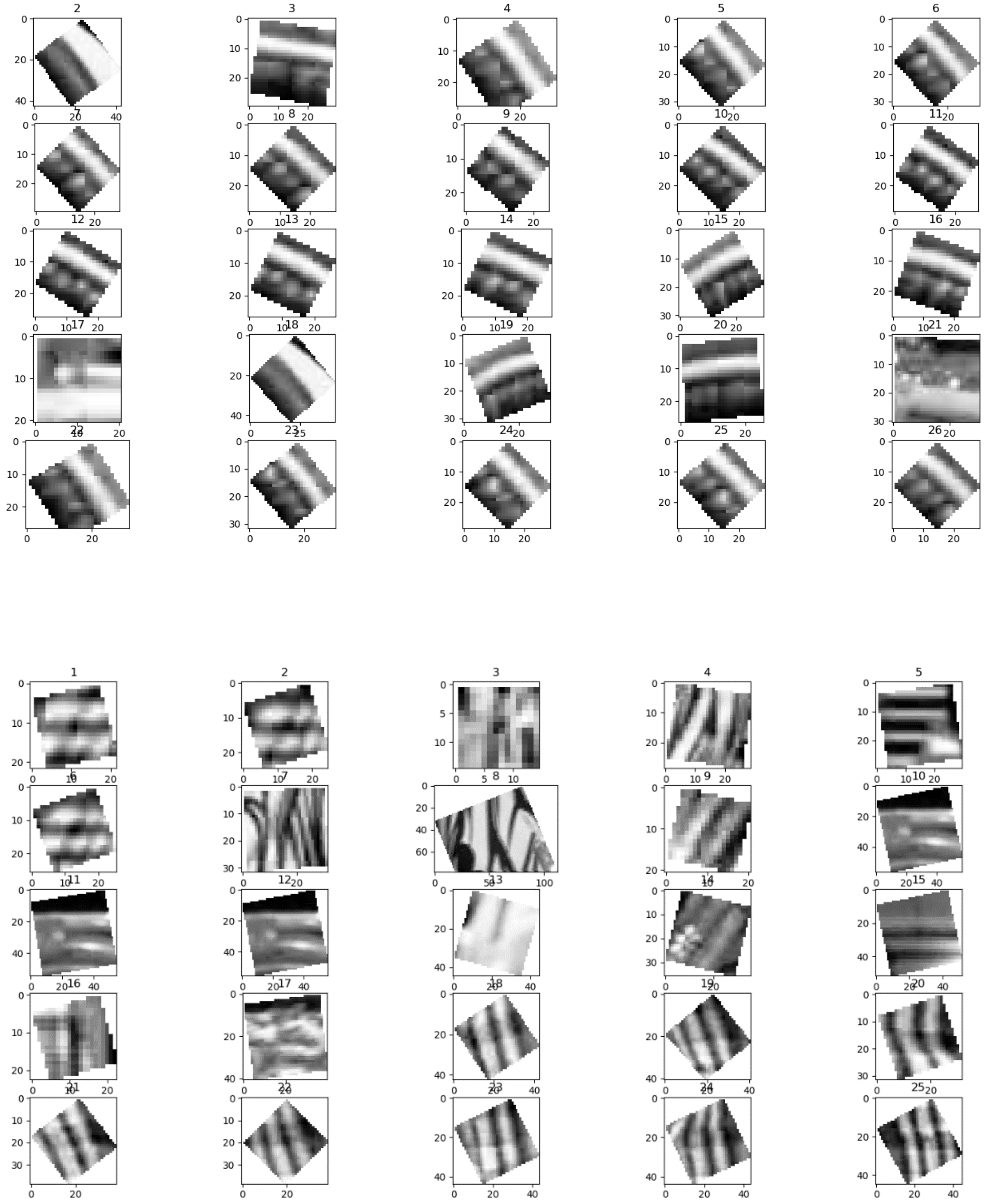

Figure 3: This is a visualization of the vocabulary used throughout the course of the images. The SIFT descriptors throughout the entire set of images were randomly sampled to build this vocabulary. Displayed on the above are two of the most common descriptors. The top shows a descriptor which appears to have a bright thick white bar to one side of the square descriptor region. The bottom images show a striped descriptor.

3. Similar Pictures

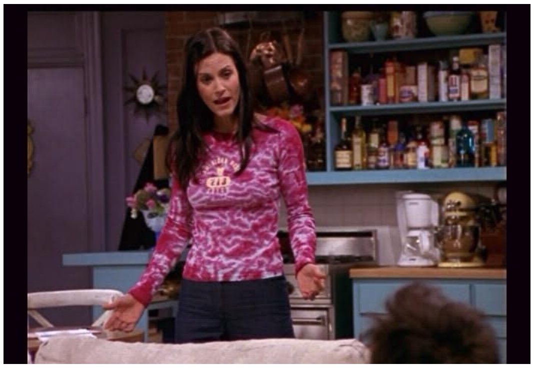

Figure 4: Original picture of Monica.

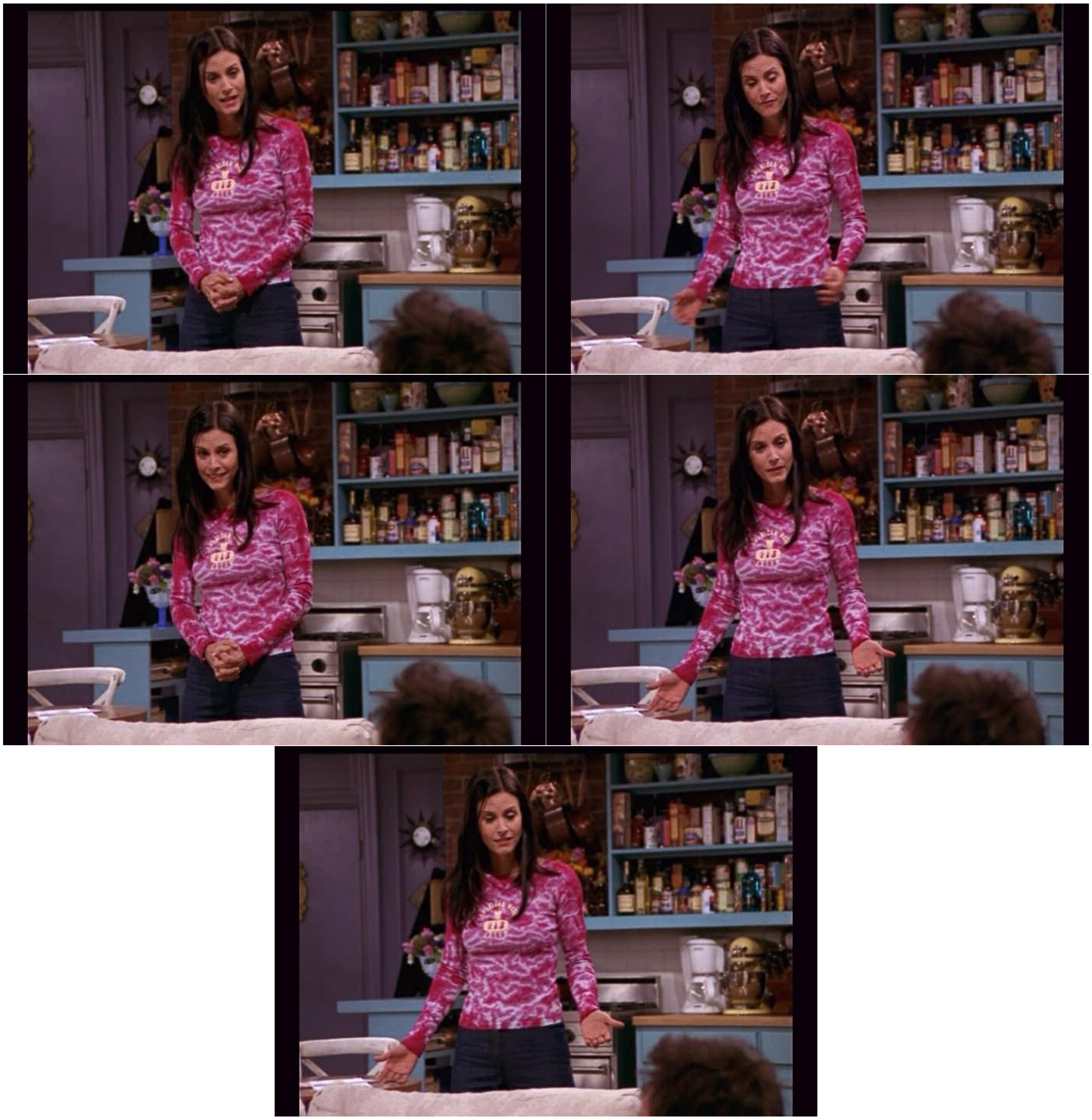

Figure 5: Above find the 5 most similar pictures to one of Monica. The top result is returned in the top left of this figure while the 5th ranked result is at the bottom. (This will be true for the remaining figures with a similar format). The descriptor does surprisingly well in this case, and I suspect it is due to the unique pattern on Monica's shirt.

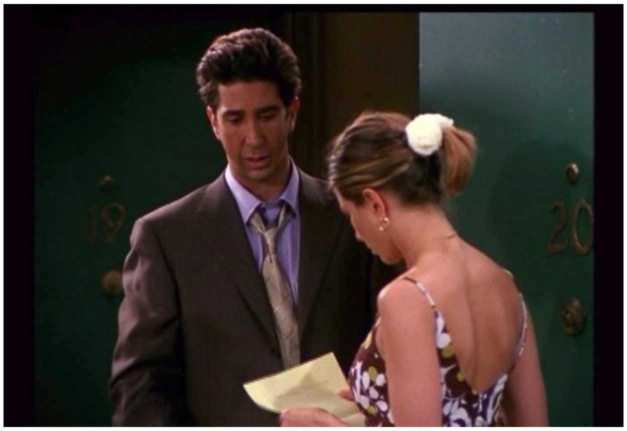

Figure 6: Original Picture of Ross and Rachel.

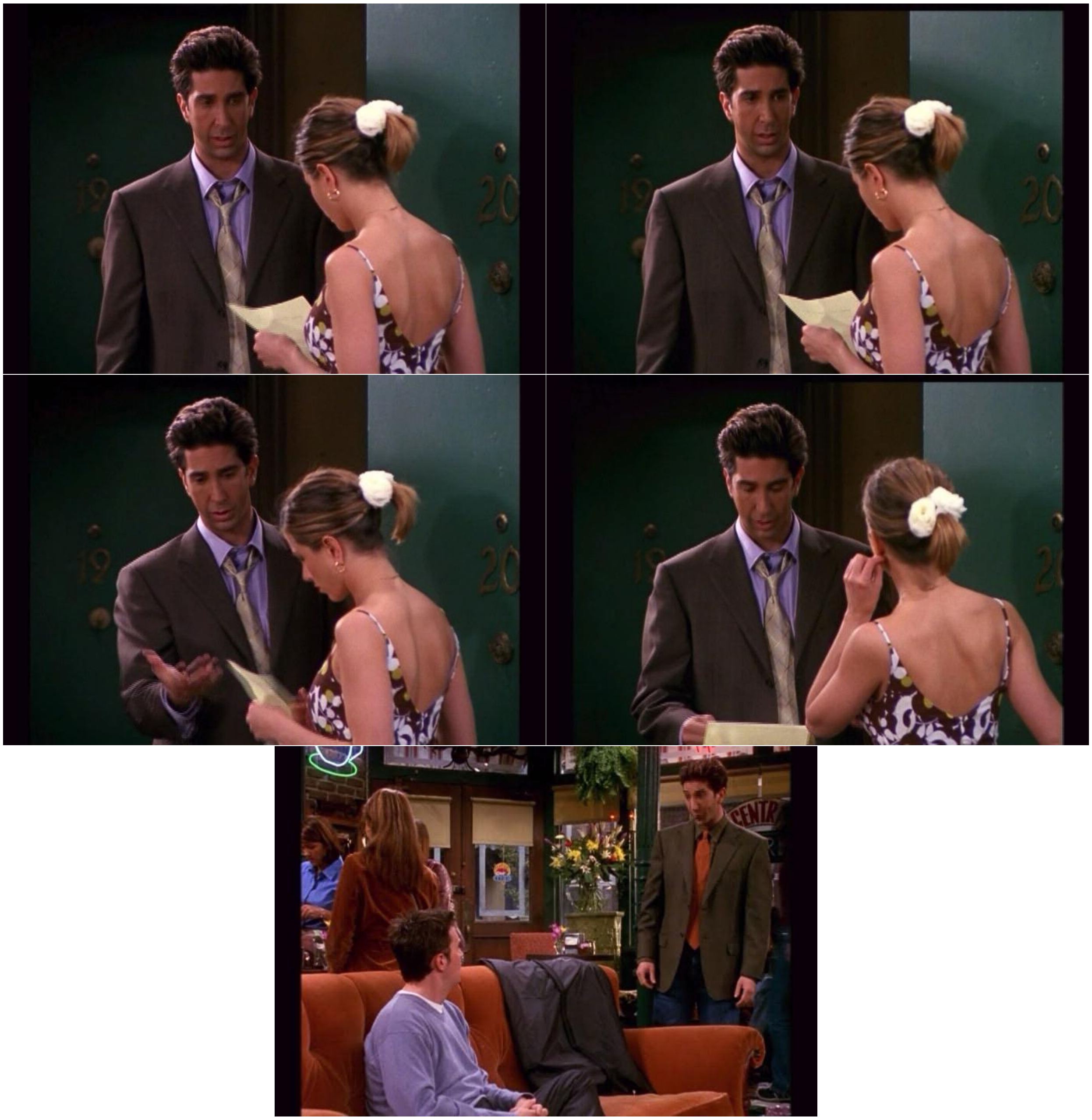

Figure 7: Five most similar pictures to Ross and Rachel. As can be seen number 5 is quite different, but it appears that the features are still picking up Ross in a suit.

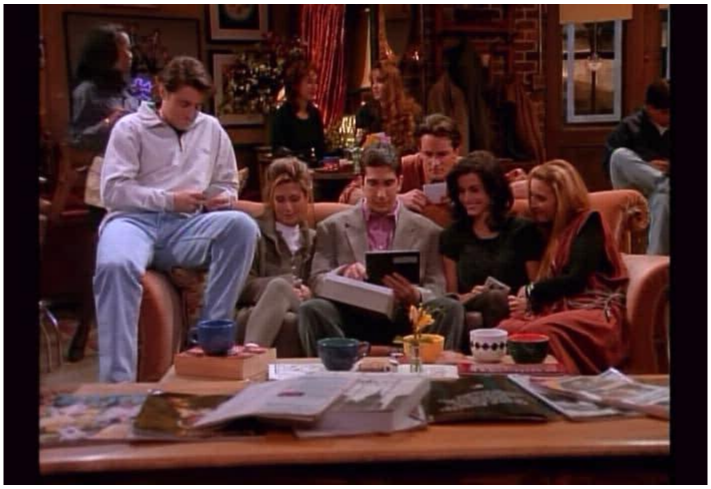

Figure 8: Original picture of the gang on the couch.

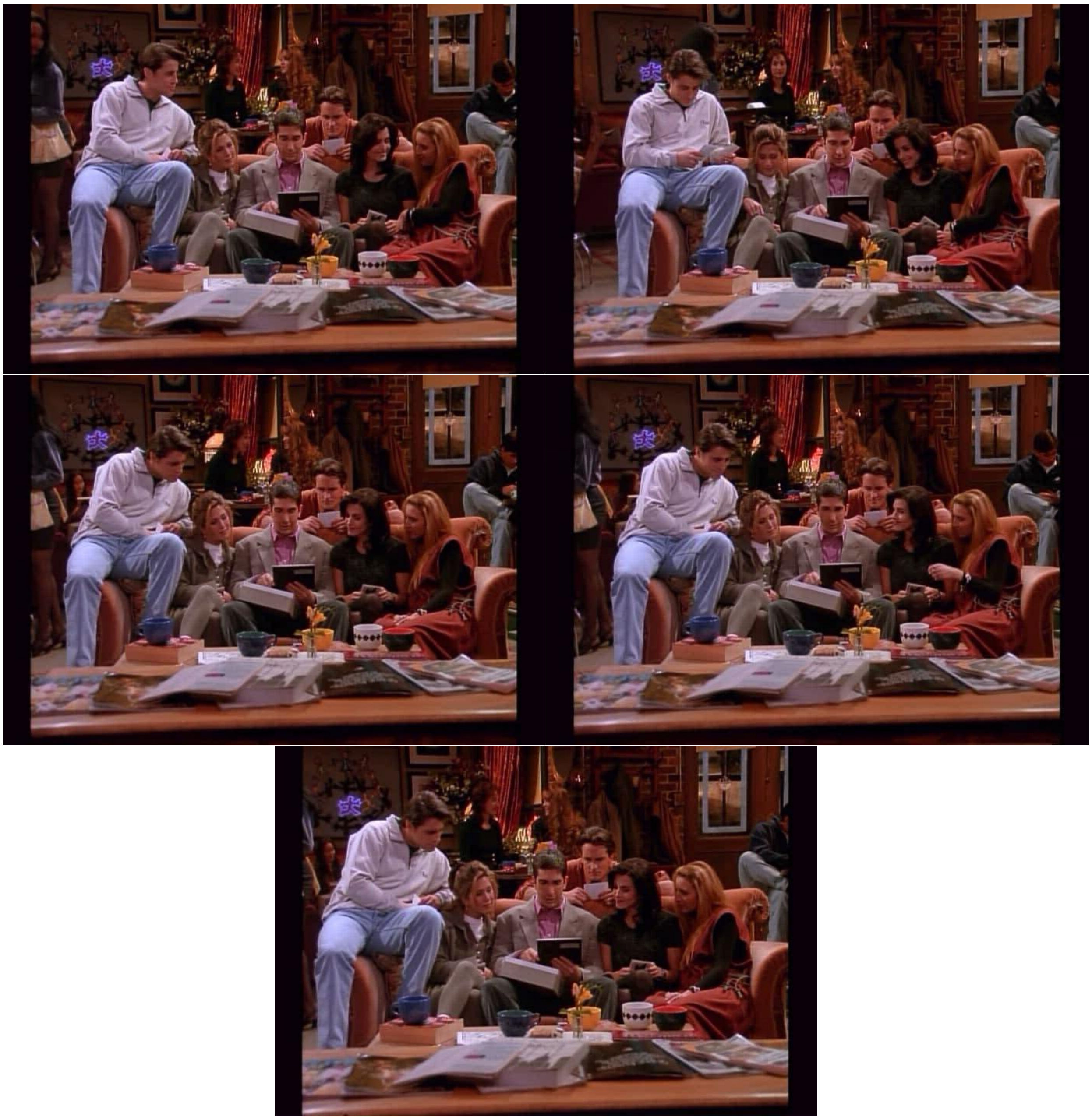

Figure 9: 5 most similar pictures to the gang. All very similar, but the distinctions can be made out in the head positions of the cast. The consistent foreground of books and magazines likely helped to improve this outcome.

4. Region of Interest matching
   Good Examples:

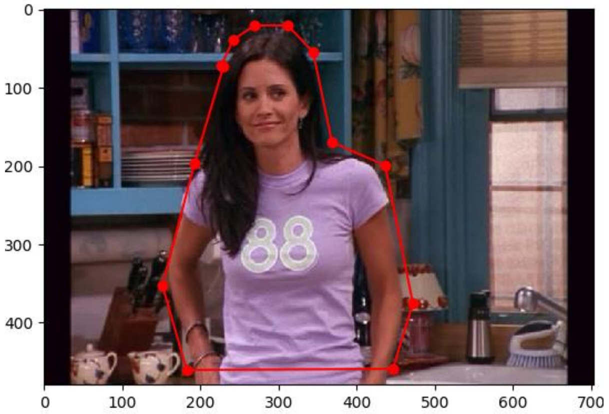

Figure 10: Region of interest Monica standing in the foreground.

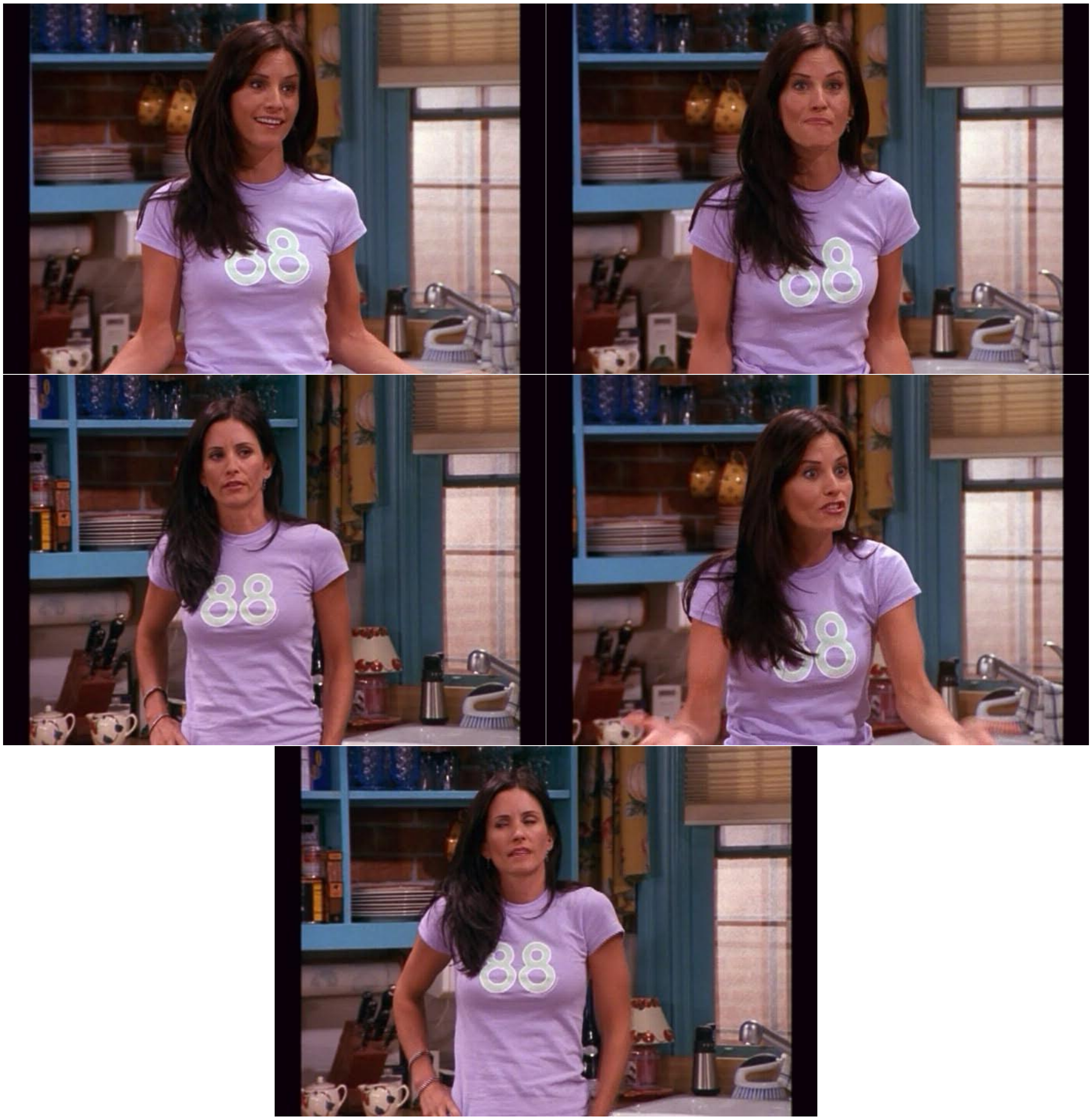

Figure 11: Top 5 results for Monica. Success is likely due to the unique 88 pattern on her shirt. We will see the same image in a failure case later.

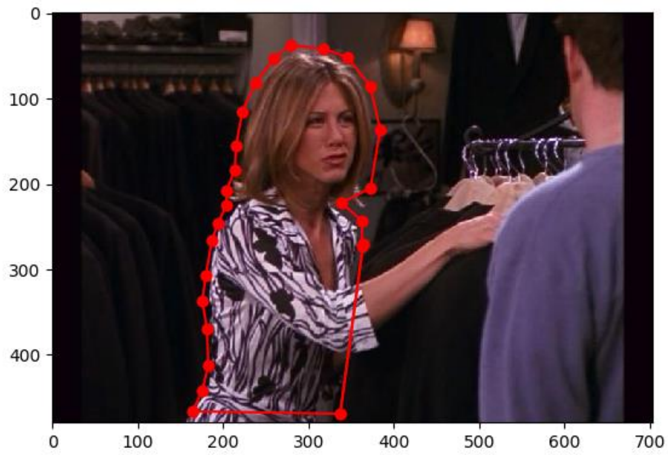

Figure 12: Region of interest around Rachel.

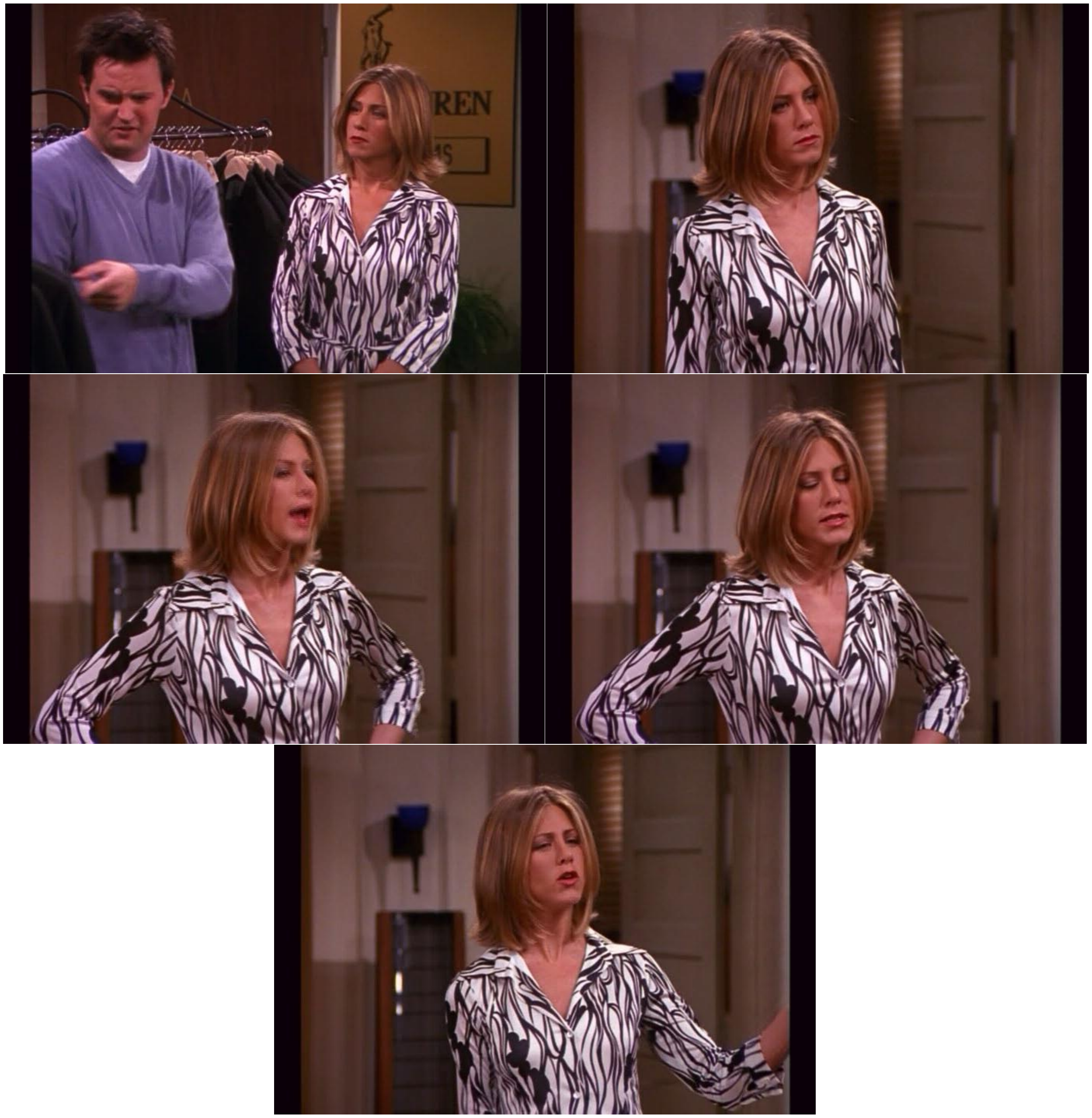

Figure 13: Again in a similar fashion the results are successful due to the unique pattern in here shirt. Here we see a different background in the first image.

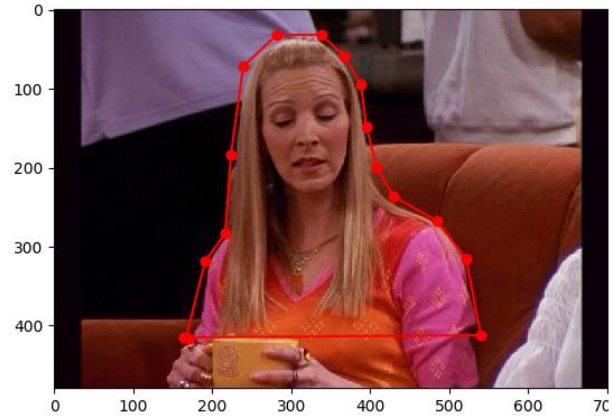

Figure 14: Original Region of interest around Pheobe.

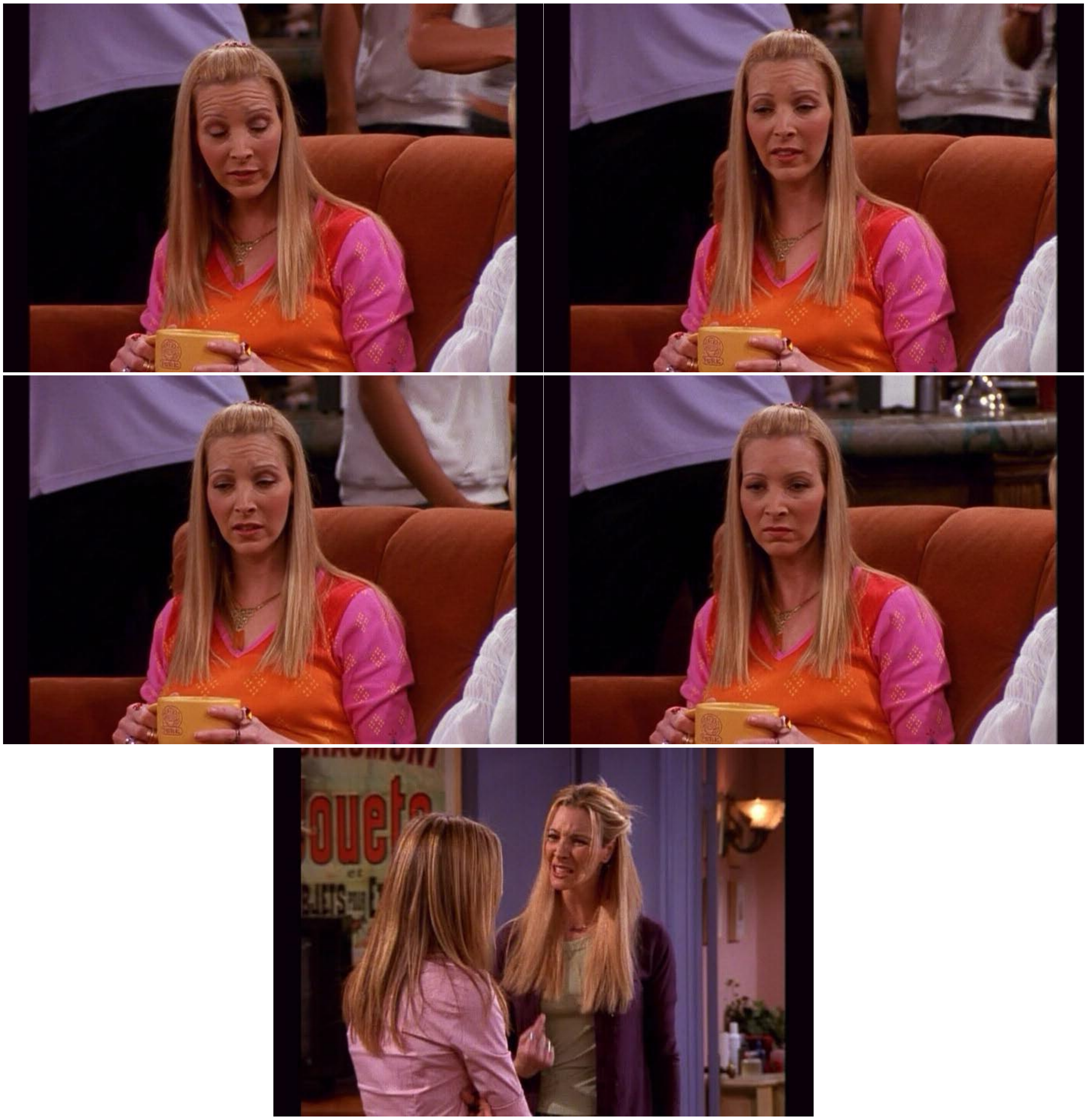

Figure 15: Here again we see positive results. However, I suspect that we are not relying on the features in the shirts as in some of the others. Here it seems that Phoebe's long blonde hair may be the defining attribute that most descriptors are attached to. In image 5, at the bottom we see a new background and different facial expression, but similar hair. Notice it did not pick up other pictures of Phoebe when her hair is up.

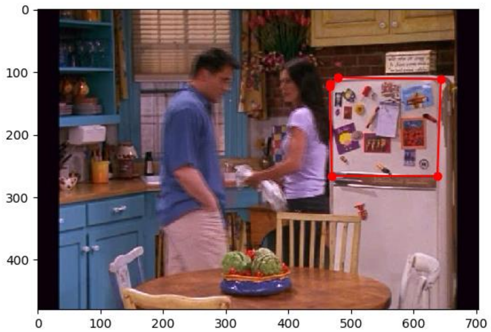

Figure 16: Show the Region of interest for the fridge magnets.

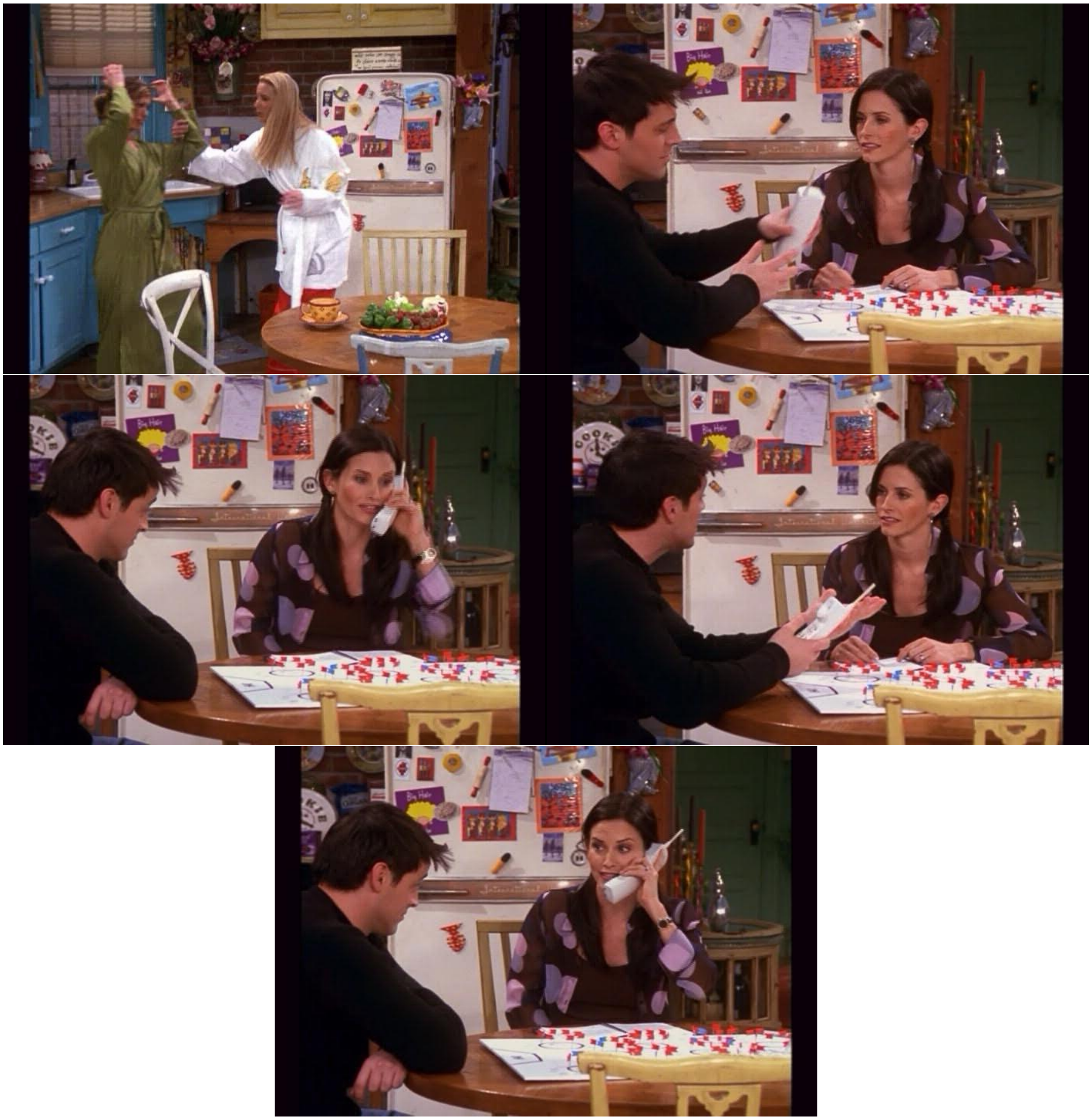

Figure 17: Shows the 5 returned images from the top 5 returned images from the region query. The results are successful as the fridge magnet descriptors seem to be unique across the vocabulary. Even with the wide changes in background and scale.

Failure Case:

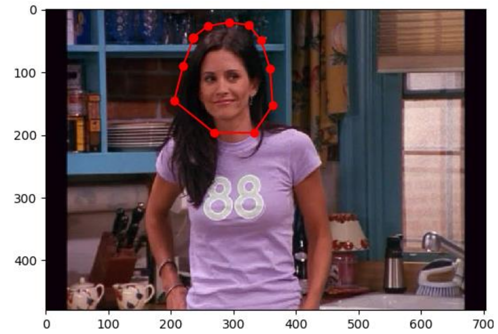

Figure 18: Region of interest is now just Monica's face. Compare with figure 10 which include's Monica's torso.

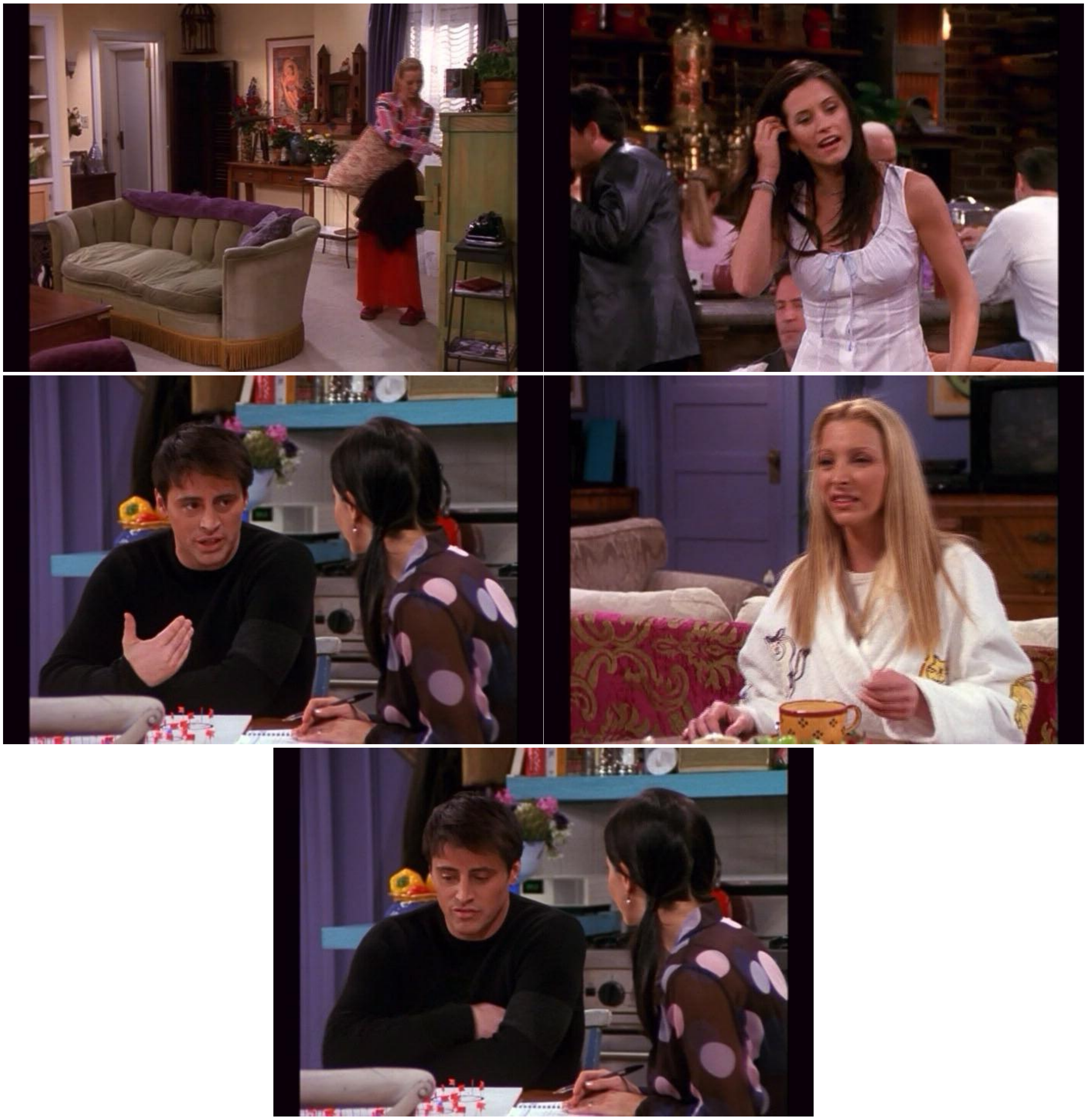

Figure 19: Top 5 result for Monica's face ROI. Here we clearly see that this did not pick up Monica's face very well. We do see that there are consistently faces present, but the descriptors do not seem to be rich enough to distinguish different faces well. For the top returned image, it appears that the drapes in the background may be returning features similar to how Monica's black hair is draped around her face in the ROI.

Comments:

It seems that the SIFT descriptors are great at picking up the unique textures and objects with the BOW approach. However, when there are similar objects with subtle differences, it struggles to differentiate. For example, with faces, they contain a similar structure but as seen in the failure case. This approach does not necessarily distinguish specific people based on their facial features.

It may be possible that this methodology can be improve and using SIFT features can be an accurate way to determine different faces. As shown below in the extra credit, tf-idf and stop list techniques offered marked improvements in performance, and I would imaging special verification would offer the same.

Extra Credit –

1. Implement tf-idf. Please find code included in zip. Shown below is the progression from the standard region query, to implementing tf-idf weighting and lastly including a stop list of the most common features in the vocabulary.

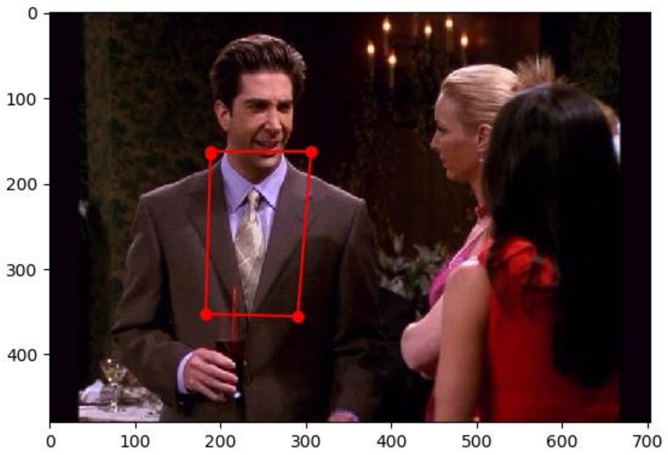

Figure 20: Region of interest for matching images. Use case to demonstrate the effectiveness of tf-idf.

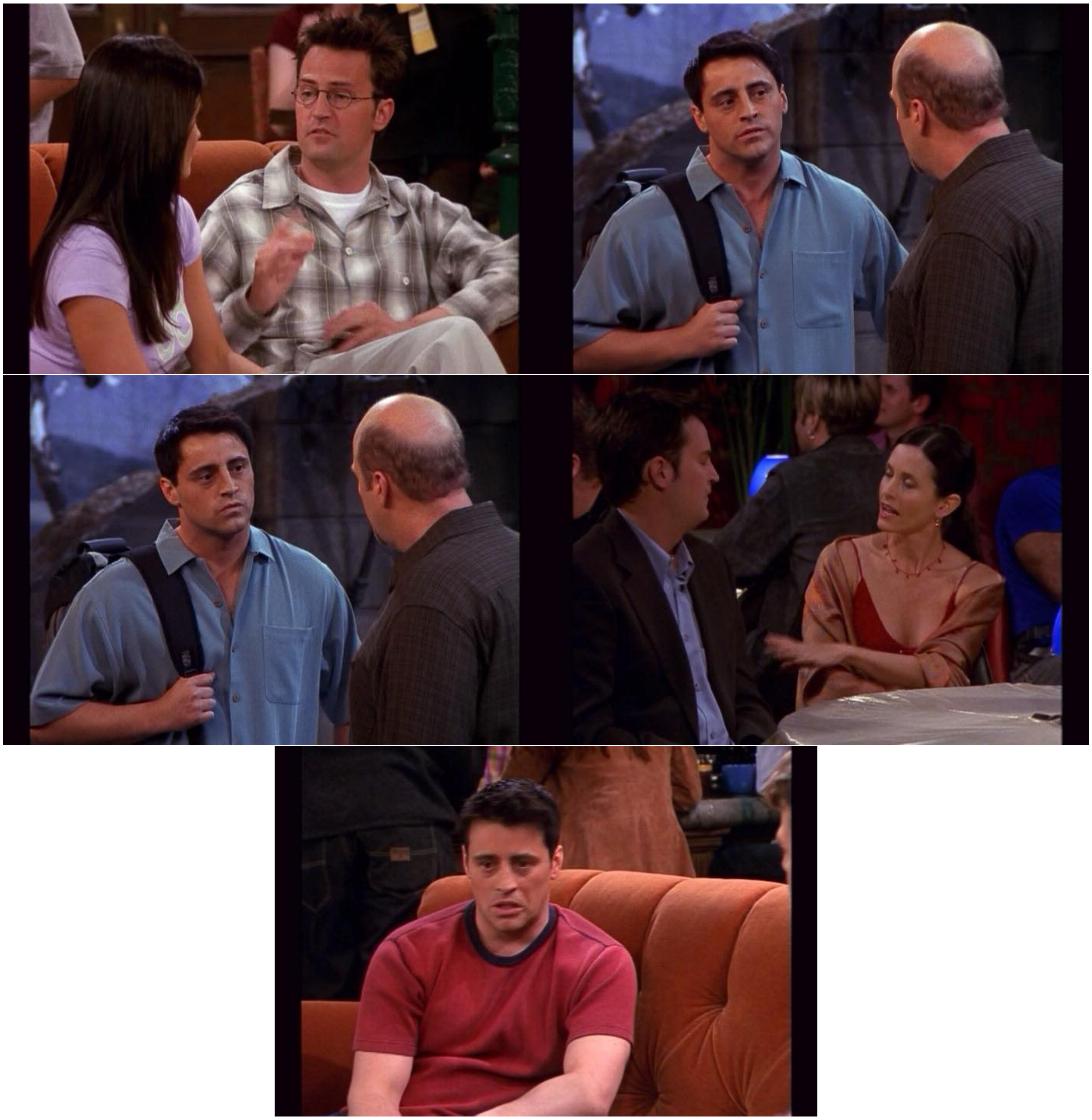

Figure 21: Top 5 results with running the regular region matching for Ross's tie. As you can see, none of these include Ross in his tie. It seems to be picking up other features like the shirt neckline.

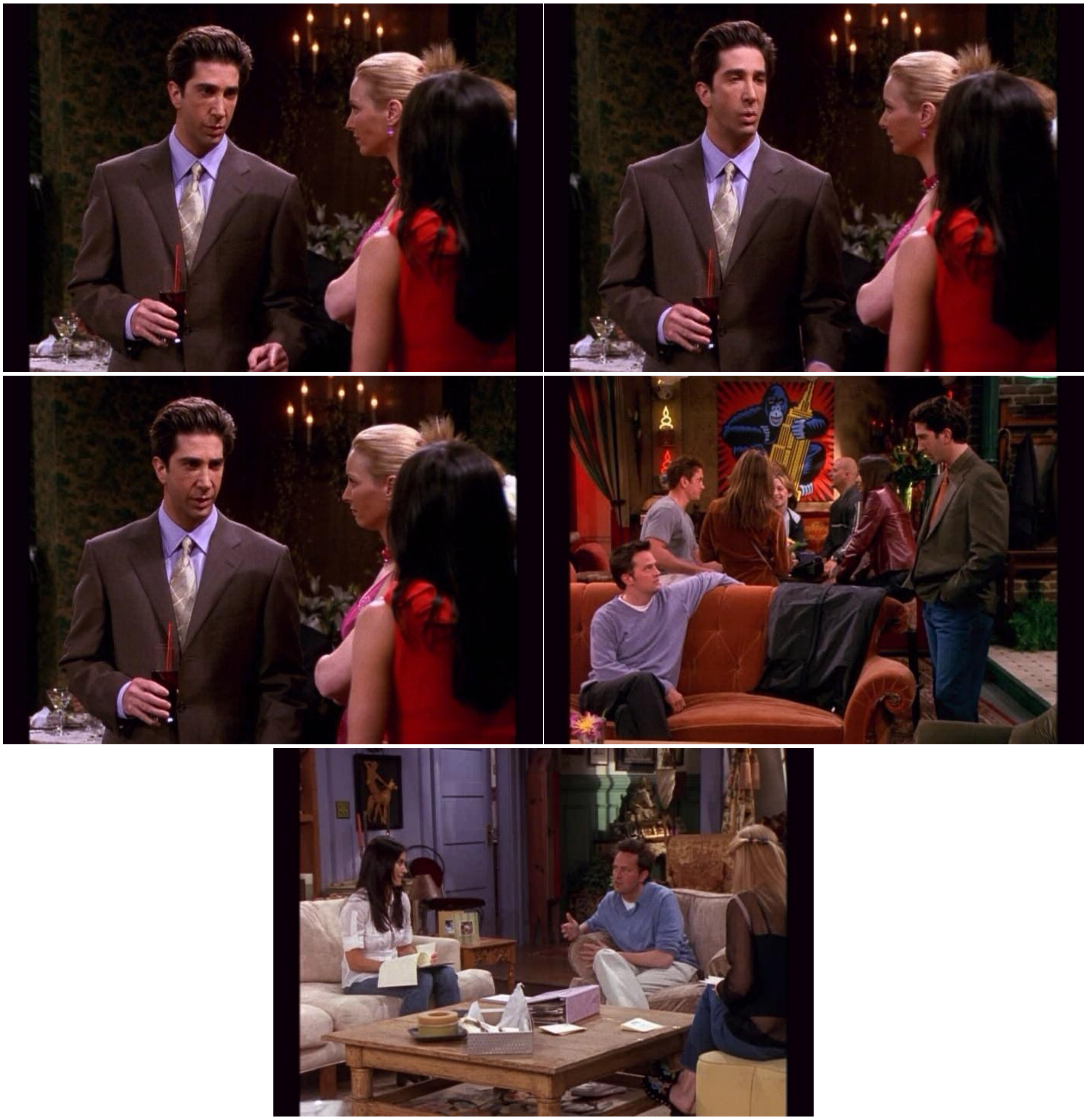

Figure 22: The above images are the top 5 matches when running the tf-idf addition on the features in Ross's tie region. As can be seen, this does significantly better than the

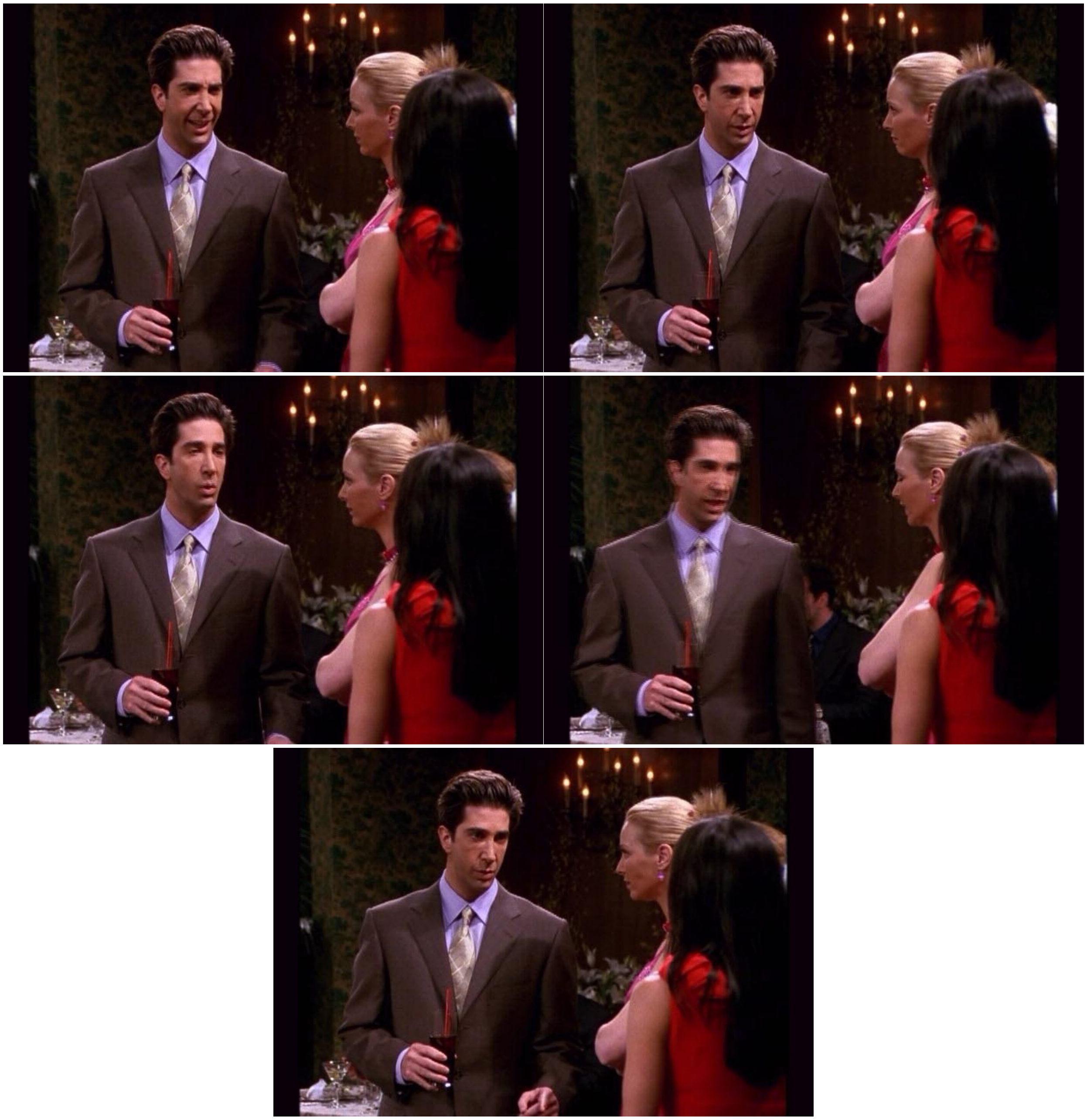

Figure 23: Here a stop list is implemented as well as the tf-idf. It show significantly improved results over the