

ECS 174: Intro to Computer Vision, Spring 2018

Problem Set 3

Ayush Surana - 916718143
Divya Ravilla - 915707281

1. Short Answer Problems

1.1

(a) Repeatability: Will be more effective in detecting the same features.

Distinctiveness: Very distinct features will be detected and used.

(b) Repeatability: Will be less effective when compared to local maxima.

Distinctiveness: Features are not that distinctive.

1.2

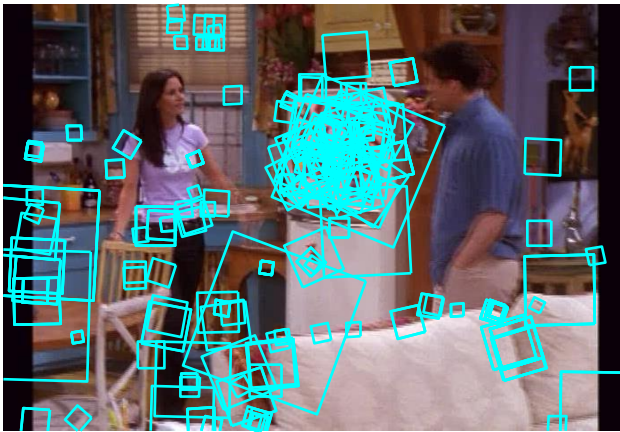
A 16×16 neighbourhood around the keypoint is taken and divided into 16 sub-blocks of 4×4 size. For each sub-block 8 bin orientation histogram is created, and stored in the descriptor matrix. For example the 1st orientation in the 1st sub-block for keypoint n is stored in $\text{descriptor}(n,1)$. Each descriptor for a frame is 128 dimensional matrix.

2. Programming: video searching with Bag of words.

2.1 The region selected by the user is depicted in the image below,



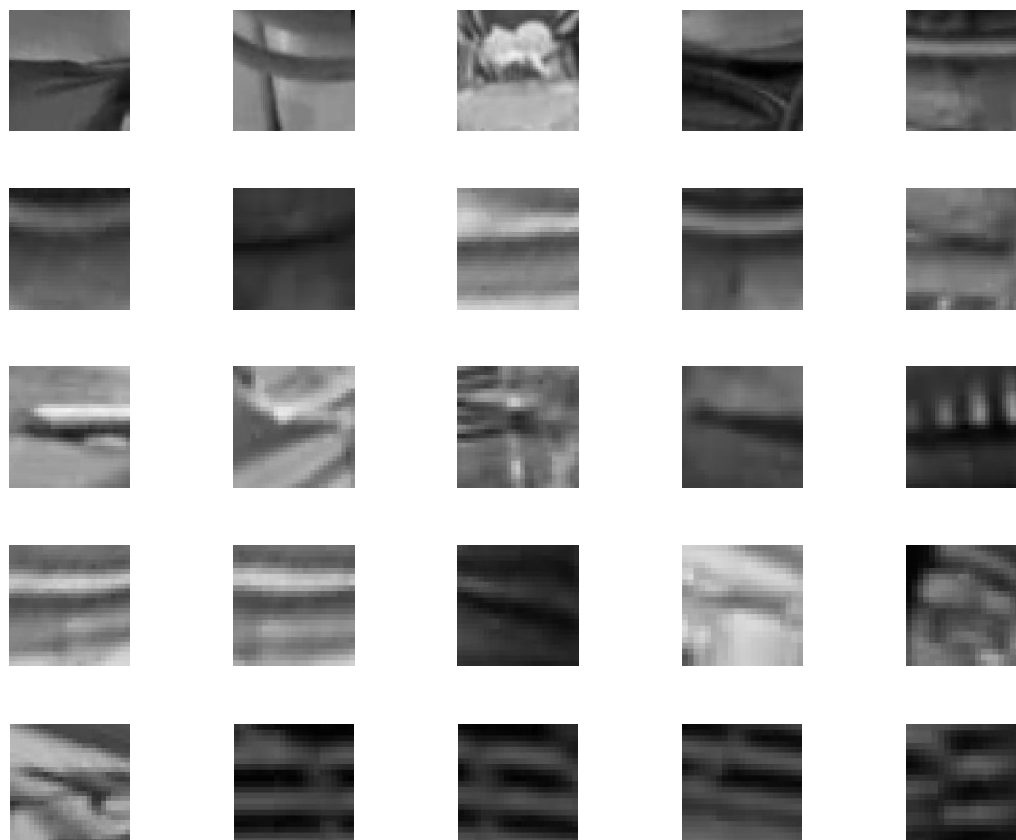
The output produced by `raw_descriptor_matches` is,



All the descriptors which are similar to the descriptors that lie inside the region selected by the user are displayed. The threshold value I used is 0.27

2.2

The following was obtained for word 1 and threshold 0.5



All the patches above show some horizontal lines. Descriptors of 25 interest points that are most similar to that of the visual word 1 are displayed.

The following was obtained for word 35 with threshold 0.6



All the patches above show similar corners. Descriptors of 25 interest points that are most similar to that of the visual word 35 are displayed.

2.3 In code: Change n for changing frame number ($n = 1 \dots 6612$)

For a particular frame number given in the frame folder(m) = $m - 59$

Results obtained for frame 1:

The histogram computed for image 1 is compared with the histograms of the other frames. 5 frames that are most similar to the image are displayed. All the images have almost similar objects. All the images contain the same visual words. The blue dress, the Friend's series character's (Rachel) face, her dress and other tiny objects are present in all the images.



Results obtained for frame 6:

The histogram computed for image 6 is compared with the histograms of the other frames. 5 frames that are most similar to the image are displayed. Rachel's face, her dress are present in all the pictures. Interest points on the chair and the other two people, may correspond to different visual words and may have been used to compute the histogram.

./frames/friends_000000065.jpeg



./frames/friends_000001237.jpeg



./frames/friends_000001913.jpeg



./frames/friends_000000065.jpeg



./frames/friends_000000066.jpeg

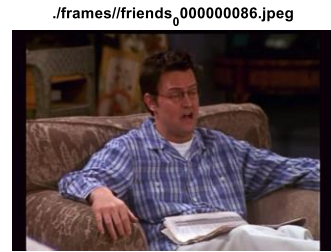
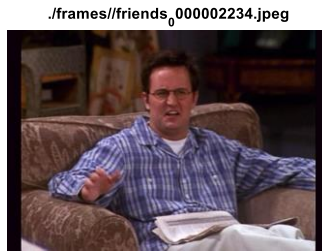


./frames/friends_000001676.jpeg



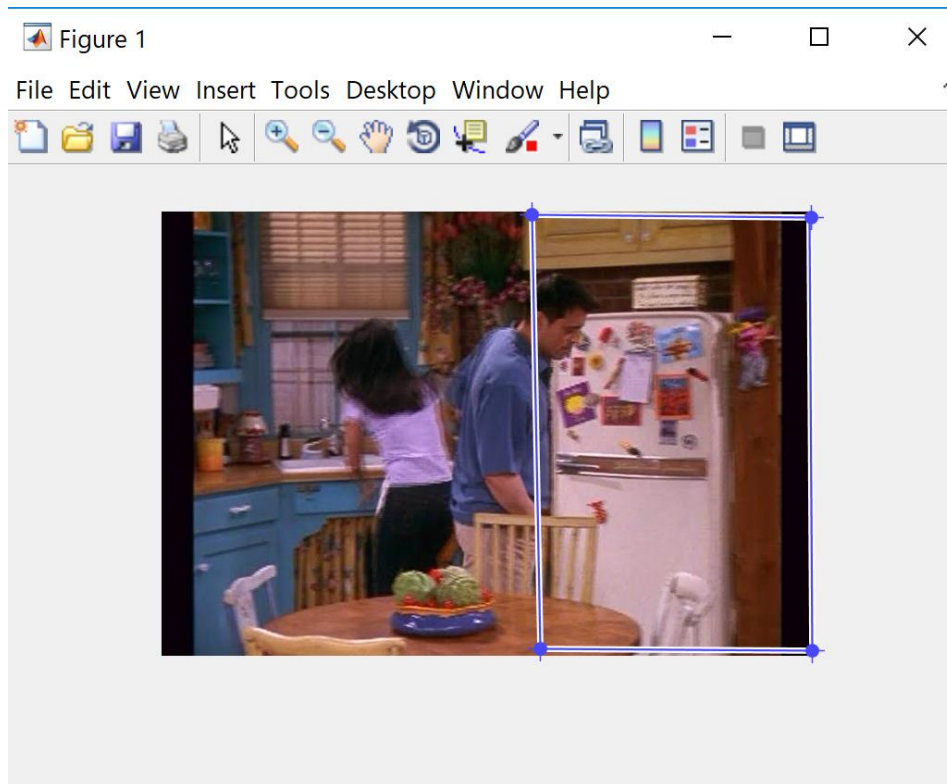
Results obtained for frame 24:

The histogram computed for image 24 is compared with the histograms of the other frames. 5 frames that are most similar to the image are displayed. All the images have Chandler in them and the background is similar in all the images. The histogram of all the frames are very similar.



2.4 Top 6 frames which contain the objects selected are displayed.

1. Region selected



./frames/friends_000000409.jpeg



./frames/friends_000000411.jpeg



./frames/friends_000006337.jpeg



./frames/friends_000000409.jpeg



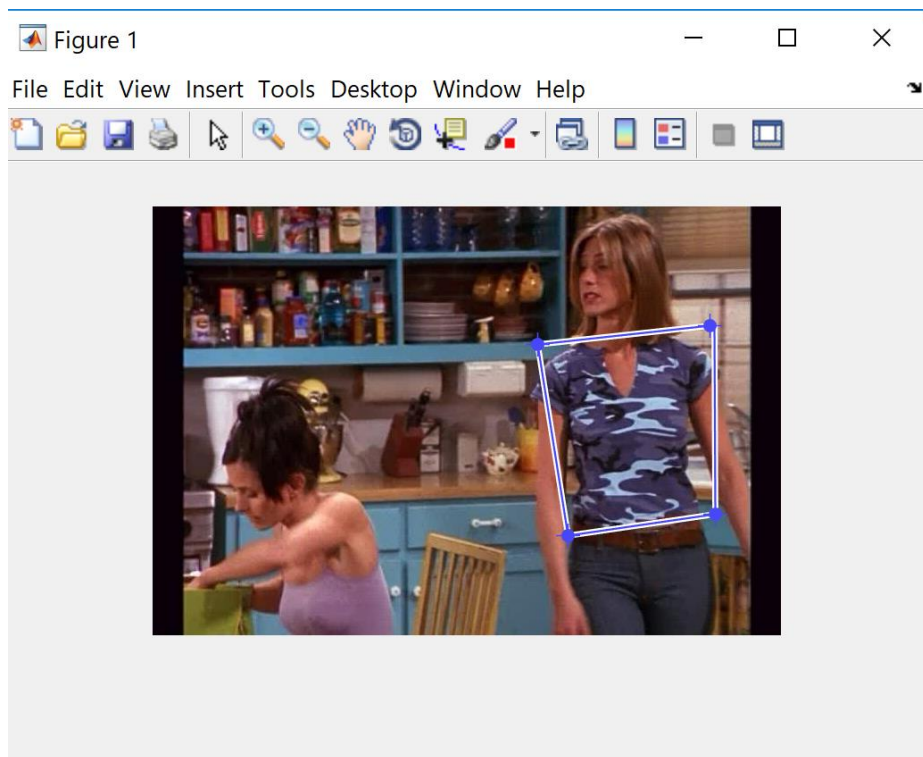
./frames/friends_000003603.jpeg



./frames/friends_000000424.jpeg

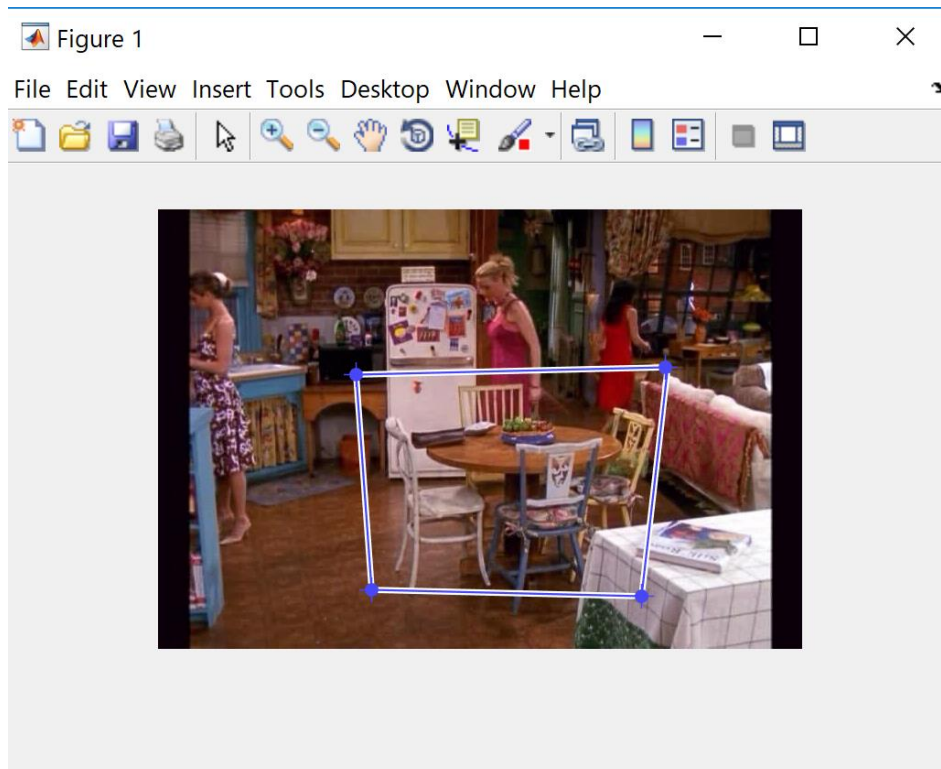


2. Region selected,



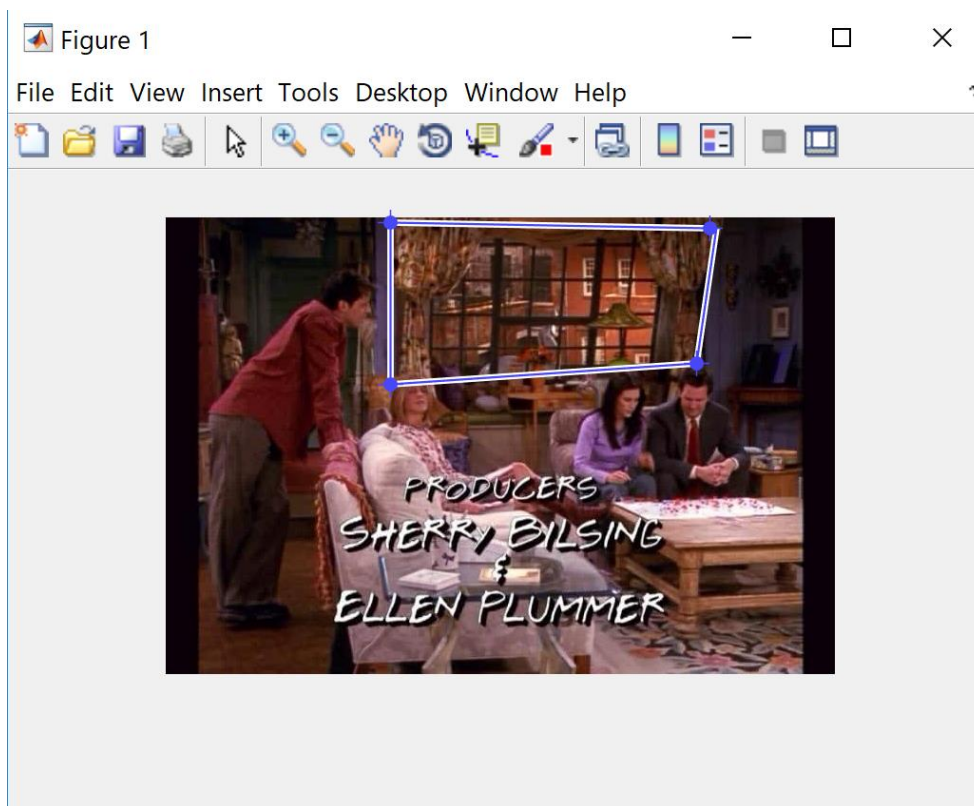


3. Selected Region





4. Selected region

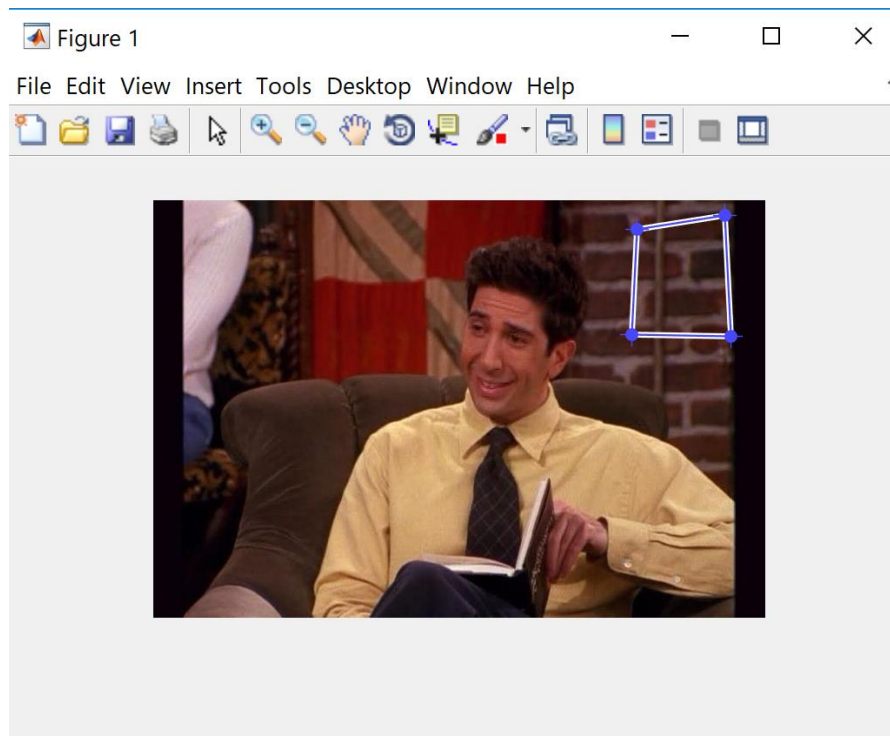


Frames with windows are shown below,



Failure case,

Selected Region



./frames/friends_000001007.jpeg



./frames/friends_000003917.jpeg



./frames/friends_000001007.jpeg



./frames/friends_000001021.jpeg



./frames/friends_000002662.jpeg



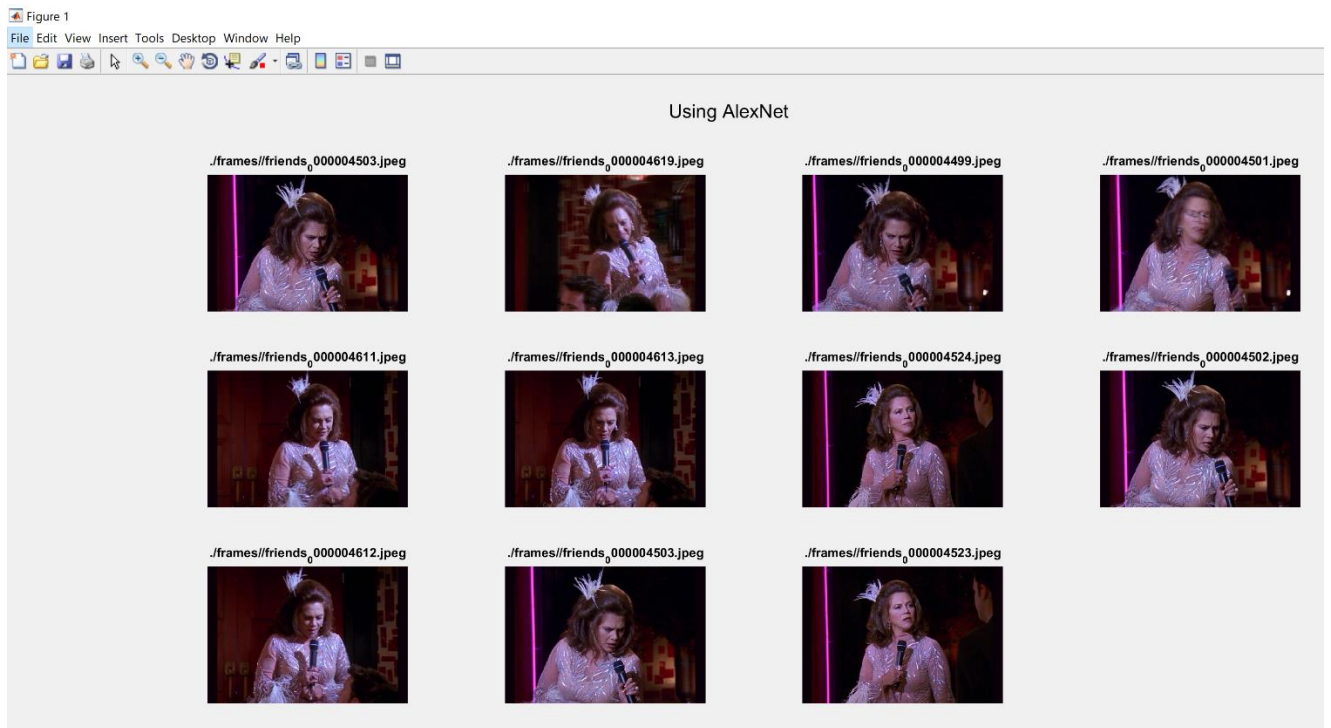
./frames/friends_000006214.jpeg



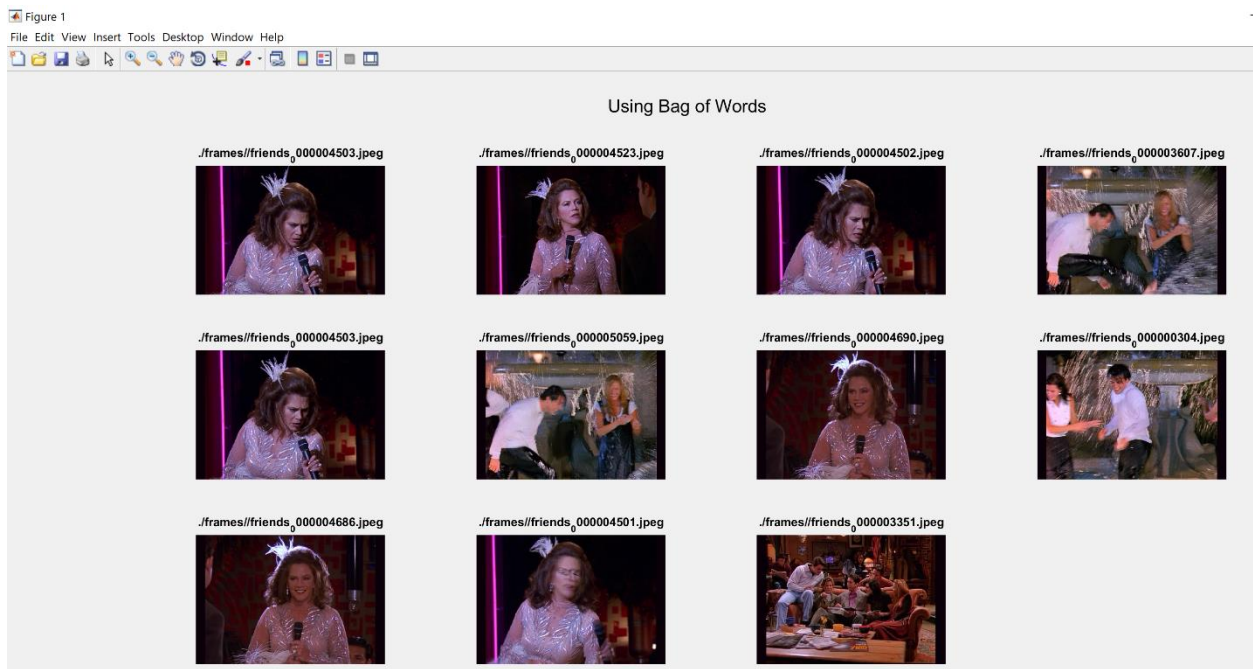
The descriptors of interest points inside the selected region are chosen for making visual word vocabulary, since the number of words are less, all the descriptors that are a little similar to the descriptors of visual words are mapped, and hence there is not much similarity between the images. If the region selected is very small and the number of interest points are very less, the descriptors are mapped to any visual word, even if it's not too similar to it. When a region with less texture content is selected the number of descriptors will be too low and hence there might not be similar pictures produced.

2.5 To run the code using AlexNet set $m = 2$ and to run the code using Bag of words set $m = 1$.

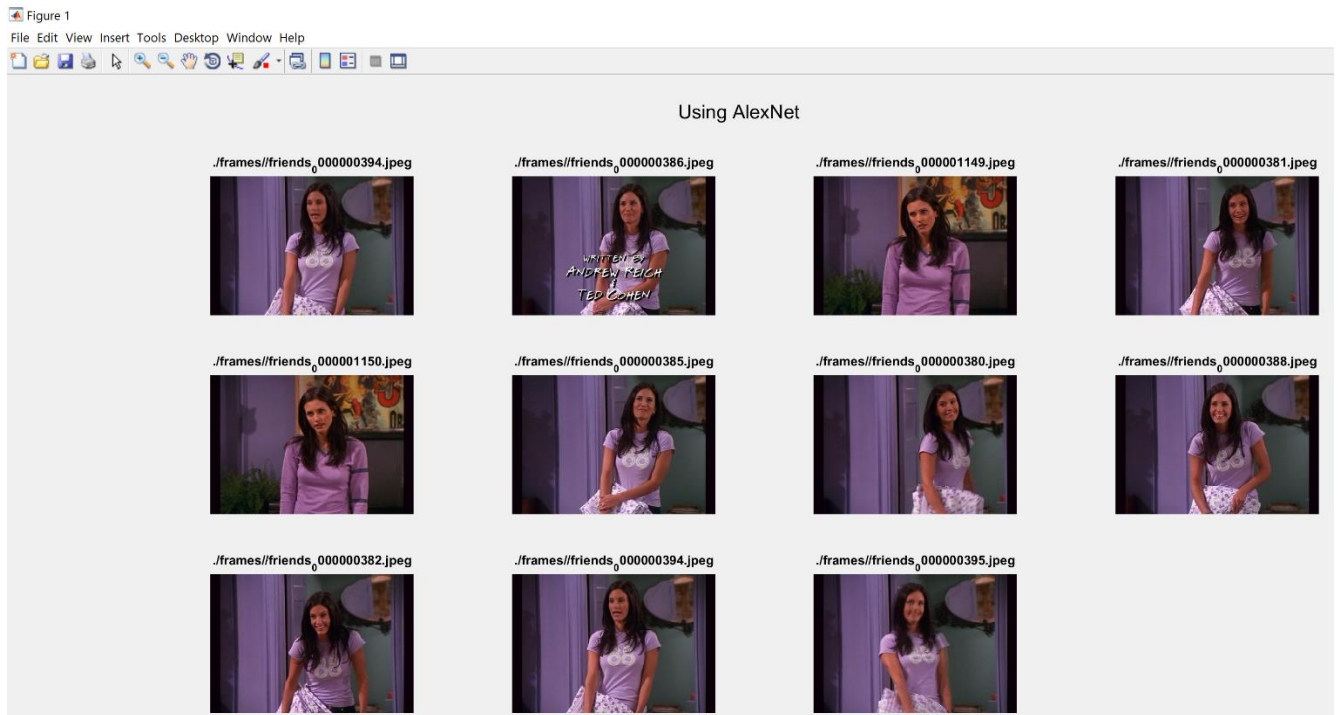
Result obtained with AlexNet for friends_0000004503.jpeg ($n = 4444$)



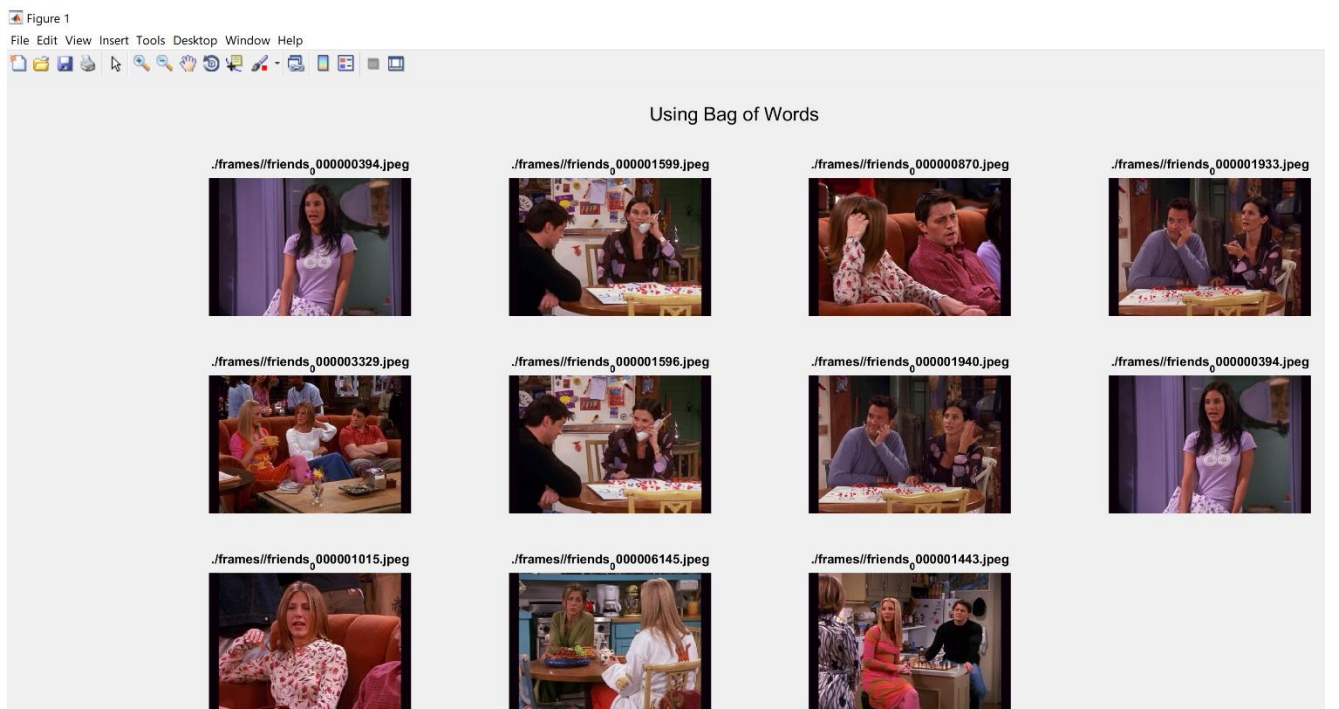
Result obtained with Bag of words for friends_0000004503.jpeg ($n = 4444$)



Result obtained with AlexNet for friends_0000000394.jpeg (n =335)



Result obtained with AlexNet for friends_0000000394.jpeg (n =335)



AlexNet outperforms the usual Bag of words method. It is faster and produces more accurate results than Bag of words. AlexNet is a Deep Convolution Neural Network. It involves multilayer processing. It

can recognize different faces while Bag of words only recognize a face in general. AlexNet contains a large dataset, hence contains more objects and is more specific.