Problem Set 3

Suhas - 915623705

June 3, 2018

Classmate collaborator(s):doing alone.

1 Short Answer Problems

1.1

Cases:-

a) Take any positions that are local maxima in Scale_Space

b) Take any positions whose filter response exceeds a threshold

properties:-

1) Repeatabililty

2) Distinctiveness

Case_a, Property_1 :-

This works really good because of we are considering local maxima across all the scales so whether the two images that we are comparing are one zoomed and other normal this repeatability property holds because we are considering local maxima across all scales as interest points.

Case_a, Property_2 :-

This works really good because of we are considering local maxima across all the scales so whether the two images that we are comparing are one zoomed and other normal this distinctiveness property holds because of sift descriptor we have histogram's across all directions to give us only if it a exact match between these two images.

Case_b, Property_1 :-

This works not so good because of we are considering certain threshold on only on one scale so whether the two images that we are comparing are one zoomed and other normal this repeatability property does not holds because we can't see matching positions having same threshold values(when very high) to be considered as interest points.

Case_b, Property_2 :-

This works not so good because of we are considering certain threshold on only on one scale so whether the two images that we are comparing are one zoomed and other normal this duplicity property holds but because of threshold values(when very high) first of all there are not even having any repetable interest points only. but wehn it is lower threshold there is a duplicity because there more number of interest points so duplicity chances are more.

so whether the threshold is more or less in this case duplicity property fails

1.2

Number of pixel gradients go in that particular direction within sub_patches.

2 Programming Video search with bag of visual words

2.1 Raw descriptor matching

The input image with selected region of interest in shown in fig 2.1.

The output image with nearest euclidean distance is shown in fig 2.2 .

The threshold used here is 0.3 to avoid the irrelevant matches even though they are minimum.

As we can see in the fig 2.2, most of the descriptors which are highlighted are the desired ones and are exactly matching with the descriptors which fallen in the selected region of the fig 2.1.

As we can even see there are some outlier's because even though the descriptors are nearby distance that doesn't mean it's a match so the solution for this is to fine tune it

by some threshold to remove those outlier's.

Figure 2.1: Input Image (showing selected region)

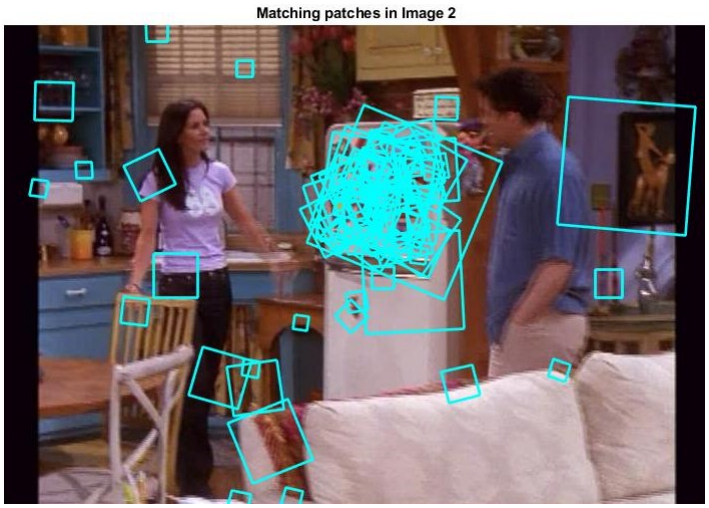

Figure 2.2: Output Image (showing nearest matching patches)

2.2 Visualizing the Vocabulary

As we can see in below two figs 2.3 and 2.4 we can clearly observe how well the Kmeans is classifying the patches based on sift parameters.

Irrespective of directions it is clearly classifying the patches across many groups. We can can observe that all the of patches that are belonging to same group are representing

similar gradients and shape even though some of them have different orientation.

Fig 2.3 is representing the similar cleavage across all the patches which are belonging to same word.

Fig 2.4 is representing the similar pattern across all the patches which are belonging to same word.

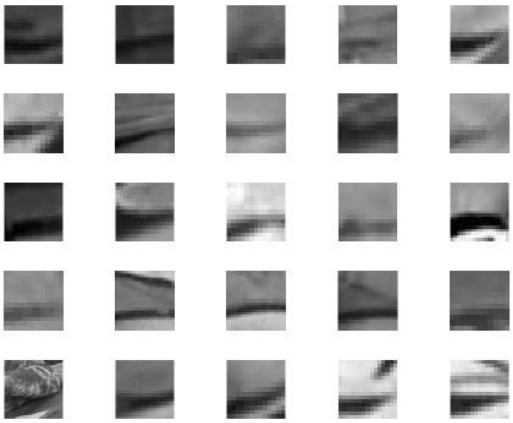

Figure 2.3: Patches belonging to same word.

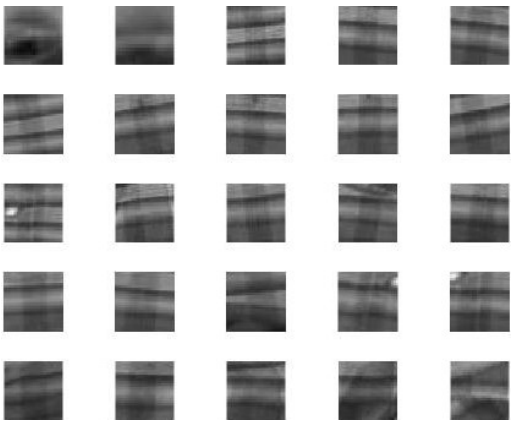

Figure 2.4: Patches belonging to same word.

2.3 Full Frame Queries

In all the below images 2.5 2.6 2.7 left most frame is query next 5 are the resulting retrieved nearest neighbour frames.

order:-
where left most is the nearer to query and similarity decreases from left to right

Because of querying by full frame we have more patches to check which means more accuracy, as expected if we observe in the below figures 2.5 , 2.6, 2.7 We can see that whether the frame near or not to the queried frame, it picks the best match by evaluating it's distance.

In figure 2.5 we can clearly observe that across all frames there is sight variation in his face direction remaining and all it is same so all the frames have similar patches to get nearest distances.

Figure 2.5: full frame matches

In figure 2.6 we can clearly observe that across all frames there is a same edges because of building windows so all the frames have similar patches to get nearest distances.

Figure 2.6: full frame matches

In figure 2.7 we can clearly observe that across all frames there is a same edges so all the frames have similar patches to get nearest distances.

Figure 2.7: full frame matches

2.4 Region queries

In all the below images 2.8, 2.9, 2.10, 2.11 left most frame is query next 5 are the resulting retrieved nearest neighbour frames.

The selected region in highlighted with yellow color line

order:-
where left most is the nearer to query and similarity decreases from left to right

Region quire's works well only when the the patches in the selected region are found in the searchable frame's too. So as suggested I selected the regions which are present in various frames.

Here I selected the stickers region on fridge 2.8 because it creates a lot of patches in that particular region because of more edges and because of this advantage as expected all the retrived frames have the same fridge.

Figure 2.8: Region Matches

Here I selected the Green Lamp which comes in many frames and as expected all the retrieved frames have same lamp in it 2.9.

Figure 2.9: Region Matches

Here I selected the fruits basket which results in retrieving only the frames which contains that fruits basket as shown in the fig 2.10

Here I selected the word 'friends' because I want to know how many times that particular word is appearing across all the frames in the same format. This clearly shows the accuracy of this model because as we can see in fig 2.11 all the frames whether they are

Figure 2.10: Region Matches

nearer are farther from quire's frame all 'friends' word consisting frames appeared in the result.

Figure 2.11: Region Matches

2.5 Full frame queries-2 : SIFT bag-of-words vs Deep Features

In all the below images 2.12, 2.13, 2.14, 2.15 left most frame is query next 5 are the resulting retrieved nearest neighbour frames.

order:-
where left most is the nearer to query and similarity decreases from left to right

DeepFC7 is really working good than the sift features as shown below two examples.

when comparing Bag of Words figure 2.12 and DeepFC7 figure 2.13 we can see that number of correct frames retrieved are 3/5 vs 4/5 respectively.

Figure 2.12: Bag of Words Retrieval

when comparing Bag of Words figure 2.14 and DeepFC7 figure 2.15 we can see that number of correct frames retrieved are 1/5 vs 5/5 respectively.

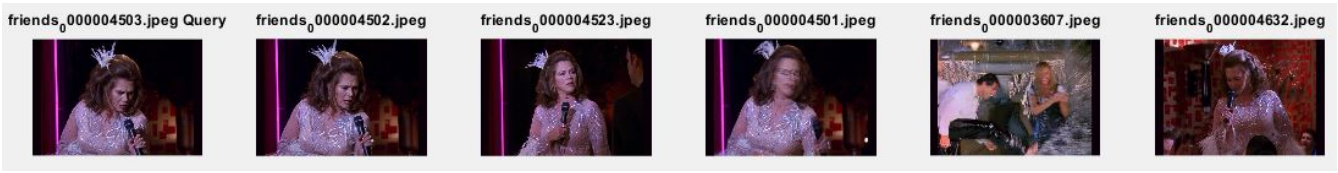

Figure 2.13: DeepFC7 Retrieval

Figure 2.14: Bag of Words Retrieval

Figure 2.15: DeepFC7 Retrieval