collaborators:

Kavitha Dee

Vandita Anand

Karishma Harry

SHORT ANSWER PROBLEMS

1. When performing interest point detection with the Laplacian of Gaussian, how would 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?

1. a) Detect many interest points in a local spatial region in this image. This would make each interest point less distinctive, since each of these interest points could produce good matches to many different interest points.

2. (b) The interest points is less repeatable, since the same image content may not be detected in different images since the threshold will be different for different images. This would also make the points less distinctive, for a similar region as in (a).

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

2.

Compared to Harris Corner detector which was not invariant to scale, to develop man interest operator that is invariant to scale and rotation is where SIFT came in. The essential idea is to address the problem of matching features with changing scale and rotation. The SIFT descriptor is very distinctive and but invariant to the variations. The key point has a particular orientation, scale and location assigned to it. Each value recorded in a 128 dimensional SIFT keypoint descriptor corresponds to one of 8 quantized gradient directions in one of the 4x4 spatial bin.

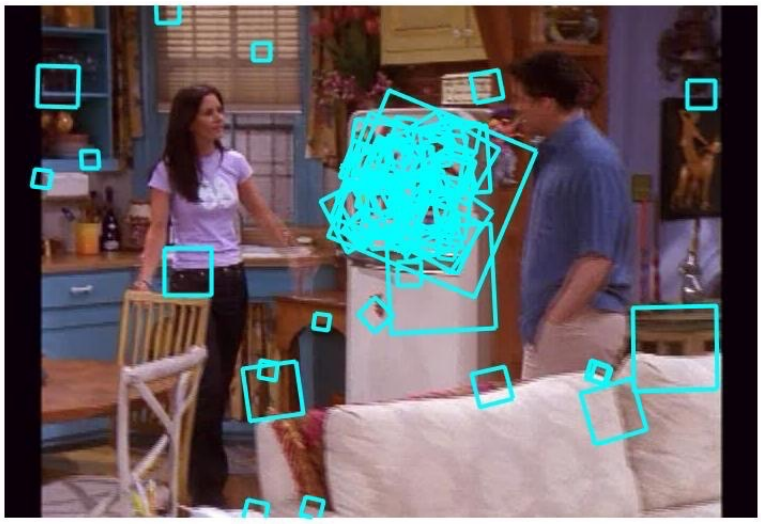
3. If using SIFT with the Generalized Hough Transform to perform recognition of an object instance, what is the dimensionality of the Hough parameter space? Explain your answer.

3. Hough Transform makes use of votes in a 4D hough Parameter Space. Voting lets features vote for all the models that are compatible with it. The algorithm cycles through all features and looks for models parameters that receive a lot of votes. For this situation, each element in the sift descriptor space should vote for keypoints. The key points would have the following parameters: x position, y position, scale and orientation. This is analogous to each point in image space voting for the centers in the center parameter space.

Raw Descriptor Matching

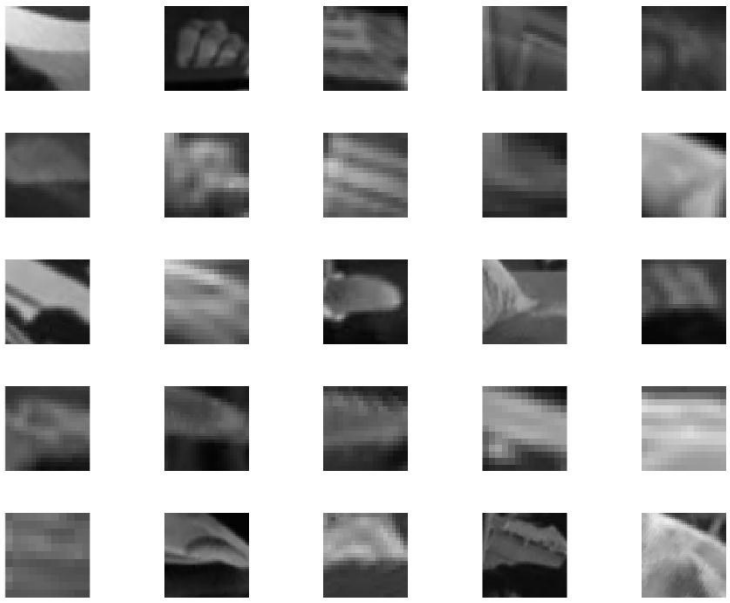
This function does the following -

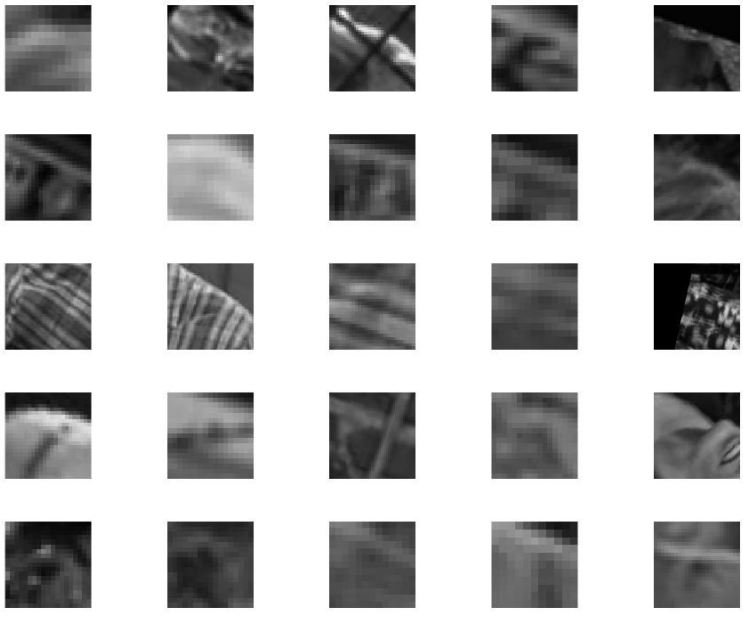
1) We use the selectRegion.m function to allow a user to select a region of interest.
2) Get the indices of the descriptors that lie in the region that is selected.
3) The distance of those descriptors from all descriptors in frame 2 is calculated.
4) This value is then thresholded by trail and error whichever shows the best result in this case its 75.



Build Visual Vocabulary

1) Sampled out 20 descriptors from each frame available.
2) Clustered all these descriptors into 1500 clusters
3) cluster centers = visual words.
4) To find 25 patches corresponding to a visual word: find sift patches belonging to that word and display the patches.





Full Frame Queries

1) Got a random frame, choose 3 of them, this is our query frame.
2) Get the sift descriptors and do the same step as the previous function as calculating the difference between the descriptor and the cluster center.
3) Made its bag-of-words histogram using a separate function of the words that match.
4) To compare the best matching frames from a given query frame, we can compare the distances between their histograms using normalized scalar product.
5) 5 most similar frames are displayed. There were some sometimes that the images were not matching and this is because

Reasons for failure:

Sometimes a scene only has a couple of frames.

It didn't not work for all random images. It worked for a good amount. If sample for kMeans was representative of the data set, then good results were obtained.






4. Region Queries - Explain the results, including possible reasons for the failure cases.

1) Instead of making a bag-of-words representation using the query image using all its descriptors.

2) We need to make it using the descriptors that lie in the region selected by the user.

3) The representations of all the other test frames remain the same. This is the only difference between this and the previous function.

There were a lot of failure cases for this one. If the query is very difficult, the best results were not obtained. When I chose something like a particular distinct patch of someone's dress, or perhaps a distinct object then some good queries ended up as results. The result below could be considered one of the vague cases where a patch is selected and that patch is searched for in the other images but the results are pretty very vague as the color appears on the other frames but it is not very obvious. Also when I selected a more contrasting pattern, the results were not that consistent.