Computer Vision
Assignment 4

Dhwanit Gupta
Ayush Tewari

1. Bag of Words model has been used to implement search in a dataset of 800 images with 11 categories. The various steps involved in implementing the BoW model are:
   a. Training
      i. Detecting keypoints and computing the corresponding features from all the images of testing set.
      ii. Clustering these features as words using k–means clustering algorithm.
      iii. Removing the top 5% and the bottom 5% of the words based on their frequency of occurrence. This is done because these words are not relevant to the search.
      iv. For every image in the testing set, computing its document vector, which is the normalized histogram of these computed words. Various matchers can be used here, like Flann or BruteForce.
      v. Building the inverse document vector i.e. for each word in the vocabulary, computing a vector which stores the number of times that word is present in every image.
   b. Testing
      i. Detecting keypoints and computing the corresponding features of the image.
      ii. Finding the corresponding word for every feature, its closest match. Compute its document vector.
      iii. From the inverted index, find all the documents containing all the words of the test image i.e. Find the insersection of corresponding inverse document vectors for these words. Alternatively, we can chose OR operation intstead of AND used here.
      iv. This set is the result set which matches the given image.
   c. Ranking
      i. From this set, we want to rank the image based on some similarity measure.
      ii. We have used cosine similarity i.e. Higher the normalized dot product of the two document vectors, higher is the rank. 8 top results are displayed.

   This has been tried using SIFT and SURF detectors and descriptors. The results are summarized below. The number of clusters used is 1000. As this number decreases, the results deteriorate because many features match to the same word resulting in noisy matches.

Descriptor used	Number of clusters	Max Number of Attempts	Average Accuracy
SIFT	100	10	66%
SIFT	1000	100	76%
SURF	100	10	63%
SURF	1000	100	74%

The precision is almost equal to the accuracy here as recall is ~1 because there are very few cases for which there is a false negative.

2. The second method used to search is by training a classifier for each class of images (11 in our dataset). The method used is as follows.

a. Training
i. Compute global feature for each image in the testing set.
ii. Train a classifier such that it can distinguish between different classes.

b. Testing
i. Compute the global feature of the test image.
ii. Using the classifier, find the corresponding class in which the image is present.
iii. Use cosine similarity to rank different images in the corresponding class.
iv. Show the top 8 matches as the result.

GIST ang PHoG have been used as the global features. The data has been trained using SVM and neural networks. As SVM is a binary classifier, for every class, a one vs rest classifier has been trained and a recursive approach has been used while testing. As neural network can train multi-class data , we have used only one classifier for classifying all the categories.

The results obtained for GIST has been summarized below.

Classifier	Average Accuracy
SVM	50%
ANN	15%

The results obtained for PHoG has been summarized below.

Classifier	Number Of Bins	Average Accuracy
SVM	25	30%
SVM	100	60%
ANN	25	15%
ANN	100	25%

The precision and acurracy are exactly equal here. Recall is equal to 1 as there is no false negative (the classifier will always declare it as belonging to one of the classes).

We can observe that the results using global features are not as good in comparison to the Bag of words model. This is mainly because when we take the whole image in consideration, we won't be able to match image which are cropped versions or have smaller parts in common with the original image. This does not happen in the case of Bag of words model.
Also, we can say that search using global features is less noisy i.e. No image from any other class will be displayed. But, in the Bag-of-words model, we have noisy matches too i.e. Match across different categories.
The results in case of PHog are better than GIST mainly because PHog encodes the Hog of the whole image as well as local information from different parts of the image upto a specific level. This makes it more robust that GIST in matching.