DU Jack
LAROYE Ambroise
SIA

Lab2: Image Classification

Introduction :

DU Jack
LAROYE Ambroise
SIA

1. Bag-of-Words Classification with Histograms of Oriented Gradients

1. Dataset

We will use the STL-10 dataset for training and testing our model. STL-10 contains 96 × 96 color images that belong to one of the following ten semantic classes : airplane, bird, car, cat, deer, dog, horse, monkey, ship, and truck. It contains a training set with 500 images from each class (5000 images in total) and a test set with 800 images from each class (8000 images in total). The training set is used to optimize the model parameters, whereas the test set is used to evaluate the model performance.

2. Bag of Words Classification Pipeline Overview

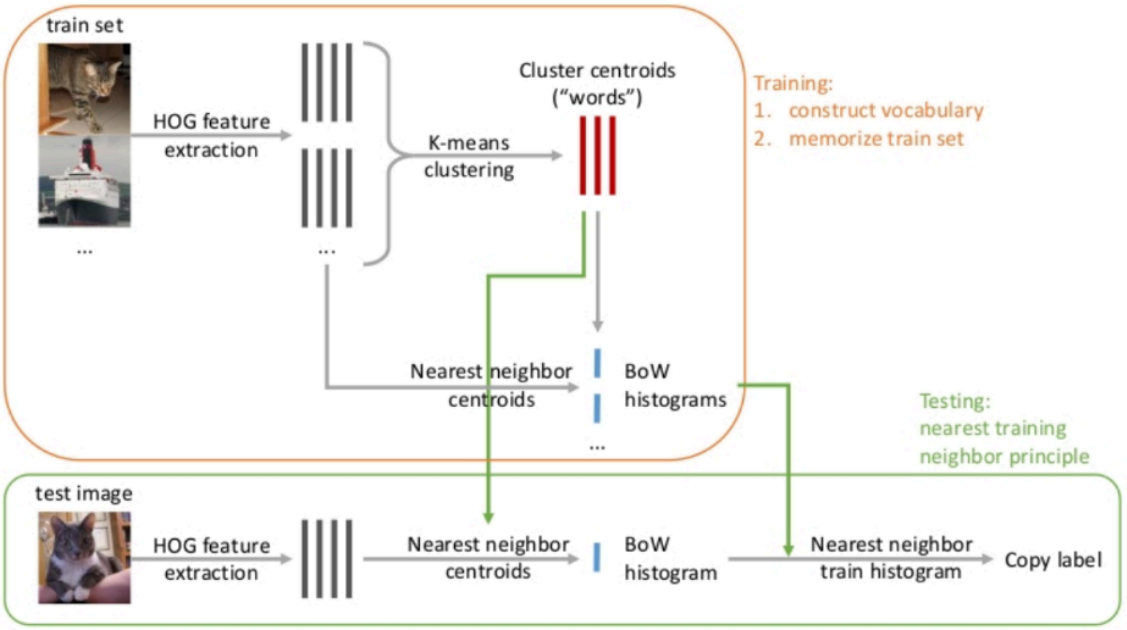


FIGURE 1 – Overview of BoW pipeline with HOG features and nearest neighbor classifier.

The algorithm we are going to implement is described in the diagram above.

Complexity :

Space Complexity:

- T is the total number of training images.
- K is the number of visual words (i.e., the BoW vocabulary size).
- Each Bag of Words (BoW) histogram has K bins, each representing a visual word.

For nearest-neighbor prediction, we need to store :

1. BoW histograms for all training images :
   Each histogram is a vector of length K, so storing the histograms for all T images requires T×K space.

2. Class labels for the training images :
   Each training image has a corresponding label. Storing these labels requires T space.

Thus, the overall space complexity is dominated by storing the histograms, leading to a total space complexity of O(T × K).
(The space required for labels is O(T), which is negligible compared to O(T × K) for the histograms.)

Time Complexity:

To predict the label of a test image, we follow these steps:

1. Compute the BoW histogram for the test image:

The time complexity for this step is not directly analyzed in your statement, but it depends on how the histogram is generated. Assuming it's precomputed or done in a different part of the process, we can focus on the prediction step.

2. Compute the distance between the test image's BoW histogram and each training image's BoW histogram:

- Each distance computation (e.g., Euclidean distance) between two vectors of length K requires O(K) time.
- Since there are T training images, computing the distance to all T histograms takes O(T×K) time.

3. Find the nearest neighbor:

- This involves selecting the minimum distance from the T computed distances, which requires O(T) operations.

Thus, the total time complexity for predicting the label of a single test image is dominated by the distance computation, which gives an overall time complexity of O(T × K).

Conclusion:

- Space complexity: O(T × K).
- Time complexity for prediction: O(T × K).

This is a standard complexity for a nearest-neighbor-based approach using BoW histograms.

DU Jack
LAROYE Ambroise
SIA

Feature Description with Histograms of Oriented Gradients (HOG)

Step 1: grid points

```
X indices: [  0 159 319 479 639   0 159 319 479 639   0 159 319 479 639
0 159 319 479 639]

Y indices: [  0   0   0   0   0 159 159 159 159 159 319 319 319 319 319
479 479 479 479 479]
```

Step 2: HOG function (look at the code)

Visual Vocabulary Construction with K-means Clustering

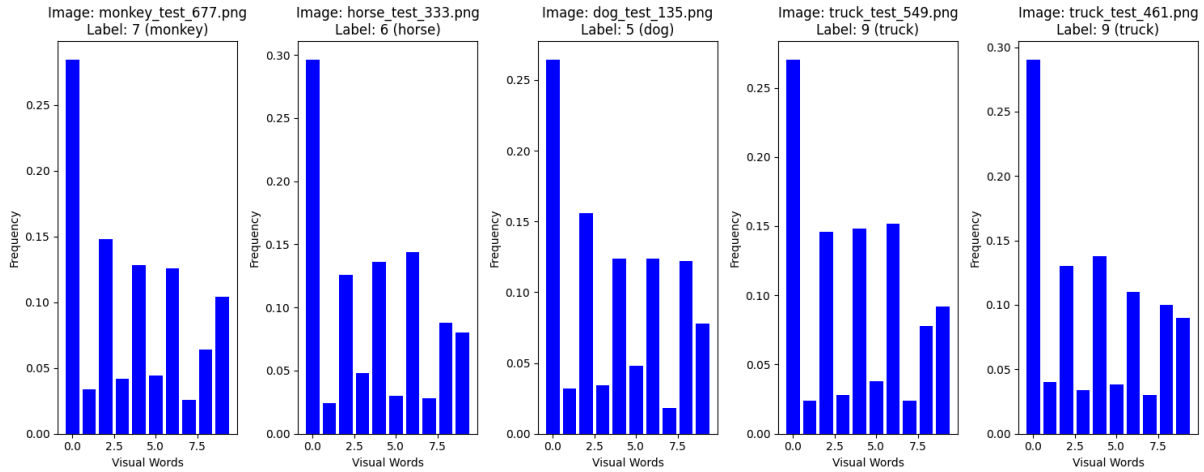
Step 1 : Random initialization of cluster centroids
Step 2 : Assign each data point to the cluster whose centroid is nearest to the point with respect to the Euclidean distance.

Dataset :

We prepare each class by identifying the picture with their corresponding label.

DU Jack
LAROYE Ambroise
SIA

Visualization of BOW histogram :



Test of our algorithm :

Training (For 2 class only)

Now we are training on 10 runs to generalize our results and verify their consistency. We are testing across multiple classes. We report our result in the following table :

Combination 1	Combination 2	Accuracy result (%)
Airplane	Car	66
Airplane	Bird	52
Deer	Truck	80
Airplane	Car	68
Airplane	Cat	52
Deer	Ship	81

How can we understand such difference between class :

- low quality images
- Some pictures do not have satisfying objects
- Similarities between class (airplanes - birds) (dogs - cats)

We then perform the same experiments as above where the L1 norm is used to measure distances between BoW histograms. And compare the obtained performance to the L2 norm baseline with different settings of K (50, 100, 200, 40).

The performance of both norms is very similar at each value of K. The mean accuracies for L1 and L2 norms are nearly identical, with minimal differences (usually 0.01 or less). This suggests that in this specific context, the L1 norm does not provide a significant improvement over the L2 norm, but it also does not degrade performance. Both distance measures appear to be equivalent for this task.

From the data provided, we can understand why the L1 and L2 norms are producing such similar results:

- BoW Histograms: The histograms for both the training and test images are extremely sparse (mostly zeros, with only one bin containing a 1). Such sparse histograms lead to very small differences in distances, causing the L1 and L2 norms to behave almost identically.
- L1 and L2 Distances: Both the L1 and L2 distance matrices are quite similar. This suggests that, in many cases, the differences between histograms are minimal and isolated to just one or two dimensions. As a result, both norms provide essentially the same relative distances between the images.

Impact of K :
Average performance slightly increases as K grows from 50 to 200, which is expected since more clusters allow for better capture of variations in HOG descriptors. However, at K=400, the accuracy does not significantly improve compared to K=200, suggesting that the benefits of increasing the number of clusters start to plateau. This may be because the model begins to overfit or the additional granularity is no longer as useful. Overall, the best accuracy is achieved at K=200, with an average of 0.85.