Computer Vision project

Michele Sanna

February 14, 2024

Abstract

In this work I'm going to implement the second problem proposed in the project assignment. The assignment requires to implement a bag-of-words approach to solve an image classification task. Every section of this report will cover the choices concerning the corresponding step in the assignment.

1 Building a visual vocabulary

For each image in the dataset I sampled 50 SIFT descriptors using the OpenCV detector, obtaining approximately 75k descriptors for the whole training dataset. To choose the number of centroids I tried to look at the curve of the average silhouette score, but it had a decreasing trend without any significant maxima. The number of centroids that it's used (26) is therefore chosen based on the best performance in terms of accuracy.

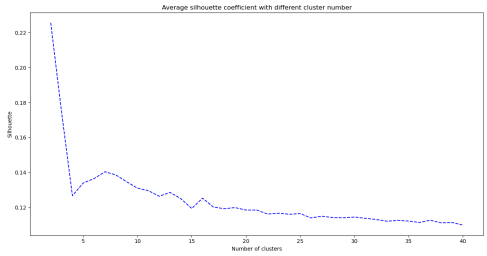

Figure 1: Average silhouette coefficient plotted

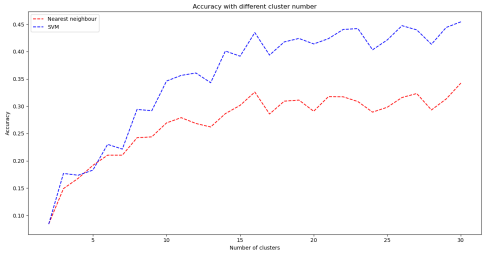

Figure 2: Accuracy plotted for different number of clusters

2 Representing each image as an histogram

To represent each image as an histogram an high number of SIFT descriptors are computed (up to 800). To increment the value of the bins corresponding to a visual word I tried two methods:

- In the first method, for each SIFT descriptor, the value of the bin corresponding to the closest centroid is increased by 1. Once that all the SIFT descriptors have "voted", the histogram vector is then normalized (every element is divided by the l2 norm of the vector).
- In the second method, for each SIFT descriptor, the value of each bin is increased by the reciprocal of the distance between the descriptor and the centroid corresponding to the bin. The histogram vector is then normalized (every element is divided by the l2 norm of the vector).

For both of the methods I used the euclidean distance, because the earth mover distance (or at least my implementation of it) was too slow to be applied. The second method has a slightly better performance.

2.1 Nearest Neighbour Classifier

The Nearest Neighbour Classifier, combined with the first method of normalized histogram construction, had an accuracy of 0.293, while with the second method had an accuracy of 0.301 (those values are an average on many runs).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
0	18.0	3.0	2.0	4.0	12.0	7.0	12.0	21.0	4.0	6.0	0.0	8.0	9.0	9.0	1.0	116.0
1	13.0	89.0	8.0	48.0	11.0	4.0	9.0	4.0	14.0	2.0	19.0	3.0	13.0	6.0	17.0	260.0
2	0.0	3.0	173.0	0.0	1.0	2.0	0.0	0.0	16.0	0.0	17.0	10.0	0.0	6.0	0.0	228.0
3	10.0	6.0	2.0	55.0	13.0	11.0	4.0	10.0	8.0	6.0	7.0	5.0	11.0	5.0	7.0	160.0
4	18.0	5.0	2.0	6.0	29.0	26.0	8.0	19.0	9.0	14.0	8.0	28.0	20.0	16.0	3.0	211.0
5	13.0	4.0	7.0	11.0	28.0	43.0	17.0	5.0	3.0	18.0	2.0	19.0	25.0	3.0	10.0	208.0
6	14.0	1.0	0.0	2.0	13.0	6.0	21.0	14.0	3.0	15.0	1.0	7.0	9.0	4.0	0.0	110.0
7	17.0	0.0	0.0	2.0	11.0	9.0	21.0	39.0	4.0	28.0	1.0	20.0	11.0	24.0	2.0	189.0
8	5.0	21.0	13.0	7.0	17.0	5.0	5.0	9.0	71.0	5.0	35.0	19.0	20.0	26.0	16.0	274.0
9	6.0	1.0	0.0	2.0	4.0	14.0	20.0	21.0	0.0	29.0	0.0	5.0	3.0	8.0	2.0	115.0
10	12.0	21.0	39.0	14.0	21.0	10.0	6.0	5.0	50.0	1.0	68.0	16.0	20.0	13.0	14.0	310.0
11	10.0	4.0	15.0	1.0	13.0	20.0	8.0	11.0	11.0	12.0	7.0	70.0	16.0	13.0	4.0	215.0
12	11.0	3.0	5.0	5.0	16.0	22.0	4.0	14.0	6.0	6.0	4.0	23.0	65.0	4.0	4.0	192.0
13	4.0	0.0	2.0	3.0	3.0	1.0	2.0	8.0	8.0	5.0	4.0	9.0	1.0	88.0	3.0	141.0
14	10.0	28.0	6.0	16.0	13.0	19.0	6.0	10.0	12.0	16.0	15.0	20.0	39.0	11.0	35.0	256.0

Table 1: Confusion matrix for the nearest neighbour classifier. Second method for construct histograms used. In this case the accuracy is 0.2991

2.2 One Vs Rest SVM Classifier

The One Vs Rest SVM Classifier, combined with the first method of normalized histogram construction, had an accuracy of 0.413, while with the second method had an accuracy of 0.442 (those values are an average on many runs).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Tot
0	6.0	2.0	0.0	1.0	3.0	5.0	19.0	18.0	6.0	19.0	5.0	11.0	12.0	9.0	0.0	116
1	2.0	105.0	8.0	33.0	4.0	2.0	0.0	3.0	21.0	1.0	46.0	4.0	18.0	4.0	9.0	260
2	0.0	0.0	203.0	0.0	0.0	0.0	0.0	1.0	7.0	0.0	6.0	7.0	1.0	2.0	1.0	228
3	0.0	5.0	0.0	76.0	3.0	12.0	3.0	4.0	4.0	3.0	10.0	2.0	21.0	12.0	5.0	160
4	4.0	4.0	1.0	2.0	7.0	27.0	14.0	20.0	7.0	12.0	12.0	37.0	48.0	15.0	1.0	211
5	1.0	0.0	0.0	6.0	1.0	110.0	7.0	6.0	2.0	15.0	2.0	25.0	22.0	4.0	7.0	208
6	0.0	0.0	0.0	1.0	0.0	11.0	29.0	16.0	1.0	27.0	1.0	7.0	12.0	5.0	0.0	110
7	1.0	0.0	2.0	0.0	2.0	7.0	24.0	33.0	2.0	35.0	1.0	22.0	27.0	33.0	0.0	189
8	2.0	3.0	16.0	7.0	2.0	0.0	1.0	2.0	154.0	2.0	46.0	5.0	13.0	21.0	0.0	274
9	0.0	1.0	0.0	0.0	1.0	2.0	12.0	9.0	0.0	75.0	0.0	4.0	6.0	5.0	0.0	115
10	0.0	12.0	50.0	9.0	5.0	2.0	1.0	0.0	51.0	0.0	133.0	5.0	14.0	20.0	8.0	310
11	0.0	0.0	5.0	0.0	1.0	8.0	7.0	12.0	3.0	11.0	0.0	135.0	21.0	10.0	2.0	215
12	0.0	0.0	0.0	0.0	1.0	14.0	6.0	9.0	10.0	5.0	4.0	21.0	114.0	3.0	5.0	192
13	0.0	0.0	0.0	1.0	0.0	0.0	2.0	2.0	4.0	4.0	0.0	3.0	2.0	122.0	1.0	141
14	1.0	19.0	3.0	12.0	3.0	33.0	5.0	3.0	6.0	17.0	21.0	25.0	34.0	7.0	67.0	256

Table 2: Confusion matrix for the one vs rest SVM. Second method for construct histograms used. In this case the accuracy is 0.4586