Computer Vision Homework 02

Qiu Yihang

Nov.27-Dec.3, 2022

1 Programming Assignment: Bag-of-Features

1.1 Main Idea

The main idea for bag-of-features is as follows.

- Extract **SIFT** features from each image. Collect features for each category.
- Divide features into $N\_Clusters$ clusters by **k-Means**, i.e. learn the "visual vocabulary".
- Quantize features in each image by the learned visual vocabulary.
- Use the histogram of visual words to represent the image.

Furthermore, since we need to classify images in **Caltech-101** Dataset, we need to further use **SVM** on the representation of images, i.e. the histogram for each image extracted by **bag-of-features**, to classify the images.

1.2 Some Details During Programming

To avoid the side effects brought by an unbalanced dataset, before k-means clustering, we calculate **the minimum number of features in each category**, and force each category to have the same number of features, i.e. the minimum number.

Specifically, since there are 102 categories, we cluster image SIFT features into 102 clusters, i.e. we choose $\boldsymbol{N\_Clusters = 102}$. In fact, different number of clusters is acceptable as long as it is not too small, since SVM is implemented on the histogram feature after bag-of-features.

1.3 Results

We also experiment on different kernel functions of SVM.

The results are shown in the following table.

Kernel	Accuracy
Linear	0.15051020408163265
Poly	0.36698250728862974
RBF	**0.37572886297376096**
Sigmoid	0.21829446064139943

It is plain to see that

- The linear kernel function seems not suitable for this problem. Thus, the data is likely to be not linearly separable.
- The polynomial kernel function performs well but not as good as RBF. Meanwhile, the number of its parameter is more than RBF.
- The **RBF kernel** performs the best, and we choose it as the kernel function for SVM.

The best accuracy of our method is 0.37572886297376096.