Report: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories

Daquan Lin 85610653
ShanghaiTech University
Shanghai, China
lindq@shanghaitech.edu.cn

Abstract

This report introduces how to implement the algorithm for Recognizing Natural Scene Categories by Spatial Pyramid Matching show in this paper[2]. I test algorithm on the publicly available Caltech-256[1].

1. Introduction

Although the Bags of Features has great performance, it is weak in spatial information. People can construct some strange images that consist of some of features without caring about the spatial structure, which can also be classified as a targeted label with high probability. In this paper, the author proposed a new method that contain the spatial information, which called Spatial Pyramid Matching(SPM).

2. Dataset

I use a batch of data from Caltech-256, choose 20 classes and each class pick up 50 images. Therefore, I get 1000 images. Then, I split it into training set and test set with ratio of 4/1.

3. SIFT and Build Codebook

For each RBG image, get gray image firstly and calculate the key points for $stepsize$: 4. Then, calculate the descriptors by key points. For all images, I got $X_train_feature$.

Set $K = 100$, use K-means to build codebook with $X_train_feature$.

4. Build Spatial Pyramid

Construct a sequence of grids at resolutions $l = 0, 1, 2$, such that the grid at level l has 2^l cells along each dimension, for a total of $D = 2^{dl}$ cells.

5. Spatial Pyramid Matching

For each level, concatenate the pyramid, and calculate the histogram with a weight.

$$\begin{aligned}\mathcal{K}^L(X,Y) &= \mathcal{I}^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}}(\mathcal{I}^l - \mathcal{I}^{l+1}) \\ &= \frac{1}{2^L}\mathcal{I}^0 + \sum_{l=1}^{L} \frac{1}{2^{L-l+1}}\mathcal{I}^l\end{aligned} \tag{1}$$

6. Result

I use SVM to train the data, When $M = 100, level = 1$, use grid search for C and γ by 5-fold Cross-validation. When $C = 100$ and $\gamma = 0.001$, I got the best model that accuracy is 0.431 ± 0.044 in test dataset.

When $M = 100, level = 2$, use grid search for C and γ by 5-fold Cross-validation. When $C = 10$ and $\gamma = 0.001$, I got the best model that accuracy is 0.398 ± 0.057 in test dataset.

Because it is very slow and cost memory heavily, so I just try a little.

7. Conclusion

In this report, I implemented the algorithm for Recognizing Natural Scene Categories by Spatial Pyramid Matching.

References

[1] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Computer vision and pattern recognition, 2006 IEEE computer society conference on, volume 2, pages 2169–2178. IEEE, 2006.