

# Algorithm and Data Structure Coursework: K-Means Feature for Image Retrieval

Qiwei Feng  
2011011250, IIS-10  
Tsinghua University  
gdfqw93@163.com

Pufan He  
2011011307, IIS-10  
Tsinghua University  
hpfdf@126.com

## ABSTRACT

This project implements a similar image search algorithm (image retrieval) based on multiclass classification and K-Means feature. Our training phase includes image resizing, image patch extraction, patch sampling, PCA whitening, K-Means for patches, feature extraction and multiclass SVM. We use 218 dimension K-Means and RGB, HSV color moment. The training phase takes no greater than one hour in time, 8GB in memory. Finally we obtained 69.49% accuracy on test data classification.

We have made our work open, and the full project codes can be found at <https://github.com/caiwaifung/lastcourse>.

## Keywords

Image Retrieval, Image Classification, SVM, Whitening, K-Means

## 1. INTRODUCTION

## 2. IMPLEMENTATION

### 2.1 Patch Extracting and Sampling

#### 2.1.1 Whitening

### 2.2 K-Means Clustering

### 2.3 Feature Extracting

### 2.4 Multiclass SVM

## 3. EXPERIMENTS

### 3.1 Data Set

### 3.2 Without Whitening

### 3.3 With Whitening

### 3.4 Final Test

Please see `a.html` under `result.zip` for a more detailed demo.

## 4. CONCLUSION

## 5. REFERENCES

- [1] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] A. Coates and A. Y. Ng. Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade*, pages 561–580. Springer, 2012.
- [3] A. Coates, A. Y. Ng, and H. Lee. An analysis of single-layer networks in unsupervised feature learning. In *International Conference on Artificial Intelligence and Statistics*, pages 215–223, 2011.
- [4] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [5] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.
- [6] J. A. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.
- [7] A. Vailaya, M. A. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for content-based indexing. *Image Processing, IEEE Transactions on*, 10(1):117–130, 2001.