

# Image Scene Classification Based on Fisher Discriminative Analysis and Sparse Coding

Jianliang Meng<sup>1, a</sup>, Rui Ni<sup>1, b, \*</sup>, Ye Wang<sup>1, c</sup>, Peng Zhao<sup>2, d</sup>

<sup>1</sup>North China Electric Power University, China

<sup>2</sup>State Grid Hebei Electric Power Company

<sup>a</sup>mjl1954@126.com, <sup>b</sup>renee1110@163.com, <sup>c</sup>yeahwangy@hotmail.com, <sup>d</sup>zpnew@sohu.com

**Keywords:** Scene classification, image representation, sparse coding.

**Abstract.** Bag of visual words (BoVW) is currently widely utilized as an image representation model. However, conventional BoVW methods usually bring drawbacks such as large representation errors, lack of spatial information and weak discrimination. In order to overcome those, this paper proposes a new image scene classification algorithm based on fisher discriminative analysis and sparse coding. Firstly, we construct the non-negative sparse locally linear coding to encode the local features with neighbors' visual vocabularies, thus to make effective use of images' spatial information. Secondly, we add fisher discriminative analysis to construct a non-negative sparse locally linear coding model based on fisher discriminative criterion constraint, thus to obtain the images' discriminative sparse representation, and promote the spatial separability of sparse coefficients and enforce the classification capability of images' sparse representation. Finally, we combine SVM classifier to perform scene classification. Experiment results show that our algorithm efficiently utilizes spatial information of images and incline to seek images' discrimination representations, and improves the classification performance, thus it's more suitable for image classification.

## 1. Introduction

Image scene classification is an important research in computer vision. One of the most important problems to be solved in scene classification is the representation of the image. In recent years, the image scene classification method based on BoVW has been widely concerned by scholars. Li proposed a Bayesian hierarchical model (model Bayesian hierarchical, BHM) for scene classification. Then Li uses BHM for target classification, and achieved good classification results. But BHM does not take the regional and spatial distribution of the image into account, and these local spatial information is important for the effective representation of the scene. In general, these common methods need to use the K-Means algorithm to cluster the local features of the image. Although the K-Means algorithm is simple, it is easy to generate large quantization error, which leads to the decrease of the classification performance. In addition, the BoVW image visual vocabulary distribution histogram representation is a linear feature, which can obtain good classification results, while the nonlinear classifier has high computational complexity and poor generalization ability.

In order to solve the problem, the Yang uses sparse coding method (SC) to quantify the local features of the image, which is based on the sparse coding based spatial pyramid matching algorithm (SCSPM) for scene classification. However, the SC model is the generative model which is based on the minimization of the signal reconstruction error. They are the minimal signal reconstruction error, and not suitable for solving the problem of image scene classification. Therefore, it is more important to find the discrimination of the image in the image scene classification task. Based on this, this paper introduces the SC model to improve the discriminant of image sparse representation.

## 2. Method

### 2.1 Sparse Coding Image Representation Based on Fisher Discriminant.

Natural images have been proved to have sparse structure, so the use of SC to represent the image is reasonable. The SC image representation is generally divided into 2 steps: dictionary learning and

sparse decomposition. This paper uses the NSLLC model based on Fisher to learn the dictionary, and on the basis of the VW library and the generated image sparse representation vector.

NSLLC model based on Fisher discriminant constraint.

In this paper, we introduce the Fisher criterion into the target function of the dictionary learning, and combine the NSLLC algorithm, and propose a NSLLC model based on Fisher discriminant constraint, and its form is

$$\min J(B, X) = \|Y - BX\|^2 + Y_1 \cdot \ln F(X) + Y_2 \sum_{i=1}^N \|d_i \otimes x_i\|^2 \quad (1)$$

Here,  $Y$  represents the input feature matrix;  $B$  represents over complete dictionary;  $X$  to  $y$  with respect to the sparse coding matrix  $B$ ;  $N$  represents the input feature vector of a number;  $Y_1$  and  $Y_2$  are adjustment factors (positive constant). In formula (1), the first item is signal fidelity, which guarantees that the signal energy is not lost in the case of coding; The second item is the discriminant term, which is used to enhance the discrimination of sparse representation coefficients  $X$ , which makes the classification of  $X$  better; The third term is the penalty term,  $d_i \otimes x_i$  is a new vector obtained by the vector  $d_i$  and  $x_i$ ,  $d_i$  is shown, the penalty term requires that the non zero coefficient of the encoding  $x_i$  is always more than the  $y_i$ , which reflects the local.

Build VW Library.

The building process of NSLLC Library Based on Fisher discriminant constraint VW model is as follows:

Step1. Randomly selects a number of images from the training image, and extracts the SIFT features, forming the SIFT feature vector set.  $F = \{f_1, f_2, \dots, f_i, \dots, f_N\}$   $f_i \in \mathbb{R}^{128}$ , and  $N$  is the number of SIFT feature vectors.

Step2. Uses the iterative algorithm in section 2.1.1 to solve the VM library of the SIFT feature Vector set  $F = \{f_1, f_2, \dots, f_i, \dots, f_N\}$  which  $V \in \mathbb{R}^{128 \times K}$ .

Image sparse representation vector's generation.

After the completion of the construction of VW library  $V$ , an image  $I$  can be represented by sparse vector. The concrete steps of image sparse representation vector generation are given below:

INPUT. Image  $I$ , VW library  $V$ .

OUTPUT. Sparse representation of images  $p$

Step1. Extract input image  $I$  SIFT features, constitute the SIFT feature vector set  $F = \{f_1, f_2, \dots, f_i, \dots, f_M\}$ ,  $M$  is the total number of SIFT feature vectors in the input image.

Step2. According to VW library  $V$ , the standard gradient optimization algorithm is used to solve the model, get each  $\alpha_i \in \mathbb{R}^k$  in  $f_i \in \mathbb{R}^{128}$ , which  $f_i \approx \alpha_i$ .

Step3. Combine by fusing all the sparse feature vectors  $\alpha = \{\alpha_i\}$  ( $i = 1, 2, \dots, M$ ) and get the sparse vector representation of the image  $p \in \mathbb{R}^k$

Taking into account the main idea of Fisher discriminant analysis is the mean distance between the largest categories, this paper uses the average value of the fusion method to fuse all the sparse feature vector  $\alpha$ .

## 2.2 Image Scene Classification Algorithm Based on Fisher Discriminant Sparse Coding.

In the training phase, the SIFT feature vector set  $F = \{f_1, f_2, \dots, f_i, \dots, f_N\}$  is extracted from the training image and the SIFT feature is extracted randomly, which  $f_i \in \mathbb{R}^{128}$  and  $N$  is the total number of SIFT feature vectors in the input image. As a training vector set, the VW library  $V$  ( $V \in \mathbb{R}^{128 \times K}$ ) is constructed by using the algorithm in section 2.1.2, and the sparse representation coefficients of the training images are generated by  $X$ . Finally, the sparse representation coefficient  $X$  of the training images is used to train the SVM classifier.

In the test phase, the SIFT feature of the test image is extracted, and the SIFT feature vector set  $F = \{f_1, f_2, \dots, f_i, \dots, f_M\}$  is obtained,  $M$  is the total number of SIFT feature vectors in the input image. Then according to the VW library  $V$ , the  $F$  is based on the Fisher discriminant constraint NSLLC, the sparse feature vector set  $\{\alpha_i\}$  ( $i = 1, 2, \dots, M$ ) is obtained. Then, the mean value of all the sparse feature vector  $\{\alpha_i\}$  is fused, the sparse representation vector  $p$  is obtained. Finally, the sparse representation of the test image is trained to recognize and get the result of the SVM.

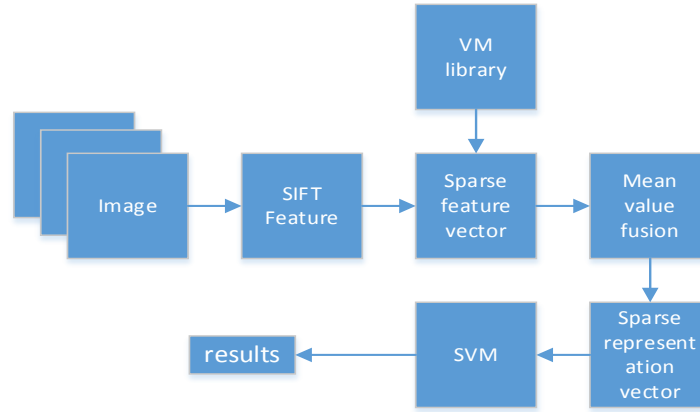


Fig. 1. Fisher based discriminant sparse coding algorithm flow chart

### 3. Results Analysis

In this paper, 2 indexes are used to evaluate the performance of scene classification, the accuracy rate and the average accuracy. In order to verify the effectiveness of the proposed algorithm, it is compared with the traditional SPM algorithm, SCSPM algorithm, LLC algorithm, NSLLC algorithm, pLSA algorithm, SPM-pLSA algorithm and BHM algorithm. The traditional SPM algorithm and the SCSPM algorithm use the Pyramid structure to extract the SIFT features of the image sub block, and the total number of Pyramid is located in the 3 layer; LLC and NSLLC algorithm directly use the LLC algorithm to provide 1024 VW as BoVW, and the nearest neighbor parameters of the 2 algorithms are  $k = 5$ ; The main questions of the meaning of the word pLSA and BHM are 40; The spatial pyramid SPM-pLSA algorithm design into three layers, respectively  $1 \times 1$ ,  $2 \times 2$  and  $4 \times 4$ , the PLSA model topic number is designed into 40. In this paper's algorithm, the convergence thresholds are designed to be  $\varepsilon = 0.001$ . 8 kinds of algorithms of the vocabulary size was set for 1024. Table 1 shows the 8 kinds of algorithms' average accuracy results on three image databases.

Table 1 Eight kinds of algorithms' average accuracy results on three image databases.

	OT	FP	LSP
SPM	63.8	61.4	57.7
SCSPM	64.4	62.3	60.2
LLC	67.2	63.1	62.5
NSLLC	73.1	71.1	71.3
PLSA	68.4	67.6	65.2
SPM-pLSA	75.4	73.5	72.8
BHM	75.2	74.0	73.5
THIS	77.0	75.3	76.2

From table 1, we can get that the average accuracy of the traditional SPM algorithm is the lowest. This is because the image histogram of the algorithm is obtained by direct statistics. It is a linear feature, which needs to be combined with the nonlinear classifier to achieve better classification results. The average accuracy of SCSPM algorithm is higher than the traditional SPM algorithm, which is due to the SCSPM algorithm for image sub block SC, with sparse vector representation of image features, thus improving the image feature representation of classification ability. And the LLC algorithm has the higher accuracy than the SCSPM method because of the locality of feature extraction; Compared with the LLC algorithm, the NSLLC algorithm has better performance than the VW algorithm, so it can improve the classification performance; Compared with the NSLLC algorithm, the proposed algorithm on the basis of it adds the Fisher criterion, which directly enhances the discriminant of the image features, so that the classification performance can be improved; The classification performance of SPM-pLSA algorithm compared with pLSA is improved, which is helpful to improve the representation of scene content. However, the absolute spatial information is used to classify the scene in Pyramid, which can be divided into different semantic categories in

mistake. The classification performance of SPM-pLSA algorithm is limited. And compared with the SPM-pLSA algorithm and BHM algorithm, the proposed algorithm considers the local neighborhood of feature points, and adds the Fisher discrimination strategy, so the classification performance is improved.

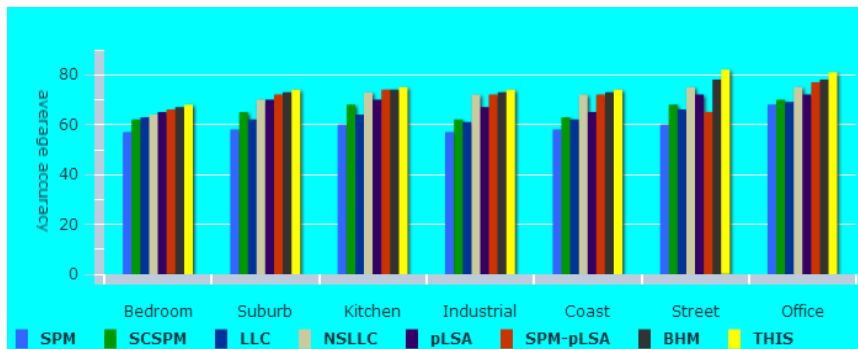


Fig. 2 Comparison of 8 kinds of algorithms with 15 kinds of natural scenes in LSP Library

From Figure 2, we can see that the classification accuracy of Street and Office is improved obviously, which is because the spatial distribution of these 2 scenes is large, and the spatial information of feature points can be increased by this algorithm. And the local spatial information of Bedroom and kitchen scene is very similar, it is difficult to distinguish them from the spatial information, so it is very difficult to improve the classification performance of the scene.

#### 4. Conclusion

In this paper, we construct the NSLLC model based on Fisher discriminant analysis, use Fisher discriminant analysis method is used to maximize the inter class distance in order to increase the discrimination between different classes, and to reduce the difference between the two classes. Therefore, the classification of image sparse representation is improved, and the performance of scene classification is improved. The experiment results show that the proposed algorithm is effective for two dimensional image scene classification. In addition, for the 3D image scene, the proposed algorithm can be applied to the 3D model after projection, so the algorithm has good scalability. However, the computational complexity of this algorithm is relatively high, how to speed up the calculation speed is an important research direction

#### References

- [1]. Huang Y Z, Wu Z F, Wang L, et al. Feature coding in image classification: a comprehensive study[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 493-506
- [2]. Ergul E, Arica N. Scene classification using spatial pyramid of latent topics[C]//Proceedings of the 20th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2010: 3603-3606
- [3]. Zhao Q, Horace HSlp. Unsupervised approximate-semantic vocabulary learning for human action and video classification[J]. Pattern Recognition Letters. 2013, 34(15): 1870-1878
- [4]. Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2009:1794-1801
- [5]. Guo Wenjing, Zhang Yanqiu, Liu Yongjin. Image category learning based on features words and shape model[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(10): 1467-1475