International Journal of Hybrid Information Technology
Vol. 10, No. 8 (2017), pp.121-136
http//dx.doi.org/10.14257/ijhit.2017.10.8.10

Unsupervised Deep Discriminative Feature Learning for Aerial Image Classification

Tao Shi, Chunlei Zhang*, Hongge Ren and Fujin Li

College of Electrical Engineering, North China University of Science and Technology, Tangshan 063009, China
randyray@126.com

Abstract

Nonnegative sparse coding (NSC) is widely utilized as an image representation model. However, conventional NSC methods usually cause large representation errors, lack of spatial information and weak discriminative. In order to overcome these drawbacks, this paper proposes an unsupervised deep discriminative feature learning framework for aerial image classification which is based on Fisher Discriminative Nonnegative Sparse Coding (FDNSC) and Deep Belief Network (DBN). First, image features are extracted by using scale-invariant feature transform (SIFT). Then fisher discriminative analysis is added to construct a NSC with fisher discriminative criterion constraint, thus to obtain the discriminative sparse representation of images. Finally, DBN is combined to perform aerial image classification. The proposed method is applied to OT data set and UC Merced data set. Experimental results show that the proposed method efficiently utilizes spatial information of images and can promote the spatial separability of sparse coefficients, thus improves the classification performance and is more suitable for aerial image classification.

Keywords: *Aerial Image Classification; Nonnegative Sparse Coding; Fisher Discriminative Criterion; Deep Belief Network*

1. Introduction

Aerial images have been widely used in city planning, coastal surveillance, military tasks, etc. It is important to understand the contents of the aerial image and study the scene types. However, aerial images contain many objects of different sizes and have the characteristics of large range, wide viewing angle, high resolution and large data, which bring great challenges to study.

One of the important problems to be solved in the aerial image classification is the representation of the image. At present, the most widely used method of image representation is bag of visual words (BoVW) [1]. First extract the local features of the image such as the scale invariant feature transform (SIFT) feature and then quantize the local feature into visual words. Finally represent the number of visual words in the form of histogram. Subsequently, Lazebnik et al. proposed a spatial pyramid matching (SPM) method for image scene classification and recognition [2]. First divide the image into different levels in space and then calculate the similarity between each layer of Pyramid. Finally get the total image similarity by weighted sum of each similarity. Yang and Newsam have computed the co-occurrence of the visual words and combined this with the bag-of-visual-words (BoVW) method, and they reported higher classification accuracy than the traditional BoVW and the spatial pyramid matching kernel (SPMK) for their extended spatial co-occurrence kernel (SPCK++) method [3]. Combining probabilistic latent semantic analysis (pLSA) and SPM, Ergul et al. proposed SPM-pLSA image scene classification algorithm and achieved good results [4]. However, these methods are easy

ISSN: 1738-9968 IJHIT

to produce large quantization errors in the construction of visual words, which leads to the decrease of the classification performance.

With the development of sparse coding (SC) [5], the image representation method has changed greatly. Sparse coding simulates the activity of the neuron's sparse type. Use a set of basis functions to obtain the input image encoding where only a small amount of the coefficient is large, while the others are small or close to zero. Yang et al. quantized local image feature by using SC method [6] and proposed sparse coding based on spatial pyramid matching (SCSPM) for image classification. Zhang et al. extended Laplasse sparse encoding and proposed kernel Laplasse sparse encoding, which reduced the feature quantization error and enhanced the performance of sparse encoding [7]. However, all the SC models are based on the objective of minimizing the signal reconstruction error and ignore the category attribute and discriminate representation of the image, which is not conducive to image classification.

In recent years, deep learning has been widely used in various fields of machine vision as a new method. Deep learning network has a hierarchical structure, which can effectively learn the features from a large number of input data. One of the representative deep networks is deep belief network (DBN) [8]. Qi Lu et al. proposed a remote sensing image classification method based on DBN model which can outperform support vector machine (SVM) and traditional neural network (NN) [9]. This paper takes DBN as an important tool for aerial image classification.

Aiming at the problem that the aerial image feature is difficult to extract and nonnegative sparse coding (NSC) methods usually cause large representation errors, lack of spatial information and weak discriminative, this paper proposes an unsupervised discriminative feature learning algorithm. On the basis of NSC, this algorithm adds the Fisher Discriminative criterion, which makes the same type of image sparse representation coefficient distance closer and the sparse representation of different types of images distance farther. Thus, it enhances the discrimination of the extracted features. In order to enhance the classification ability of aerial image, this paper also proposed an aerial image classification method based on Fisher Discriminative Nonnegative Sparse Coding (FDNSC) and DBN. The method is applied to OT data set and UC Merced data set and is compared with other methods. Experimental results show that our method can promote the spatial separability of sparse coefficients and improves the classification performance.

2. Relevant Theory

2.1 Nonnegative Sparse Coding

Barlow’s effective encoding hypothesis proposed that the response of the optic nerve cells to the external environment in the primary visual cortex is consistent with the characteristics of sparse encoding. Sparse coding is to make a few nerve cells activate on the basis of a complete representation of the input stimulus pattern. In 1996, Olshausen et al. used a linear superposition method to further study the relationship between the nature of simple cell receptive fields and sparse encoding and proposed a sparse encoding model [10]. He used a linear superposition of basis functions to represent input image. Under the condition of minimum mean square error, the result of linear superposition is similar to the original image as much as possible. Meanwhile, he used sparse penalty function to make the representation of the feature as sparse as possible. This paper use I and S to represent the original image and sparse vector. The linear superposition model and optimization criteria are expressed as follows:

$$I = AS \tag{1}$$

$$\min_{A,S} E(A,S) = \sum_{x,y}[I(x,y) - \sum_{i}\alpha_i\phi_i(x,y)]^2 + \lambda\sum_{i}S(\frac{\alpha_i}{\sigma_i}) \quad (2)$$

where A is basis function matrix, α_i is the i response value of the column vector S and $\phi_i(x,y)$ is the i column vector in A. The first item in formula (2) is information preserving degree of reconstructed image calculated by error sum of squares. The second item is the sparse characteristics of the encoding, which is to use fewer basis functions to represent the original image.

Neurophysiologic studies show that primary visual cortex VI area receives the nonnegative data from lateral geniculate nucleus and the stimulation of neurons in VI region can not be negative, which indicates that it is not accurate to use SC to simulate the receptive field of simple cells. The non negative of the image input data can better simulate the receptive field behavior of mammals [11]. In 2002, Hoyer proposed NSC algorithm which can not only reflect the statistical characteristics of the data but can also obtain the local feature representation of the interested object [12]. Hoyer uses the method of iterative updating dictionary A and sparse efficient S. First, fixing A, consider the optimization of S and then fixing S, update dictionary A. Due to the good expression ability of NSC, it has been successfully applied to face recognition, image retrieval and other areas of computer vision.

2.2 Fisher Discriminative Analysis

The main idea of Fisher discriminative analysis [13] is to establish a new discriminative criterion for the linear combination of multiple observations and minimize the ratio of variance within group to variance between groups. Assume that $X = [x_1, x_2, \cdots, x_l, \cdots, x_C]$ is a collection of training samples that contain C classes, where x_l is class l training sample and each training sample contains N_l feature points $x_l^k (k = 1, 2, \cdots, N_l)$. The total number of feature points for all types of samples is N. The Fisher discriminative is

$$F(x) = S_W / S_B \quad (3)$$

where S_W is within-class scatter of X and S_B is between-class scatter of X.

$$S_W = \frac{1}{N}\sum_{l=1}^{C}\sum_{k=1}^{N_1}(x_l^k - \mu_l)^T(x_l^k - \mu_l) \quad (4)$$

$$S_B = \frac{1}{N}\sum_{l=1}^{C}N_l(\mu_l - \mu)^T(\mu_l - \mu) \quad (5)$$

where μ_l is the sample mean of class l and μ is the sample mean of all class.

$$\mu_l = \frac{1}{N_l}\sum_{k=1}^{N_l}x_l^k \quad (6)$$

$$\mu = \frac{1}{N}\sum_{l=1}^{C}\sum_{k=1}^{N_l}x_l^k \quad (7)$$

2.3 Deep Belief Network

Deep learning network has a hierarchical structure, which can automatically learn high-level features from low-level features. This paper uses deep belief networks to represent the relationship between low-level features and high-level semantics of aerial image and then perform image classification.

In 2006, Hinton et al proposed a model of DBN and successfully applied it to the recognition of handwritten font. The basic structure of DBN is restricted boltzmann machine (RBM), as shown in figure 1:

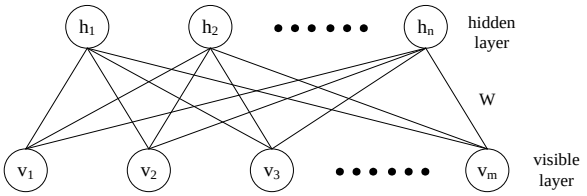


Figure 1. RBM Model

RBM is a type of two layer neural networks comprised of a visible layer that represents the observed data and a hidden layer that represents the hidden variables. Connections only exist between the visible layer and the hidden layer.

RBM is an energy based model, and its energy function is defined in formula (8):

$$E(v,h) = -\sum_{i=1}^{I}\sum_{j=1}^{J} v_i h_j w_{ij} - \sum_{i=1}^{I} a_i v_i - \sum_{j=1}^{J} b_j h_j \tag{8}$$

where w is weight matrix, b are visible unit biases and a are hidden unit biases.

Based on the energy function, the definition of the joint distribution is in formula (9):

$$P(v,h) = \frac{\exp(-E(v,h))}{Z} \tag{9}$$

where Z is called the partition function and

$$Z = \sum_{v}\sum_{h} \exp(-E(v,h)) \tag{10}$$

Conditional probability distributions are as follows:

$$p(h_j = 1|v) = \delta(b_j + \sum_{i=1}^{I} v_i w_{ji}) \tag{11}$$

$$p(v_i = 1|h) = \delta(a_i + \sum_{j=1}^{J} h_j w_{ji}) \tag{12}$$

where $\delta(x)$ is sigmoid function.

Each two hidden layers constitute a RBM network and the top layer is back propagation (BP) neural network, as shown in figure 2:

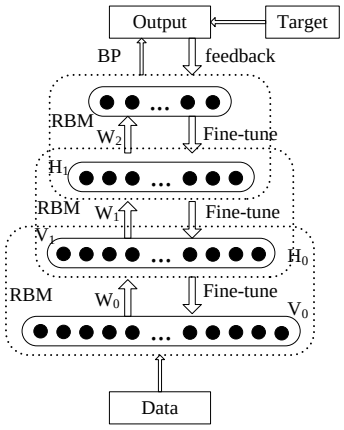


Figure 2. DBN Structure

The lower layer of the RBM extracts and abstracts the input data and put it as the high layer input. The DBN is trained by the combination of pre-training and fine tuning. First, the unsupervised training of each layer of RBM is carried out in a bottom-up way and then the supervised BP neural network is used to fine tuning the whole model in a top-down way.

After training DBN, this paper put the sparse feature of the image as input and the category of image as output to train a new neural network structure, which is suitable to classify aerial images.

3. Proposed Method

In order to enhance the discrimination of sparse representation vectors, this paper adds the Fisher discriminative criterion based on NSC and proposes nonnegative Fisher discriminative sparse coding (FDNSC) algorithm. Combined with DBN, this paper also proposed an aerial image classification method based on FDNSC and DBN.

3.1 Fisher Discriminative Nonnegative Sparse Coding (FDNSC) Algorithm

In general, image representation based on SC can be divided into dictionary learning and sparse decomposition. This paper uses FDNSC to learn dictionary and on this basis to construct visual word library and generate image sparse representation vector.

3.1.1 The Model of FDNSC

This paper added fisher discriminative criterion into the objective function of the dictionary learning and proposed FDNSC model. For the convenience of description, this paper divide the original image blocks into n dimensional column vector I. $I_i(i=1,...,N)$ represents each pixel point. The reconstructed image block is represented by $Y_i(i=1,...,N)$. The objective function of FDNSC is

$$\min_{A,S} E(A,S)=\sum_{i=1}^{N}(I_i-Y_i)^2+\lambda_1\cdot\ln F(x)+\lambda_2\cdot\sum_{i=1}^{M}S(\frac{\alpha_i}{\sigma_i}) \tag{13}$$

where λ_1 and λ_2 is weight coefficient, $F(x)$ is shown in formula (3) and $S(\alpha_i/\sigma_i)$ is sparse penalty function. The first and the third item are consistent with SC model. The first item is the error sum of squares of the reconstructed image and the original image and the third item is the sparsity of the function. The second item is a judgment used to

enhance the discriminative of sparse representation coefficients X and make X have better class discrimination. For the convenience of calculation, this paper uses $\ln F(x)$.

The over complete dictionary A and the sparse representation coefficient S can be obtained by solving the optimization problem of the formula (13). To ensure the non negative characteristics of sparse coding, this paper uses a learning algorithm similar to NSC, which is to solve one variable by fixing another variable.

Step 1 Fixing A, the optimization problem of formula (13) can be rewritten as

$$\min_{S} E(A,S) = \sum_{i=1}^{N}(I_i - Y_i)^2 + \lambda_1 \cdot \ln F(x) + \lambda_2 \cdot \sum_{i=1}^{M} S(\frac{\alpha_i}{\sigma_i}) \tag{14}$$

Use gradient optimization algorithm [14] to solve sparse representation coefficients S. The iterative formula is

$$\nabla E(s_i) = \frac{\partial E(A,S)}{\partial s_i} = -2\sum_{k=1}^{N}(I_k - Y_k)\phi_{k,i} + \lambda_3 \frac{(s_i - \mu_l)}{S_W} - \lambda_4 \frac{(\mu_l - \mu)}{S_B} + \lambda_2 \nabla \sum_{i=1}^{M} S(\frac{\alpha_i}{\sigma_i}) \tag{15}$$

where $\lambda_3 = \frac{2\lambda_1(N_l - 1)}{N \cdot N_l}$, $\lambda_4 = \frac{2\lambda_1(N - N_l)}{N^2}$. $\phi_{k,i}$ is the element in basis function matrix A. S_W and S_B are shown in formula (4) and (5). μ_l and μ are shown in formula (6) and (7).

Step 2 Fixing S, the optimization problem of formula (13) can be rewritten as

$$\min_{A} E(A,S) = \sum_{i=1}^{N}(I_i - Y_i)^2 + \lambda_2 \cdot \sum_{i=1}^{M} S(\frac{\alpha_i}{\sigma_i}) \tag{16}$$

Use incremental codebook optimization algorithm [15] to solve over complete dictionary A.

To sum up, the FDNSC learning algorithm is:

Step 1 Initialize dictionary A by randomly sampling feature set Y.

Step 2 Fix A and use gradient optimization algorithm and formula (15) to generate sparse representation coefficient S.

Step 3 Fix S and use incremental codebook optimization algorithm to generate over complete dictionary A.

Step 4 If E (A, S) < threshold ε, end iteration or turn to step 2.

3.1.2 Visual Words Library Construction

In this stage, this paper is going to find a set of basis functions and corresponding sparse weights that can be used to reproduce the original feature matrix X with least reconstruction error.

To construct a basis function (dictionary), first, randomly sample low-level features from the entire data set to generate matrix $X = [x_1, x_2, ..., x_n]$. Next, given the feature matrix X, this paper learn the basis functions by finding best solution for a minimization problem which is similar to the sparse coding framework. The basis function D is learned using alternate minimization of formula (17):

$$\min_{D,s_i} \sum_i \left\| Ds_i - x_i \right\|_2^2$$

subject to $\|D_j\|_2 = 1, \forall j$

and $\|s_i\|_0 \le k, \forall i$ (17)

where $\|s_i\|_0$ is the number of nonzero elements in column vector s_i.

The constructive process of the visual words library based on FDNSC is as follows:

Step 1 Randomly select a number of images from the training image and extract SIFT features $F = \{f_1, f_2, ..., f_N\}$, where N is the number of SIFT feature vectors. $F = \{f_1, f_2, ..., f_N\}$ is corresponding to training set $Y = \{y_1, y_2, ..., y_N\}$ in formula (13).

Step 2 Solve the sparse representation coefficient S and the visual words library V of the SIFT feature vector set by using the iterative algorithm in section 3.1.1. The visual words library V is corresponding to the over complete dictionary A.

3.1.3 Sparse Representation

After the construction of visual words library V, the sparse representation of the image can be made. The process is as follows:

Step 1 Input image I and visual words library V.

Step 2 Extract SIFT features and generate SIFT feature vector set $F = \{f_1, f_2, ..., f_M\}$, where M is the number of SIFT feature vectors.

Step 3 According to visual words library V, use gradient optimization algorithm to solve formula (14) and generate sparse feature vector α_i:

$$f_i \approx V\alpha_i \tag{18}$$

Step 4 Get all sparse feature vectors $\alpha = \{\alpha_i\}(i = 1, 2, ..., M)$.

After sparse representation, the new feature representation for an image scene will usually have a very high dimensionality. For computational efficiency and storage volume, it is standard practice to use a pooling strategy to reduce the dimensionality of the image representation. With the sparse features vector α_i computed for an image patch, this paper estimate the final feature representation as follows:

$$P = \frac{1}{M}\sum_{i=1}^{M}\alpha_i \tag{19}$$

At this point, the sparse feature representation of aerial image p is got. p is used as the input vector of DBN to complete the high-level image classification.

3.2 Aerial Image Classification Algorithm Based on FDNSC

After sparse representation of the image, this paper combine DBN to complete the classification. The flow chart of the algorithm based on FDNSC is shown in figure 3. From the flow chart, the aerial image classification algorithm includes training stage and testing stage.

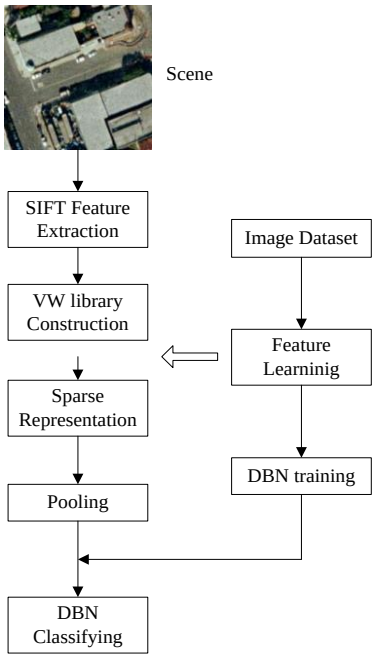


Figure 3. Framework of the Proposed Aerial Image Classification

The training stage includes feature learning and DBN training. First, extract SIFT features and generate SIFT feature vector set $F = \{f_1, f_2, ..., f_N\}$ where N is the number of SIFT feature vectors and then construct visual words library V and sparse feature vector S. Finally use sparse feature vector S to train DBN. The training steps of DBN are as follows:

Step 1 Input sparse feature vector S.

Step 2 Train the first layer of RBM.

Step 3 Use the output of the first layer RBM as the input of the second layer RBM and train the second layer RBM.

Step 4 Repeat step 2 and step 3 until training all RBM.

Step 5 Use BP algorithm to fine tuning the parameters of the whole network.

The testing stage includes SIFT feature extraction, VW library construction, sparse representation, pooling and DBN classifying. First, extract SIFT features and generate SIFT feature vector set $F = \{f_1, f_2, ..., f_M\}$ where M is the number of SIFT feature vectors and then generate sparse feature vector $\alpha = \{\alpha_i\}(i = 1, 2, ..., M)$ according to visual words library V and generate sparse feature representation P by mean pooling. Finally, put P into DBN to identify and get the final classification results.

4. Experiment

4.1 Experimental Setup

This paper uses dense-SIFT algorithm [16]. First, divide the image into image blocks with the size of 16×16 pixels and interval of 8 pixels. Then divide the image block into 4×4 sub regions and calculate the gradient histogram of 8 directions in each sub region as the seed point. Finally, the seed points are connected to the 128 dimensional feature vectors.

In order to verify the effectiveness of the proposed algorithm, this paper compare it with the BoVW method reported in [1], the SPM-PLSA method reported in [4], the SCSPM and the BHM method proposed in [17], the NSLLC method proposed in [18],

the SPMK method described in [19] and the SC+SVM method described in [20]. The SCSPM method use pyramid structures to extract SIFT features and the layer of the pyramid is set to 3. The nearest neighbor parameter k of the NSLLC is set to 5. The semantic subject number of the BHM method is set to 40. The pyramid layer of the SPM-PLSA is set to 1×1, 2×2, 4×4.

This paper validate the aerial scene classification on two data sets, OT data set and UC Merced data set. With OT data set and UC Merced data set, randomly select 80% samples from each class to initialize the training set and the remaining 20% as the testing set. Repeated 10 classification experiments on each data set and the classification accuracy of the 10 experiments were averaged as the final classification accuracy.

4.2 Experiment On OT Data Set

OT data set contains 8 aerial scene categories. (1) Forest. (2) Mountain. (3) Open Country. (4) Coast. (5) Highway. (6) City. (7) Tall Building. (8) Street. This paper select 100 images per class. Figure 4 shows some of the images in the OT data set.

Figure 4. Few Example Images form the OT Data Set

In order to study the sensitivity of the sparsity parameter on OT data set, this paper varied their values over a wide range. Figure 5 shows the classification performance with different sparsity parameter values. The results showed that there was a wide range of sparsity values for which the classification performance was consistent, and the best classification performance was obtained at a sparsity value close to 0.4. Based on this analysis, this paper set the sparsity value as equal to 0.4 to generate sparse features.

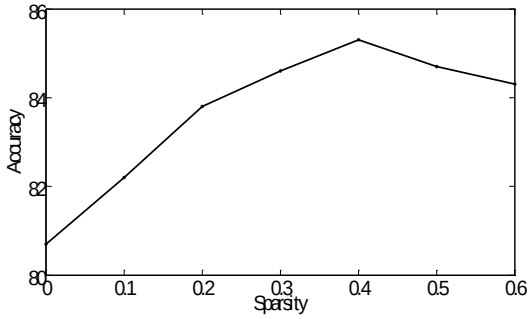


Figure 5. Effect of the Sparsity Parameter Value on the Classification Accuracy with the OT Data Set

To test the effects of different layers of DBN and different iterations of RBM on the classification performance, this paper verify the classification accuracy of three different layers of DBN as shown in figure 6.

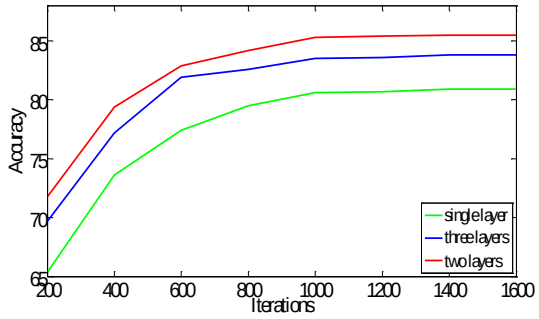


Figure 6. Effect of the Layer of DBN and Iterations of RBM on the Classification Accuracy

From figure 6, in the initial stage, the classification accuracy is significantly improved with the increase of the number of iterations. When the number of iterations is greater than 1000, the classification accuracy is almost unchanged. So the number of iterations is set to 1000. Besides, the 2-layers DBN outperforms the single-layer DBN and 3-layers DBN. So this paper select 2-layers DBN.

To evaluate classification performance under different feature sizes, this paper measured the overall classification accuracy with the OT data set for values of dictionary sizes ranging from 100 to 2000 as shown in figure 7. From figure 7, the classification accuracy improves with the increase of feature sizes. When the number of feature sizes is greater than 600, the classification accuracy tends to be gentle. The experimental analysis shows that values of feature sizes around 1300 produced excellent accuracy across all the data sets.

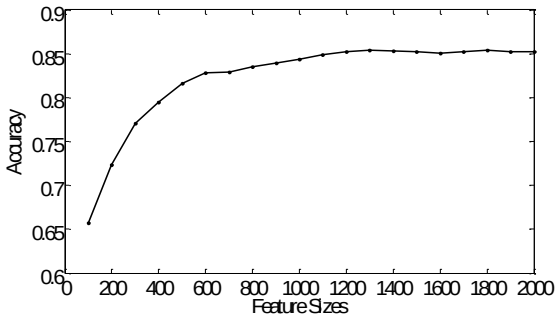


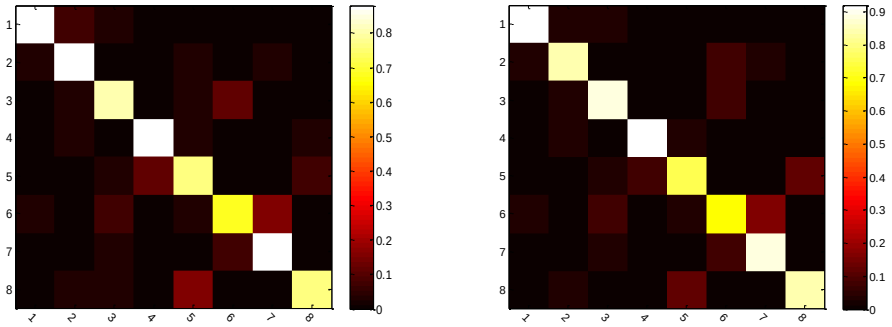
Figure 7. Effect of Feature Sizes on the Classification Accuracy with the OT Data Set

To evaluate the classification performance of our proposed method, this paper compared it with seven other methods with the OT data set. Of the eight strategies, our method produced better performance, as shown in table 1. This paper also compared the classification performances with and without the Fisher Discrimination. The results illustrated that using Fisher Discrimination is an efficient way to increase the scene classification accuracy. The reason is that sparse representation provides a simple representation of redundant information and represent features in a more concise and effective way.

Table 1. Comparison with the Previous Reported Accuracies on OT data set

Methods	SCSPM	NSLLC	BoVW	BHM	SPM-PLSA	SPMK	SC+SVM	Non FD	With FD
Accuracy	65.5%	74.2%	75.5%	76.3%	76.5%	78.3%	84.7%	83.9%	**85.3%**

In order to find out the wrong classification, this paper made the confusion matrices which are reported in figure 8. The confusion matrix generated for the proposed method in figure 8(b) shows that the classification errors were mainly from scenes that share similar structures, such as street and highway. The overall accuracies of OT data set are reported in figure 9. From figure 9, the highest accuracy for the classification is forest, coast, and tall building, which have a regular textural and spatial structure.



(a) (b)

Figure 8. Confusion Matrix Showing the Classification Performance with the OT data set: (a) non Fisher Discrimination (b) with Fisher Discrimination

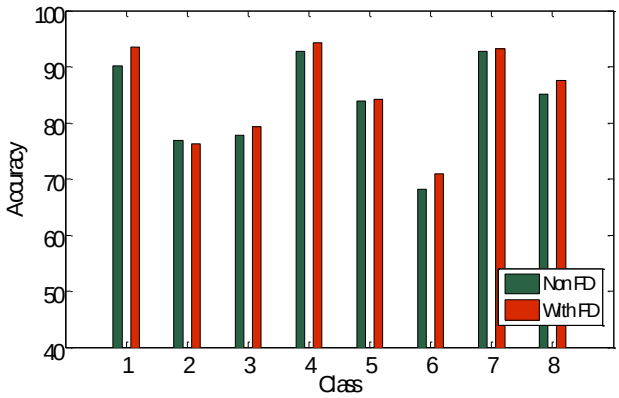


Figure 9. The Overall Accuracies with the OT Data Set for the Proposed Method

4.3 Experiment On UC Merced Data Set

UC Merced data set contains manually extracted images from the USGS National Map Urban Area Imagery collection. UC Merced data set consists of 256 × 256 color images from 21 aerial scene categories. (1) Agricultural. (2) Airplane. (3) Base-ball diamond. (4) Beach. (5) Buildings. (6) Chaparral. (7) Dense residential. (8) Forest. (9) Freeway. (10) Golf course. (11) Harbor. (12) Intersection. (13) Medium residential. (14) Mobile home park. (15) Overpasses. (16) Parking lot. (17) River. (18) Runway. (19) Sparse residential. (20) Storage tanks. (21) Tennis court. The data set contains highly overlapping classes and has 100 images per class. Figure 10 shows some of the images in the UC Merced data set.

Figure 10. Few Example Images form the UC Merced Data Set

This paper first compared the classification accuracies for varied sparsity parameter values on the UC Merced data set. Set the sparsity from 0 to 0.6 and measure the classification accuracy in the same way as before, which is shown in figure 11. The results showed that a sparsity value close to 0.3 generated the best accuracy. Based on this analysis, this paper set the sparsity value equal to 0.3 to generate the feature extractors.

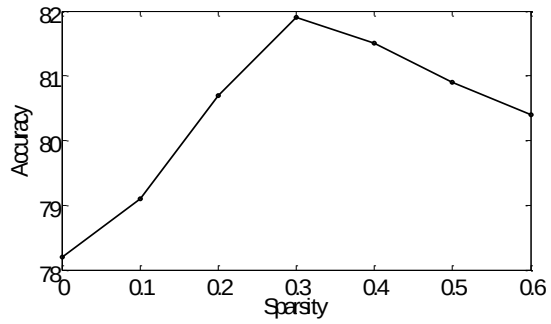


Figure 11. Effect of the Sparsity Parameter Value on the Classification Accuracy with the UC Merced Data Set

To evaluate the classification performance with different feature sizes, this paper measured the overall classification accuracy with the UC Merced data set for values of feature sizes ranging from 100 to 2000, which is shown in figure 12. From figure 12, in the initial stage, the classification accuracy improves with the increase of feature sizes. When the number of feature sizes is greater than 1200, the classification accuracy is almost unchanged and the feature size of 1400 produced an excellent accuracy with this data set.

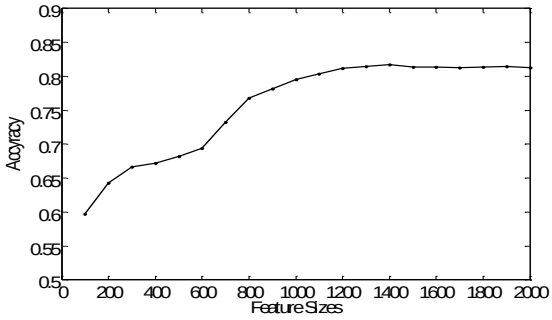


Figure 12. Effect of Feature Sizes on the Classification Accuracy with the UC Merced Data Set

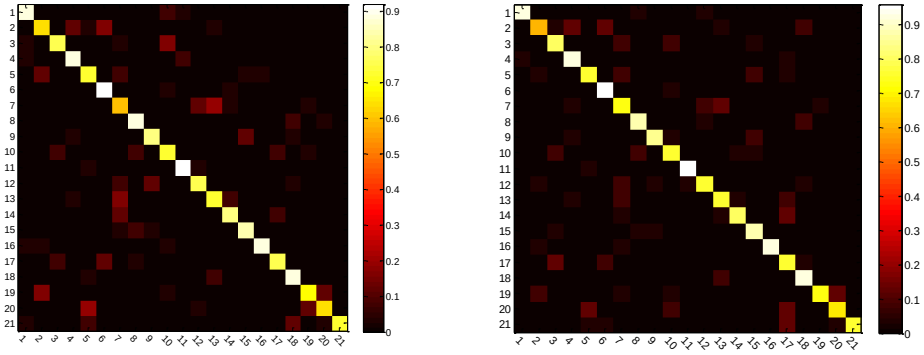
To test the classification performance of our proposed method in a larger data set, this paper measured the classification performance with the UC Merced data set. Compared with seven other methods, the results are shown in table 2. Our proposed method also has

a better performance on larger data set. This paper also compared the classification performance with and without the Fisher Discrimination to validate that sparse representation is a required step to characterize the scene effectively.

Table 2. Comparison with the Previous Reported Accuracies on UC Merced Data Set

Methods	SCSPM	NSLLC	BoVW	BHM	SPM-PLSA	SPMK	SC+SVM	Non FD	With FD
Accuracy	64.3%	70.4%	71.9%	73.2%	73.5%	74%	81.7%	80.8%	**81.9%**

As before, this paper made the confusion matrices to find out the wrong classification as shown in figure 13. From figure 13, the classification errors were mainly from scenes that share similar structures, such as Buildings, Sparse residential and Storage tanks. Also, the results show that local edge based features lacks the ability to capture distinguishing shape patterns which are important for discriminating classes such as the airplane, baseball diamond, and storage tanks. The overall accuracies of UC Merced data set is shown in figure 14. From figure 14, the highest accuracy for the classification is Chaparral, Harbor and runway, which have a regular textural and spatial structure.



(a) (b)

Figure 13. Confusion Matrix Showing the Classification Performance with the UC Merced Data Set: (a) non Fisher Discrimination (b) with Fisher Discrimination

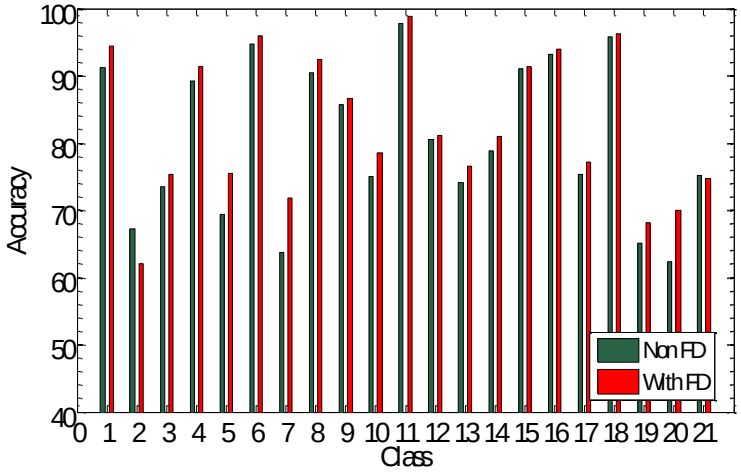


Figure 14. The Overall Accuracies with the UC Merced Data Set for the Proposed Method

According to the above two experiments, the proposed method is more advantageous and have higher classification accuracy compared with other methods. The reason is that the FDNSC can effectively extract the concise feature of aerial image and can make the same type of image sparse representation coefficient distance closer and the sparse representation of different types of images distance farther, which is more conducive to image classification. Besides, combined with DBN, the proposed method has a stronger ability to image classification, which makes the image classification accuracy higher.

According to table 1 and table 2, the proposed method does not have obvious advantages compared with SC+SVM. The reason is the limited training data. There is not enough data in each category to train a DBN network which can well express the relationship between low-level features and high-level semantic of aerial image. With the increase of training data, deep network will have more obvious advantages than shallow network.

5. Conclusion

In order to overcome the problem of large representation errors, lack of spatial information and weak discriminative of NSC, this paper added the Fisher Discriminative criterion into NSC and proposed FDNSC algorithm, which can make the same type of image sparse representation coefficient distance closer and the sparse representation of different types of images distance farther. In order to enhance the classification ability of aerial image, this paper also proposed an aerial image classification method based on FDNSC and DBN. Experimental results on OT data set and UC Merced data set indicate that the proposed method can promote the spatial separability of sparse coefficients and enforce the discrimination and classification capability for aerial images. The proposed method obtains results that are equal to or even better than the previous results and is more suitable for aerial image classification. Besides, under the premise of meeting the amount of data, deep learning will have a greater advantage on image classification and machine learning. As future extensions, we plan to apply this method to different scale data set and further optimize feature extraction algorithm and sparse encoding mode.

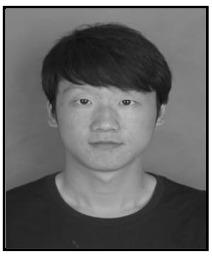
References

[1] Gong Cheng and Lei Guo, Tianyun Zhao, "Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and Plsa" , International Journal of Remote Sensing, vol. 34, no. 34, **(2013)**, pp. 45-59.
[2] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", IEEE Transactions on Computer Vision and Pattern Recognition, no. 2, **(2006)**, pp. 2169-2178.
[3] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification", Proceedings of IEEE International Conference on computer vision, **(2011)**, pp. 1465–1472.
[4] E. Ergul and N. Arica, "Scene classification using spatial pyramid of latent topics", Proceedings of IEEE International Conference on Pattern Recognition, **(2010)**, pp. 3603-3606.
[5] T. Kato, H. Hino and N. Murata, "Multi-frame image super resolution based on sparse coding", Neural Networks the Official Journal of the International Neural Network Society, no. 66, **(2015)**, pp. 64–78.
[6] Jianchao Yang, Kai Yu, Yihong Gong and Thomas Huang , "Linear spatial pyramid matching using sparse coding for image classification", IEEE transactions on Computer Vision and Pattern Recognition, no. 9, **(2009)**, pp. 1794-1801.
[7] Lihe Zhang, Lei Pan and Tao Liu, "Image Classification based on kernel Laplasse sparse encoding", Journal of Dalian University of Technology, vol. 2, no. 55, **(2015)**, pp. 192-197.
[8] G. E. Hinton, S. Osindero and Y. W. Teh, "A fast learning algorithm for deep belief nets", Neural Computation, vol. 18, no. 7, **(2006)**.
[9] Qi Lu, Yong Dou, Xin Niu, Jiaqing Xu and Fei Xia, "Remote sensing image classification based on DBN model", Journal of Computer Research and Development, vol. 9, no. 51, **(2014)**, pp. 1911-1918.
[10] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images", Nature, vol. 6583, no. 381, **(1996)**, 381(6583):607-609.

[11] Chunjie Zhang, Jing Liu and Chao Liang, "Image classification by non-negative sparse coding, correlation constrained low-rank and sparse decomposition", Computer Vision & Image Understanding, vol. 7, no. 123, **(2014)**, pp. 14-22.
[12] Xinzheng Zhang, Qizheng Wu and Shujun Liu, "HRR profiles time-frequency non-negative sparse coding for SAR target classification", Progress in Electromagnetics Research B, vol. 1, no. 60, **(2014)**, pp. 63-77.
[13] Guoqiang Wang , Nianfeng Shi, Yunxing Shu and Dianting Liu, "Embedded Manifold-Based Kernel Fisher Discriminant Analysis for Face Recognition", Neural Processing Letters, vol. 1, no. 43, **(2016)**, pp. 1-16.
[14] H. Lee, A. Battle and R Raina, "Efficient sparse coding algorithms", Advances in Neural Information Processing Systems, no 19. **(2006)**, pp. 801-808.
[15] Jinjun Wang, Jianchao Yang, Kai Yu, Thomas Huang and Yihong Gong, "Locality-constrained Linear Coding for image classification", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2010), pp. 3360-3367.
[16] Yu Liu, Shuping Liu and Zengfu Wang, "Multi-focus image fusion with dense SIFT", Information Fusion, no. 23, **(2015)**, pp. 139-155.
[17] Fei-Fei L, Perona P, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition, **(2005)**, pp. 524-531.
[18] Hoyer P O, "Non-negative sparse coding", Proceedings of IEEE Workshop on Neural Networks for Signal Processing, **(2002)**, pp. 557-565.
[19] Fan Zhang, Bo Du, Liangpei Zhang, "Saliency-Guided Unsupervised Feature Learning for Scene Classification", IEEE Transactions on Geoscience & Remote Sensing, vol. 4, no. 53, **(2015)**, pp. 2175-2184.
[20] A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification", IEEE Transactions on Geoscience & Remote Sensing, vol. 1, no. 52, **(2014)**, pp. 439-451.

Authors

Tao Shi, he received the Ph.D. degree from University of Science and Technology Beijing, China. Now, he is an associate professor at the College of Electrical Engineering, North China University of Science and Technology, China. His research interests include pattern recognition, intelligent control and cognitive robot.



Chun-Lei Zhang, he is now a master student in College of Electrical Engineering, North China University of Science and Technology, China. His research interests include pattern recognition and image processing.

Hong-Ge Ren, she received the Ph.D. degree from Beijing University of Technology, China. Now, she is an associate professor at the College of Electrical Engineering, North China University of Science and Technology, China. Her research interests include artificial intelligence and cognitive robot.

Fu-Jin Li, he received the Ph.D. degree from China University of Mining and Technology, China. Now, he is an professor at the College of Electrical Engineering, North China University of Science and Technology, China. His research interests include intelligent control and intelligent instrument.