LETTER

Compact Sparse Coding for Ground-Based Cloud Classification

Shuang LIU¹, Nonmember, Zhong ZHANG¹(✉), Member, and Xiaozhong CAO¹, Nonmember

SUMMARY Although sparse coding has emerged as an extremely powerful tool for texture and image classification, it neglects the relationship of coding coefficients from the same class in the training stage, which may cause a decline in the classification performance. In this paper, we propose a novel coding strategy named compact sparse coding for ground-based cloud classification. We add a constraint on coding coefficients into the objective function of traditional sparse coding. In this way, coding coefficients from the same class can be forced to their mean vector, making them more compact and discriminative. Experiments demonstrate that our method achieves better performance than the state-of-the-art methods.

1. Introduction

Clouds play an essential role in the earth radiation balance and climate change, and are critical inputs for flight planning and aviation. Ground-based cloud classification, as an important cloud observation technique, has received more and more attention from the research community. This is because successful cloud classification can improve the precision of weather prediction and help us to understand atmospheric conditions. At present, clouds are classified by human observers who have received professional training [1]. Despite their unquestionable usefulness, different observers will output different classification results and it is time-consuming. Hence, an automatic ground-based cloud classification technique is urgently needed.

Automatic cloud classification is a challenging task due to the extreme appearance variations under different atmospheric conditions, which makes this technique under-developed. At present, existing cloud classification techniques are generally based on the characteristics of structure and texture in cloud images. The algorithms based on structure features include cloud fraction and edge sharpness [2], and fourier transformation [3]. While the algorithms based on texture features include co-occurrence and auto-correlation matrices [4], and local binary patterns and it extension [5], [6]. In addition, several algorithms [1], [7] have been proposed that fuse these two characteristics as the final representation. Although these works are suggestive, they fail to extract really useful information from cloud images. It is because cloud images, as one kind of natural texture, usually possess very large intra-class variations due to the large variation in illumination, climate and deformation. Therefore, it demands a more powerful algorithm to extract discriminative information of cloud images.

Fortunately, the sparse coding model has been proposed as a very effective method for texture and image classification [8] [9]. In the training stage, the dictionary and coding coefficients of features are iteratively learned class by class. The training stage, however, neglects the relationship of coding coefficients from the same class. We expect that the coding coefficients from the same class possess the characteristic of compactness. Namely, the coding coefficients from the same class should be similar.

In this paper, a novel algorithm named compact sparse coding is proposed for ground-based cloud classification, it considers the compactness of samples from the same class. Specifically, a constraint on the coding coefficients is added into the objective function of traditional sparse coding model. We force the coding coefficients from the same class to their mean vector and therefore they are more compact and discriminative. Our method is verified on two challenging cloud image databases, and experiments demonstrate that our method achieves better results than previous methods on ground-based cloud classification.

2. The Proposed Algorithm

In this section, we first briefly review the traditional sparse coding; then describe the proposed compact sparse coding algorithm and give the optimization procedures. Finally, the feature representation is applied for ground-based cloud classification.

2.1 Traditional Sparse Coding

Sparse coding has emerged as an extremely successful tool for statistical signals processing. For a signal $x \in \mathbb{R}^{M \times 1}$, we say that $x$ has a sparse approximation over a dictionary $D = \{d_1, d_2, \ldots, d_N\} \in \mathbb{R}^{M \times N}$, where $M \ll N$. Then, the sparsest coding of $x$ over $D$ is the solution of

$$\min_{a} \|a\|_0 \quad \text{s.t.} \quad \|x - Da\|^2 \leq \varepsilon$$

(1)

where $\| \cdot \|_0$ denotes the $l_0$-norm which counts the number of
non-zero elements in a vector, and $\varepsilon$ is a pre-defined parameter with a small value. Given the dictionary $D$, the model tries to seek the sparsest representation for the signal $x$. However, the solution of Eq. (1) is NP-hard. Some recent work shows that this problem can be tackled by replacing the $l_0$-norm with $l_1$-norm regularization [10]. In many applications, the dictionary $D$ is unknown and we need to construct it from training data $X = \{x_1, x_2, \ldots, x_L\}$. The dictionary $D$, as well as the sparse coding coefficient $a_i$, can be learned by optimizing the following objective function:

$$
\min_{D,A} \sum_{i=1}^{L} \|x_i - DA_i\|^2 + \lambda \|a_i\|_1
$$

s.t. $\|d_i\|^2 \leq 1, \forall k = 1, 2, \ldots, N$

where $A = [a_1, a_2, \ldots, a_L]$ ($a_i \in \mathbb{R}^{N \times 1}$) and $\lambda$ is the regularization parameter controlling the sparsity of the coefficient vector. $\| \cdot \|_1$ denotes the $l_1$-norm which counts the sum of the absolute value of each element in $a_i$, and the unit $l_2$-norm constraint on $d_i$ is to avoid trivial solutions.

2.2 Compact Sparse Coding Model

The traditional sparse coding, however, neglects the relationship of coding coefficients from the same class, which results in suboptimal performance. Figure 1(a) shows the distribution of coding coefficients produced by traditional sparse coding model (red dots represent the distribution of one class, and green squares represent the distribution of another class). We can see that part of them confuse together, so it is difficult to classify. In order to alleviate the confusion of different classes, the compactness of samples from the same class needs to be considered. We force the coding coefficients from the same class to their mean vector. Based on the above consideration, the constraint on the coding coefficients can be formulated as:

$$
\| a_i - \mu \|^2, \quad i = 1, 2, \ldots, L
$$

where $a_i$ denotes the coding coefficient of feature vector $x_i$; $\mu$ is the mean vector of all the coding coefficients from the same class, i.e., $\mu = \sum_{i=1}^{L} a_i$; $L$ is the number of feature vectors in this class. Equation (3) is actually served as a penalty term, which results in a high penalty if the coding coefficients from the same class distribute dispersively.

By adding Eq. (3) into the traditional sparse coding model, the objective function of the proposed compact sparse coding is formulated as:

$$
\min_{D,A} \sum_{i=1}^{L} \|x_i - DA_i\|^2 + \lambda \|a_i\|_1 + \rho \|a_i - \mu\|^2
$$

s.t. $\|d_i\|^2 \leq 1, \forall k = 1, 2, \ldots, N$

where $\rho$ is a regularization parameter which controls the compactness of coding coefficients.

The distribution of coding coefficients is close to their mean vector by utilizing the proposed method, and therefore possess the characteristic of compactness. As shown in Fig. 1(b), the intra-class variation is reduced by considering the relationship of coding coefficients from the same class. The large red dot and green square denote the mean vectors of two different classes respectively. As a result, different classes can be easily separated and the performance of classification will be improved.

2.3 Solution of the Compact Sparse Coding

The optimization of compact sparse coding model can be conducted by alternatively optimizing $D$ and $A$. When $D$ is fixed, this optimization problem Eq. (4) can be executed by optimizing over each coefficient $a_i$ individually:

$$
\min_{a_i} \|x - DA_i\|^2 + \lambda \|a_i\|_1 + \rho \|a_i - \mu\|^2
$$

This is a linear regression problem with $l_1$-norm regularization on the coefficients. The optimization can be solved very efficiently by the feature-sign search algorithm [11]. After optimizing each $a_i$, $i = 1, \ldots, L$, the coefficient matrix $A$ is updated. It should be noted that the mean vector $\mu$ keeps invariant when optimizing each $a_i$, and $\mu$ is updated after optimizing all the $a_i$. Once the coefficient matrix $A$ is updated, we update the dictionary $D$ which can be handled by a least square problem with quadratic constraints as:

$$
\min_{D} \|X - DA\|_F^2
$$

s.t. $\|d_i\|^2 \leq 1, \forall k = 1, 2, \ldots, N$

where $\| \cdot \|_F$ is the Frobenius norm. This can be efficiently solved by using the Lagrange dual method [11].

In summary, the whole optimization process can be described in Algorithm 1.

2.4 Feature Representation

We learn a dictionary $D_k$ ($k = 1, 2, \ldots, C$) for each class by utilizing Algorithm 1. Then the intra-class dictionaries learned from all the $C$ classes can be concatenated into a dictionary $D$:
Algorithm 1: Compact Sparse coding

Input: \( x_i, i = 1, \cdots, L, L \) is the number of feature vectors from a class; parameters \( \lambda, \rho \)

Output: \( D, A \)

1. Initialize Obtain dictionary \( D \) by k-means clustering algorithm;
2. while \( D \neq D_{\text{new}} \) do
   1. Fix \( D \) and then optimize \( A \),
   2. update \( \mu = \sum_{i} a_i \),
   3. Fix \( A \), and optimize \( D \), which can be optimized by Eq. (6). \( D_{\text{new}} \) is obtained.
3. end

\[
D = \{D_1, D_2, \ldots, D_C\} = \{d_1, d_2, \ldots, d_N\}
\]  
(7)

where \( N \) is the total number of textons learned from \( C \) classes.

After obtaining \( D \), we can encode the input features for classification. For a feature vector \( x_n \), we solve the coefficient \( a_n \) of \( x_n \) over \( D \) by Eq. (2). Let \( A = [a_1, a_2, \ldots, a_W] \) denote the coefficient matrix of a cloud image, where \( W \) is the number of feature vectors extracted from this cloud image. The final representation for the cloud image is obtained by max pooling function on each row of \( A \).

3. Experimental Results

To verify the effectiveness of the proposed method, we carry out a series of experiments on two ground-based cloud databases: Kiel database and IapCAS-E database, and compare our method with several excellent published approaches. These methods include local binary pattern [5], spectral-texture feature [7], census transform-structure feature [1], traditional sparse coding (T_SC) [8], the non-negative sparse coding (NN_SC) [12], laplacian sparse coding (LSC) [13], and K-SVD sparse coding [14]. For all the coding models, we use SVM as the classifier. In the following experiments, each image is normalized to have zero mean and unit standard deviation. The image feature is an 81 dimensional vector, which is formed by stretching a 9 \times 9 neighborhood around each pixel. Then, all the patch vectors are normalized via weber’s law [15] as follows, which can further enhance intensity invariance

\[
x' = x \frac{\log(1 + ||x||_2/0.03)}{||x||_2}
\]  
(8)

Note that all the patches are normalized according to weber’s law in the proposed method and all compared methods.

The first cloud database is the Kiel database which is provided by Kiel University in Germany. According to the international cloud classification system published in WMO, the database is divided into seven classes. This database contains 1500 cloud images and each class samples are shown in Fig. 2(a). The dictionary size for each class is set to 200, which results in a 1400 dimensional histogram for each cloud image. In this experiment, 1/2, 1/3, 1/4 samples are randomly chosen from each class as training data while the remaining cloud images are used for testing, and the partition process is implemented 50 times independently. Prior to the discussion of experimental results, we first study the influence of parameters \( \lambda \) and \( \rho \) in Eq. (4). Figure 3 shows the average accuracy values under different \( \lambda \) and \( \rho \), from which we can see that when \( \lambda = 0.1 \) and \( \rho = 0.04 \) results are the best.

With the optimal parameters, we randomly select 1000 local patches from the same class in the Kiel database, and then calculate the distance pairs between raw local patches (purple identifiers) and the distance pairs between the coding coefficients (blue identifiers). From Fig. 4, we can see that the distances between coding coefficients are more similar than that of raw patches, which demonstrates the effectiveness of the constraint on coding coefficients of the local patches from the same cloud class. In addition, we also select local patches from two different classes in the Kiel database. Figure 5 shows the distance pairs between raw local patches and between the coded sparse coefficients from the different class. From Fig. 5, we can see that the distances between the coded coefficients from different classes are larger than that of the raw cases, which verifies the feasibility of the proposed algorithm again.

The experiment results of different algorithms on the Kiel database are listed in Table 1. From Table 1, several
conclusions can be drawn through analyzing the experimental results. First, our method achieves the highest classification accuracy in each case. Second, the performance of our method is over 3% higher than that of \( T_{SC} \) model in all the cases because our method considers the compactness of samples from the same class. Finally, when 1/4 samples are chosen as training data, the improvement of the proposed method over other methods is more obvious: over 7%. It demonstrates that our model can also obtain discriminative information when the training data is insufficient.

We also verify our method on the IapCAS-E database which is provided by the Institute of Atmospheric Physics, Chinese Academy of Sciences. Herein, the database is also separated into seven classes and contains 2000 ground-based cloud images. The database is more challenging because it has large intra-class variation. Figure 2(b) shows samples from different classes. The experiment results in Table 2 show that our method achieves the best results even on a challenging database.

The improvement of our method is quite reasonable. The cloud images usually possess very large intra-class variations, due to factors such as illumination, climate and deformation. This demands a more powerful algorithm to extract discriminative information from cloud images. By considering the compactness of samples from the same class, our method can obtain more discriminative information, and therefore the representation is more precise, making the classification accuracy significantly improved.

4. Conclusions

In this paper, a novel coding method named compact sparse coding is proposed for ground-based cloud classification. This model considers the compactness of samples from the same class. Specifically, we force the coding coefficients from the same class to their mean vector making them similar and more discriminative. The experimental results show that our method achieves better results than the state-of-the-art methods in ground-based cloud classification.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant No. 61401309, Natural Science Foundation of Tianjin under Grant No. 15JCQNJC01700, and Doctoral Fund of Tianjin Normal University under Grant No. 5RL134 and No. 52XB1405.

References

[1] W. Zhuo, Z. Cao, and Y. Xiao, “Cloud Classification of Ground-Based Images Using Texture–Structure Features,” J. Atmos. Ocean. Tech., vol.31, no.1, pp.79–92, 2014.
[2] L. Liu, X. Sun, T. Gao, and S. Zhao, “Comparison of Cloud Properties from Ground-Based Infrared Cloud Measurement and Visual Observations,” J. Atmos. Ocean. Tech., vol.30, no.6, pp.1171–1179, 2013.
[3] J. Calbó and J. Sabburg, “Feature extraction from wholesky ground-based images for cloud-type recognition,” J. Atmos. Ocean. Tech., vol.25, no.1, pp.3–14, 2008.
[4] M. Singh and M. Glennen, “Automated ground-based cloud recognition,” Pattern Anal. Appl., vol.8, no.3, pp.258–271, 2005.
[5] A. Isosalo, M. Turunen, and M. Pietikäinen, “Cloud characterization using local texture information,” Proc. Finnish Signal Processing Symposium, 2007.
[6] S. Liu, C. Wang, B. Xiao, Z. Zhang, and Y. Shao, “Salient local binary pattern for ground-based cloud classification," Acta Meteorologica Sinica, vol.27, no.2, pp.211–220, 2013.

[7] A. Heinle, A. Macke, and A. Srivastav, “Automatic cloud classification of whole sky images,” Atmospheric Measurement Techniques, vol.3, no.3, pp.557–567, 2010.

[8] J. Xie, L. Zhang, J. You, and D. Zhang, “Texture classification via patch-based sparse texton learning,” In Proceeding ICIP, pp.2737–2740, 2010.

[9] S. Liu, C. Wang, B. Xiao, Z. Zhang, and Y. Shao, “Soft-signed sparse coding for ground-based cloud classification,” In Proceeding ICPR, pp.2214–2217, 2012.

[10] D.L. Donoho, “For most large underdetermined systems of linear equations the minimal $l_1$-norm solution is also the sparsest solution,” Comm. Pure Appl. Math, vol.59, no.6, pp.797–829, 2004.

[11] H. Lee, A. Battle, R. Raina, and A.Y. Ng, “Efficient sparse coding algorithms,” In Proceeding NIPS, pp.801–808, 2006.

[12] C. Zhang, J. Liu, Q. Tian, C. Xu, H. Lu, and S. Ma, “Image classification by non-negative sparse coding, lowrank and sparse decomposition,” In Proceeding CVPR, pp.1673–1680, 2011.

[13] S. Gao, I.W. Tsang, L.T. Chia, and P. Zhao, “Local features are not lonely-Laplacian sparse coding for image classification,” In Proceeding CVPR, pp.3555–3561, 2010.

[14] Z. Jiang, Z. Lin, and L.S Davis, “Learning a discriminative dictionary for sparse coding via label consistent K-SVD,” In Proceeding CVPR, pp.1697–1704, 2011.

[15] L. Liu, P. Fieguth, D. Clausi, and G. Kuang, “Sorted random projections for robust rotation-invariant texture classification,” Pattern Recognit., vol.45, no.6, pp.2405–2418, 2012.