CLUSTER STRUCTURED SPARSE REPRESENTATION FOR HIGH RESOLUTION SATELLITE IMAGE CLASSIFICATION

Guofeng Sheng[†], *Wen Yang*[†], *Lei Yu*[‡], *Hong Sun*[†]

[†]School of Electronic Information, Wuhan University, Wuhan 430072, China
[‡]VisAGeS U746 Research Unit,IRISA–Campus Universitaire de Beaulieu 35042 Rennes Cedex– France

ABSTRACT

Sparse Representation based model has achieved great success for image classification. The classical approach represents each visual descriptor as a sparse weighted combination of codebook words. While offering a sparse and robust representation for each single descriptor, this method however does not ensure that similar descriptors lead to similar representations. In this paper, we present a cluster structured sparse coding (CSSC) method by unifying sparse coding and structural clustering. This approach can encourage using the same codebook words for all similar descriptors in a group, providing a discriminative representation for the task of image classification. We evaluate our method on a challenging ground truth image dataset of 21 land-use classes manually extracted from high-resolution satellite imagery. Experimental results show that structural sparse representation yields higher accuracies in classification.

Index Terms— satellite image classification, sparse coding, structural clustering, structural sparse representation

1. INTRODUCTION

Recently, codebook based models have achieved the-state-of-the-art performance on many challenging image classification benchmarks such as 15-Scenes, Caltech-101, Caltech-256, and Pascal VOC. The paper investigates codebook based models to satellite image classification and presents a cluster structured sparse coding (CSSC) method to further improve the performance.

The codebook based model is firstly introduced by the bag-of-features (BoF) method [1, 2]. This method quantizes local feature descriptors using a visual codebook typically constructed through k-means clustering. The histogram of visual words is then used to perform image classification. However, the BoF method disregards the spatial information of feature descriptors, therefore it is incapable of capturing structure, shapes or objects. Spatial Pyramid Matching (SPM) [3] was the most successful work to address this problem. Motivated by the earlier work termed pyramid matching [2], SPM partitions the image into a sequence of increasingly finer grids and then computes histograms of visual words within each grid. Yan and Newsam [4] further extended SPM to Spatial Co-occurrence Kernel (SCK), and showed promising results on land-use classification.

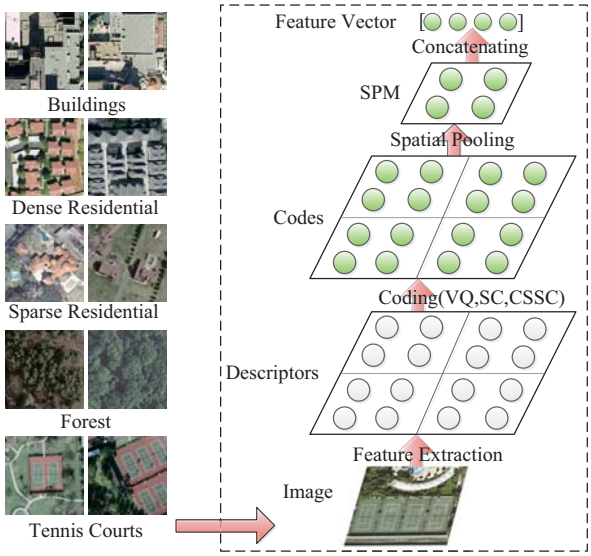

Fig. 1. Left: examples of five land-use classes. Right: flowchart of codebook based model for image classification.

Although SPM approach works well for image classification, in order to obtain good performance, SPM has to use classifier with nonlinear kernels, such as the intersection kernel or Chi-square kernel. Accordingly, the nonlinear classifier has to pay additional computational complexity, which implies a poor scalability for real application. Furthermore, the vector quantization (VQ) representation used in SPM is not so robust with respect to descriptor variability [5, 6]. To improve the scalability and alleviate the drawback of VQ , Yang et al. [5] proposed the ScSPM method where sparse coding (SC) of feature descriptors is used to replace the vector quantization (VQ). This method encourages sparsity to determine a small set of visual words from the codebook that can be used to efficiently represent each feature descriptor. The method achieved state-of-the-art performances on several natural image datasets with the linear classifier.

978-1-4673-2197-6/12/$31.00 ©2012 IEEE

In this work, we extend the sparse coding to cluster structured sparse coding (CSSC) and apply it for the image classification. The basic idea behind our CSSC approach is to ensure similar descriptors possessing similar representations. The ScSPM method considers each feature descriptor as a separate coding problem, and might select quite different codebook words for similar descriptors to favor sparsity [6, 7]. Therefore, ScSPM may lead to large variance between similar descriptors. On the other hand, previous work has shown the effectiveness of preserving locality or similarity during coding [6, 8]. To address this problem, there has been some work on structural sparse coding[9, 10], however, unlike their complex algorithms, our CSSC method is simple and efficient: descriptors are first clustered into one of K subspaces using k-means; the similar descriptors in a group are then decomposed by sparse coding over a codebook of size D simultaneously. The mixed-norm regularization is used to promote same codebook words for all descriptors within-group but different between-groups, thus making the coding coefficients more discriminative.

We evaluate the methods on a ground truth image dataset of 21 land-use classes manually extracted from publicly available high-resolution satellite imagery [4]. To the best of our knowledge, this is the first time sparse representation based classification methods are evaluated on this challenging dataset. Apart from comparing our CSSC method to standard approaches, namely BoF, SPM, SCK, ScSPM, we perform extensive evaluation of different configurations such as the numbers of subspaces for structural clustering and the sizes of codebook for sparse coding. We show that our CSSC approach results in higher classification accuracies and can work surprising well with sparse representation based classification scheme (SRC) [11], especially when the codebook size is small. A typical flowchart of codebook based image classification model and some examples of different land-use classes are shown in Fig. 1.

2. CLUSTER STRUCTURED SPARSE CODING

Let $x \in \mathbb{R}^n$ be a feature descriptor, e.g., 128 dimensional SIFT descriptor, $B = [b_1, b_2, \cdots, b_M] \in \mathbb{R}^{n \times M}$ be a codebook with M visual words, and $V = [v_1, v_2, \cdots, v_M]^T \in \mathbb{R}^M$ be the corresponding code of x over codebook B. In this section, we first review two classical coding schemes, namely, VQ and SC, and then introduce our proposed CSSC method.

2.1. Coding descriptors in vector quantization

Traditional BoF, SPM models uses VQ coding to represent x, which assigns 1 to the nearest neighbor and 0 to the others:

$$v_i = \begin{cases} 1, & \text{if } i = \arg\min_j (\|x - b_j\|_2) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

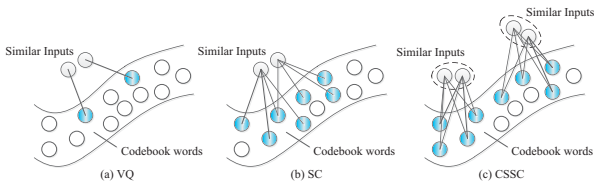

Fig. 2. Comparison between VQ, SC and CSSC representations. The atoms highlighted in blue represent the selected codebook words.

The constraint of VQ representation may be too restrictive since it only picks a single codebook word for each descriptor, giving rise to large quantization loss.

2.2. Coding descriptors in sparse coding

To ameliorate the quantization loss of VQ, ScSPM [5] relaxes the constraint by putting a l_1-norm on V instead, which enforces V to have a small set of nonzero elements. Thus, coding the descriptor x becomes a standard sparse coding (SC)[12] problem:

$$\begin{aligned} &V = \arg\min_V \|x - BV\|_2^2 + \lambda \|V\|_1 \\ &s.t. \sum_{i=1}^M v_i = 1 \end{aligned} \quad (2)$$

where $\|\cdot\|_1$ denotes the l_1-norm.

Normally, the codebook B is over-complete ($i.e., M > n$), and hence l_1 regularization is necessary to ensure that the under-determined system has a unique and sparse solution. Sparse coding can achieve a much lower quantization error than VQ, and allow the representation to capture salient properties of descriptors. However, this method considers each descriptor as a separate coding problem, and thus does not guarantee similar representations for similar descriptors.

2.3. Coding descriptors in structural sparse coding

In this paper, we present a new coding algorithm called cluster structured sparse coding (CSSC). As suggested in [6, 8], the idea of ensuring that similar descriptors will have similar representations plays an important role in the codebook based image classification models. Our CSSC method implements this idea by unifying sparse coding and structural clustering. Specifically, each descriptor is first quantized into one of K subspaces by k-means clustering, and all the descriptors within each subspace are considered to be similar. Then, similar descriptors within an image are decomposed by the following criteria simultaneously:

$$A = \arg\min_A \|X - DA\|_2^2 + \lambda \|A\|_{1,2} \quad (3)$$

where $X = [x_1, x_2, \cdots, x_P] \in \mathbb{R}^{n \times P}$ is a set of P similar descriptors, and the columns of $A = [a_1, a_2, \cdots, a_P] \in \mathbb{R}^{M \times P}$ are the codes for each descriptor. The mixed l_1/l_2

Table 1. Classification accuracies of various methods under different visual codebook sizes (D).

D	50	100	150	200	300	400	500
BoF [4]	66.81	71.86	72.81	73.62	74.43	74.81	75.76
SPM [4]	71.29	74.00	74.52	74.62	74.19	74.29	75.29
BoF+SCK [4]	68.00	72.10	74.76	75.76	75.95	76.43	76.86
SPCK+ [13]	-	76.05	-	-	-	-	-
SPCK++ [13]	-	77.38	-	-	-	-	-
ScSPM	66.76	72.71	75.23	77.67	80.26	80.71	80.90
CSSC-SVM	69.04	75.57	77.19	80.04	80.90	80.76	81.29
CSSC-SRC	**80.05**	**83.67**	**85.21**	**86.62**	**86.81**	**87.01**	**87.29**

norm of A measures reconstruction complexity, with the regularization parameter λ that balances reconstruction error (the first term) and reconstruction complexity.

CSSC decomposes a set of descriptors simultaneously, and thus can capture the correlations between similar descriptors by sharing the codebook words. The comparison between VQ, SC, CSSC representations is illustrated in Fig. 2. Due to large quantization errors, VQ representation for similar descriptors might be quite different. Each descriptor is more accurately represented by multiple codebook words in SC and CSSC. Nevertheless, as shown in Fig. 2b, the SC representation might select very different codebook words for similar descriptors to favor sparsity, thus losing correlations between codes. On the other side, the structural clustering in CSSC ensures that similar descriptors will have similar codes.

3. SPARSE REPRESENTATION BASED CLASSIFICATION SCHEME

Recently, Wright [11] et al. applied sparse coding to Face Recognition(FR) and proposed the sparse representation based classification (SRC) scheme, which achieves impressive FR performance. let y be a test sample, $A = [A_1, A_2, ..., A_k]$ be a matrix of training samples for k classes, then the SRC algorithm is as follows. 1. Solve the l_1-minimization problem:

$$\alpha = \arg\min_{\alpha} \|y - A\alpha\|_2^2 + \lambda\|\alpha\|_1 \tag{4}$$

2. Compute the residuals on every class:

$$r_i(y) = \|y - A\delta(\alpha)\|_2 \tag{5}$$

where $i = 1, ..., k$, and $\delta(\alpha)$ is a vector whose only nonzero entries are the entries in α that are associated with class i. 3. Classify y by assigning it to the class the minimizes the residual:

$$\text{identity}(y) = \arg\min_{i} \quad r_i(y) \tag{6}$$

4. EXPERIMENTAL RESULTS

In the experiments, we evaluated our method on the land-use dataset [4]. We used only a single descriptor, the Scale Invariant Feature Transform (SIFT)[14], throughout the experiments. The parameters of the subspace number (K) for structural clustering and the codebook size (D) for sparse decomposition will be discussed or stated in each experiment. In the "spatial pooling" layer, the pyramid levels were set to $0, 1, 2$ according to the best performance reported in [3]. For training the classifiers, we used nonlinear LIBSVM for BoF, SPM, BoF+SCK, SPCK+, SPCK++, linear LIBLINEAR for ScSPM, CSSC-SVM, and SRC for CSSC-SRC.

4.1. Dataset Description

The ground truth image dataset of 21 land-use classes was first evaluated in [4], and now is available to other researchers. This dataset was manually extracted from aerial orthoimagery with a pixel resolution of one foot. It consists of 21 land-use classes (agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts), with 100 images (256×256 pixels) for each class. Two examples of five classes are shown on the left of Fig. 1.

The results reported in [4, 13] are used as baselines. Following the procedure in [4, 13], five-fold cross-validation is performed. The dataset is first randomly split into five equal sets and then the classifier is trained on four of the sets and evaluated on the held-out set. The classification accuracy is the average over the five evaluations.

4.2. Results

Table 1 shows the average classification accuracies of various methods under different codebook sizes (D). For this work [13], the results are reported with only one dictionary size(100). The number of subspace (K) for structural clustering in CSSC was fixed at 10 in this experiment. As we can see, the sparse representation based approaches (ScSPM, CSSC-SVM and CSSC-SRC) outperform other approaches in most cases. In addition, CSSC works surprisingly well with simple SRC classifier, which dramatically reduces computational cost compared with SVM. The compatibility of SRC

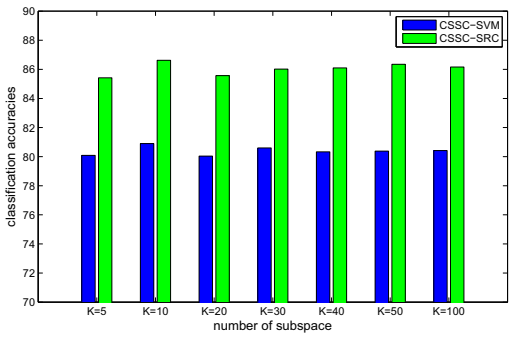

Fig. 3. Performance of CSSC under different subspace numbers (K). The size of codebook (D) was fixed at 200.

with structural sparse representation in satellite image classification is a very interesting phenomenon.

When the codebook size is small, e.g., 50, the results of ScSPM and CSSC-SVM are worse than that of SPM, because only a small number of codebook words cannot represent local feature descriptors sufficiently. But for vector quantization based method (SPM), the performance does not decrease so fast because the response of the codebook can still reflect the distribution of local descriptors.

Besides, our CSSC-SVM method improves performance over ScSPM for all codebook size, indicating the importance of preserving the correlations between similar descriptors during coding. The improvement is much more evident in the case of smaller codebook sizes $(50, 100, 150, 200)$. This is significant from a computational viewpoint since larger codebook size means heavier computational cost during sparse coding.

4.3. Discussion

To provide more comprehensive analysis of the proposed CSSC method, we further evaluated its performance with respect to subspace number (K).

Fig. 3 shows the performance when $K = 5, 10, 20, 30, 40, 50, 100$ respectively under codebook size (D) of 200. The experimental results indicate there is no substantial differences between 5 and 100 subspace numbers. When $K = 10$, our method performs best, thus we fixed K at 10 in the previous experiments.

5. CONCLUSION

This paper presents a promising image representation method called cluster structured sparse coding (CSSC) by unifying sparse coding and structural coding. CSSC can capture the correlations between similar descriptors by sharing the same codebook words. Experiments on 21 land-use class dataset show that CSSC gives superior image classification performance with sparse representation based classification scheme.

6. REFERENCES

[1] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[2] K. Grauman and T. Darrell, "Pyramid match kernels: Discriminative classification with sets of image features," in *ICCV*, vol. 2, Oct. 2005, pp. 1458–1465.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, vol. 2, 2006, pp. 2169–2178.

[4] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *ACM SIGSPATIAL GIS*, California, America, Nov. 2010.

[5] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *CVPR*, Jun. 2009, pp. 1794–1801.

[6] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *CVPR*, Jun. 2010, pp. 3360–3367.

[7] S. Bengio, F. Pereira, Y. Ponce, and G. Singer, "Group sparse coding," in *NIPS*, Dec. 2009.

[8] Y. Boureau, N. L. Roux, F. Bach, J. Ponce, , and Y. LeCun, "Locality-constrained linear coding for image classification," in *ICCV*, Nov. 2011.

[9] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *ICML*, Jun. 2009.

[10] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *CVPR*, Jun. 2010, pp. 3501–3508.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," in *TPAMI*, vol. 31, no. 2, 2009, pp. 210–227.

[12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *ICML*, Jun. 2009.

[13] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification," in *ICCV*, 2011, pp. 1465–1472.

[14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.