Discriminative Tensor Sparse Coding for Image Classification

Yangmuzi Zhang[1]
ymzhang@umiacs.umd.edu
Zhuolin Jiang[2]
zhuolin.jiang@huawei.com
Larry S. Davis[1]
lsd@umiacs.umd.edu

[1] University of Maryland, College Park, USA
[2] Noah's Ark Lab, Huawei Technologies

Sparse models have been successfully applied to many problems in image processing, computer vision, and machine learning. Many algorithms [3, 12] have been proposed to learn an over-complete and compact dictionary based on such models. In general, the input feature representations to these approaches are based on traditional vector descriptors. As pointed out in recent work [2, 7], vectorizing the original data structure, however, may destroy some inherent ordering information in the data. In computer vision, the region covariance feature, introduced in [9], is an image descriptor that captures natural correlations amongst multiple features. Hence, there has been growing interest in the development of sparse coding for positive definite descriptors. In [10], the problem of sparse coding within the space of symmetric positive definite matrices is tackled by embedding Riemannian manifolds into kernel Hilbert spaces. [7] proposed tensor sparse coding on positive definite matrices, which keeps descriptors in their original space and uses a set of randomly selected training samples as the dictionary. It successfully extended sparse coding techniques to the space of positive definite matrices. However, little research has been done to learn a discriminative and compact dictionary over such spaces.

We present a discriminative dictionary learning method for tensor sparse coding. Rather than simply using a subset of region covariance descriptors for training images as the dictionary [7], we learn a discriminative dictionary from the training set. A structural incoherence term is introduced into the dictionary learning process to regularize the incoherence between different sub-dictionaries, which increases the discriminativeness of the learned dictionary. We further incorporate classification error into the objective function to make the learned dictionary effective for classification tasks. Instead of learning multiple classifiers for each pair of classes [4, 5, 11], a linear multi-class classifier can be easily obtained during the training process. Unlike [8], which focuses on the reconstructive capability of a dictionary, the dictionary learned by our approach has both good reconstruction and discrimination capabilities. Based on this learned high-quality dictionary, we are able to obtain discriminative tensor sparse representations. Classification can be efficiently performed on these representations using the learned multi-classifier as it only involves matrix multiplication.

The quality of the dictionary influences the discriminativeness of the tensor sparse representations. Updating each dictionary atoms separately does not result in sufficient discriminating information in the sub-dictionaries. Following [1, 6], we introduce structural incoherence into sub-dictionary atoms. Incoherence will promote dictionary atoms from different classes to be independent from each other; thus it leads to sparse and discriminating representations for images. Given a training data set $S = \{S_l\}_{l=1}^{N}$, we will learn a dictionary $A = \{A^i\}_{i=1}^{K}$, with sub-dictionary $A^i = [\mathbf{a}_1^i, \mathbf{a}_2^i, ..., \mathbf{a}_{K_i}^i]$ for class i. The problem is formulated as:

$$
\begin{aligned}
min_{A,X} \quad & \Sigma_{l=1}^{N} D_{ld}(\mathbf{x}_l \otimes A, S_l) + \lambda ||\mathbf{x}||_1 + \eta \Sigma_{i \neq j,s,t} ||(\mathbf{a}_s^j)^T \mathbf{a}_t^i||_F^2 \qquad (1)\\
s.t. \quad & \mathbf{x}_l \geq \mathbf{0} \quad \forall l \\
& \mathbf{a}_t^i,\ \mathbf{a}_s^j \succeq 0 \quad \forall i,t,j,s \\
& 0 \preceq \mathbf{x}_l \otimes A \preceq S_l \quad \forall l
\end{aligned}
$$

The first two terms are the reconstruction error and the sparsity regularization. The last term sums up the Frobenius norms between every two dictionary atoms $\mathbf{a}_s^j, \mathbf{a}_t^i$ which belong to different sub-dictionaries A^j and A^i. λ, η are penalty parameters balancing reconstruction error, sparsity, and dictionary structural incoherence.

To make the learned dictionary to be more adaptive to classification tasks, we minimize the classification error in the objective function of dictionary learning as pointed out in [3]. A linear multi-classifier $f(x;W) = Wx$ is used for classification. W denotes the linear classifier's parameters. Hence, the classification error can be explicitly included in the objective function during the dictionary learning. The classifier will be learned through the training process, as well. The objective function is formulated as below:

$$
\begin{aligned}
min_{A,X,W} \quad & \Sigma_{l=1}^{N} D_{ld}(\mathbf{x}_l \otimes A, S_l) + \lambda ||\mathbf{x}||_1 + \eta \Sigma_{i \neq j,s,t} ||(\mathbf{a}_s^j)^T \mathbf{a}_t^i||_F^2 \\
& + \gamma ||H - WX||_2^2 \qquad (2)\\
s.t. \quad & \mathbf{x}_l \geq \mathbf{0} \quad \forall l \\
& \mathbf{a}_t^i,\ \mathbf{a}_s^j \succeq 0 \quad \forall i,t,j,s \\
& 0 \preceq \mathbf{x}_l \otimes A \preceq S_l \quad \forall l
\end{aligned}
$$

where the term $||H - WX||_2^2$ represents the classification error. $H = [h_1, h_2, ..., h_N] \in R^{K \times N}$ denotes the label matrix. The column vector $h_i = [0, 0, ...1...0, 0]^T \in R^K$ is a label vector for sample i. The position of 1 indicates its class index. γ controls the contribution of the classification error regularization term in the optimization process.

We have tested our approach on three types of public datasets, including texture, digit, and face database. Experiment results demonstrate the effectiveness of our approach.

[1] C. Chen, C. Wei, and Y. Wang. Low-rank matrix recovery with structural incoherence for robust face recognition, 2012. *CVPR*.

[2] T. Hazan, S. Polak, and A. Shashua. Sparse image coding using a 3d non-negative tensor factorization, 2005. *ICCV*.

[3] Z. Jiang, Z. Lin, and L. Davis. Learning a discriminative dictioanry for sparse coding via label consistent k-svd, 2011. *CVPR*.

[4] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis, 2008. *CVPR*.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Supervised dictionary learning, 2009. *NIPS*.

[6] I. Ramirez, P. Sprechmann, and G. Sapiro. Classification and clustering via dictionary learning with structural incoherence and shared features, 2010. *CVPR*.

[7] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos. Tensor sparse coding for region covariances, 2010. *ECCV*.

[8] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos. Positive definite dictionary learning for region covariances, 2011. *ICCV*.

[9] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification, 2006. *ECCV*.

[10] O. Tuzel, F. Porikli, and P. Meer. Human detectio via classification on riemannian manifolds, 2007. *CVPR*.

[11] J. Yang, K. Yu, and T. Huang. Supervised translation-invariant sparse coding, 2010. *CVPR*.

[12] Y. Zhang, Z. Jiang, and L. Davis. Learning structured low-rank representations for image classification, 2013. *CVPR*.