MULTISCALE DICTIONARY LEARNING FOR HIERARCHICAL SPARSE REPRESENTATION

Yangmei Shen, Hongkai Xiong

Department of Electronic Engineering
Shanghai Jiao Tong University, China
{shenyangmei0214, xionghongkai}@sjtu.edu.cn

Wenrui Dai

Department of Biomedical Informatics
University of California, USA
wed004@ucsd.edu

ABSTRACT

In this paper, we propose a multiscale dictionary learning framework for hierarchical sparse representation of natural images. The proposed framework leverages an adaptive quadtree decomposition to represent structured sparsity in different scales. In dictionary learning, a tree-structured regularized optimization is formulated to distinguish and represent high-frequency details based on varying local statistics and group low-frequency components for local smoothness and structural consistency. In comparison to traditional proximal gradient method, block-coordinate descent is adopted to improve the efficiency of dictionary learning with a guarantee of recovery performance. The proposed framework enables hierarchical sparse representation by naturally organizing the trained dictionary atoms in a prespecified arborescent structure with descending scales from root to leaves. Consequently, the approximation of high-frequency details can be improved with progressive refinement from coarser to finer scales. Employed into image denoising, the proposed framework is demonstrated to be competitive with the state-of-the-art methods in terms of objective and visual restoration quality.

Index Terms— dictionary learning, multiscale representation , structured sparsity, hierarchical structure, image denoising

1. INTRODUCTION

Sparse representation over redundant dictionary is a powerful model to adapt real-world signals, which is validated by well-established theoretical frameworks and state-of-the-art empirical results [1]. Its basic assumption suggests that a natural signal $\mathbf{x} \in \mathbb{R}^m$ is approximately represented by a *sparse* linear combination of atoms selected from an *overcomplete dictionary* $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_p] \in \mathbb{R}^{m\times p}$ $(m < p)$, with the corresponding sparse representation vector $\alpha \in \mathbb{R}^p$. In general, a sparse coding problem is formulated to derive the sparse representation model.

$$\min_{\alpha\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda\|\alpha\|_1, \tag{1}$$

where λ is a regularization parameter balancing fidelity and sparsity, and $\|\alpha\|_1$ is a sparsity-inducing norm leading to the well-known Lasso or basis pursuit problems.

In comparison to pre-defined analytical dictionaries e.g., wavelets, the trained dictionaries can significantly improve the approximation performance for natural images by capturing the varying structures [2, 3]. ℓ_1-regularized optimization was formulated to adaptively derive the dictionaries from the sampled signals with batch gradient descent. These batch procedures would fail for large-scale high-dimensional signals, due to high computational complexity and low convergence speed. Thus, *online* dictionary learning methods have been widely concerned to achieve faster convergence with guaranteed accuracy [4, 5]. However, ℓ_1-regularized optimization is still restricted for sparse representation of multiscale high-dimensional signals like images and videos, as it independently generates the atoms by ignoring their structural relationship [6, 7].

To sufficiently exploit prior knowledge, *structured sparsity* methods were developed to adopt sparsity-inducing regularization capable for the higher-order information about the patterns of nonzero coefficients. One such possibility is the search for group dictionaries, where group structures of bags of visual descriptors at image level are considered for image classification [8]. Another alternative has been the pursuit of hierarchical dictionaries, which involve a tree-structured group-Lasso penalty addressed efficiently by dedicated proximal methods [9]. Inspired by independent component analysis, [10] goes beyond one-dimensional patterns and puts a 2-D grid structure on decomposition coefficients to infer topographic dictionaries by network flow optimization. Since all the above algorithms work off-line, [11] develops an online structured learning scheme using variational methods, making it possible to efficiently process large and partially observable training data. However, all of these structured dictio-

The work was supported in part by the NSFC grants 61501294, 61622112, 61529101, 61472234, 61425011, China Postdoctoral Science Foundation 2015M581617 and Program of Shanghai Academic Research Leader 17XD1401900.

978-1-5090-6067-2/17/$31.00 ©2017 IEEE

naries have been traditionally restricted to fixed atom scales, which is insufficient to characterize the diverse and complex natural phenomena.

On the other hand, multiscale dictionaries have been considered to take advantage of multiscale property and data matching capability. Mairal *et al*. [12] fully decomposed images through a quadtree structure and learn multiple sub-dictionaries from patches with different scales using optimization methods like K-SVD. Modeling dictionary as a multiplication of Discrete Cosine Transform by a learned sparse matrix, the double-sparsity formulation made the first successful attempt towards the harmonic analysis [13]. In the context of Wavelet, learning process was applied into the analysis domain of Wavelet decomposition, where separate sub-dictionaries at different bands are trained by K-SVD [14]. Recently, Sulam *et al*. [15] extended the double-sparsity model by replacing the DCT dictionary with a new cropped Wavelet decomposition, which enabled dictionary learning to be up-scaled to a relative higher dimension. However, multiscale dictionaries would degrade the performance of sparse representation without considering the underlying dependencies or hidden structures between dictionary atoms.

In this paper, we develop a multiscale dictionary learning framework to enable hierarchical sparse representation, which naturally organizes atoms in a hierarchy with a descending order for node-sizes and increasing frequencies from root to leaves. Adaptive quadtree decomposition is proposed to recursively partition the images based on local statistics, which groups low-frequency components into large patches and distinguishes high-frequency details in small ones. A hierarchical regularized optimization is formulated to enforce sparsity patterns with rooted and connected subtrees. Thus, effective decomposition of image content is achieved using atoms from different scales. To improve learning efficiency, a joint hierarchical sparse coding step by proximal gradient method and a separate multiscale atom update procedure via block-coordinate descent are alternately performed. The learned dictionary is able to make sparse representation based on patches across multiple scales under a constraint of the tree-structured prior for nonzero patterns. For signal approximation, large atoms near the root provide low-frequency components, whereas fine details are hierarchically refined by small atoms in finer scales. In a nut-shell, the proposed framework can effectively represent multiscale signals with hierarchical trained dictionaries from sampled signals with multiple scales constrained by tree-structured sparsity. To validate the efficacy, the proposed framework was employed into image denoising task. Experimental results show that it is competitive with the state-of-the-art methods and allows practical applications to take a more global outlook over the diversity of real world signals.

The rest of this paper is organized as follows. Section 2 presents the proposed framework of multiscale dictionary learning for hierarchical sparse representation. Experimental results are shown in Section 3 for validation. Finally, we conclude the contributions in Section 4.

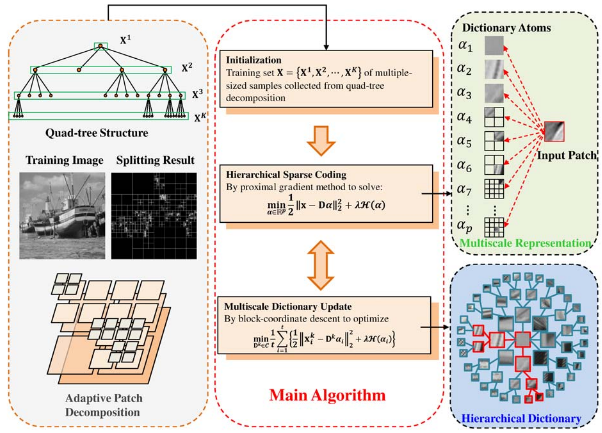

Fig. 1. The proposed framework for dictionary learning.

2. LEARNING MULTISCALE DICTIONARY FOR HIERARCHICAL SPARSE REPRESENTATION

It is widely known that natural image information spreads across multiple scales. Depending on specific structures, different images prefer different patch sizes for optimal representations. As depicts in Fig. 1, we present an attempt to explicitly exploit multiple scales simultaneously: using an efficient quadtree (QT) decomposition, an input image is recursively split into quadrants up to the selected sizes based on local features; by alternating between a hierarchical sparse coding step and a multiscale dictionary update procedure, dictionary is learned to sparsify and finely adapted to the training data; besides, a tree-structured sparsity prior is enforced to organize the learned atoms in a prespecified dendriform fashion, with larger atoms close to the root whereas the smaller near the leaves.

2.1. Adaptive Quadtree Decomposition

To achieve a variable size partition while avoiding the cost of more sophisticated techniques, an efficient quadtree decomposition is employed due to its effective balance between adaptivity of segmentation and simplicity of implementation. As for our setting, the primary purpose is to isolate the high detail regions into small sizes while grouping the low frequency regions into patches as large as possible, expecting to enhance the potential expressive force of the dictionary. Inspired by [16], local residual mean and variance values are jointly used as a simple yet effective measurement to assess the amount of details in a patch.

Given an input image, it is broken into fully overlapping patches of $\sqrt{m} \times \sqrt{m}$ pixels which are treated as independent