Engineering Applications of Artificial Intelligence 25 (2012) 1259–1264

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Evolution-enhanced multiscale overcomplete dictionaries learning for image denoising

Shuyuan Yang [a,*], Min Wang [b], Meirong Wei [a], Licheng Jiao [a]

[a] Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China
[b] National Key Lab of Radar Signal Processing, Department of Electrical Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 7 May 2011
Received in revised form
29 November 2011
Accepted 24 January 2012
Available online 20 February 2012

Keywords:
Multiscale overcomplete dictionaries learning
Evolution-enhanced
Image denoising

ABSTRACT

In this paper, a multiscale overcomplete dictionary learning approach is proposed for image denoising by exploiting the multiscale property and sparse representation of images. The images are firstly sparsely represented by a translation invariant dictionary and then the coefficients are denoised using some learned multiscale dictionaries. Dictionaries learning can be reduced to a non-convex l_0-norm minimization problem with multiple variables, so an evolution-enhanced algorithm is proposed to alternately optimize the variables. Some experiments are taken on comparing the performance of our proposed method with its counterparts on some benchmark natural images, and the superiorities of our proposed method to its counterparts can be observed in both the visual result and some numerical guidelines.

1. Introduction

Overcomplete (or redundancy) is important in transformation-based image denoising methods to have the shift invariance property (Coifman and Donoho, 1994; Bui and Chen, 1998). For example, with the growing realization of deficiencies of orthogonal wavelets in denoising images, some redundant multiscale transforms have been introduced, including undecimated Wavelets (Lang et al., 1996), Curvelet (Starck et al., 2002), Contourlet (Do and Vetterli, 2002), Wedgelet (Demaret et al., 2005), Bandelet (Zhang et al., 2010), Shearlet (Blu and Luisier, 2007) and so on. In the past decade, using spatial overcomplete representation and sparsity for images denoising has drawn much attention of researches (Elad et al., 2006; Elad and Aharon, 2006; Elad and Aharon, 2006). Its basic idea is that the sparse representation (SR) of images will help in automatically selecting the primary components in images while reducing the noise components, as long as the dictionary can well describe the characteristics of images. In more recent works (Aharon et al., 2006; Protter and Elad, 2009; Chatterjee and Milanfar, 2009; Turek et al., 2010; Dong et al., 2011), image patches prove to well represent the statistical properties of the whole image, so a large number of image patches are taken as the examples from which a dictionary can be learned. The patches are taken from the noisy image and then sparsely represented and restored, which lead to state-of-the-art denoising result.

Although the SR-based denoising methods have proved to work well on the natural images (Aharon et al., 2006; Protter and Elad, 2009; Chatterjee and Milanfar, 2009; Turek et al., 2010; Dong et al., 2011), the SR is all executed in the spatial domain. On the other hand, it has been aware that making avail of the multiscale properties of images will obtain better denoising result (Bui and Chen, 1998; Lang et al., 1996; Starck et al., 2002; Do and Vetterli, 2002; Demaret et al., 2005; Zhang et al., 2010; Blu and Luisier, 2007). Therefore, in this study we take both the overcomplete representations of images in spatial domain and transformation domain into account, and propose a multiscale dictionaries learning approach for image denoising. Some multiscale overcomplete dictionaries are learned from example patches, and then used to reduce the noise distributed in the different scales of images. We reduce the image denoising to an l_0-norm minimization problem with multiple variables. The available optimization schemes for this NP-hard problem can be mainly divided into two catalogs: approximation method and relaxation method. The approximation method includes greedy algorithms and shrinkage algorithms. A greedy strategy abandons exhaustive search in favor of a series of locally optimal single-term updates. Its basic idea is to represent a signal as a weighted sum of atoms taken from a dictionary, such as matching pursuit (Mallat and Zhang, 1993), orthogonal matching pursuit (Tropp and Gilbert, 2007) and their variants (Donoho et al., 2006; Needell, 2009). The approximation method can correctly pick up atoms in the case of existing sparse solution and the selection rule is simple to understand. However, it is characteristics of heavy computation, slow convergence, and can only work well in the noiseless case. The shrinkage strategy iterates between shrinkage/thresholding operation and projection onto

* Corresponding author.
E-mail address: syyang@xidian.edu.cn (S. Yang).

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2012.01.021

perfect reconstruction, so they are characteristic of low computation complexity (Chen et al., 2001; Bioucas-Dias and Figueiredo, 2007). However, they commonly require much iteration when the support of the solutions cannot be determined. The relaxation method includes l_1-norm and l_p-norm methods. Basis pursuit (BP) (Blumensath and Davies, 2008) approximate the solution that minimizes l_1 -norm and reduce the problem to a linear programming (LP) structure, which is solvable comparing to the l_0-norm minimization and is easy to be integrated into other variational model. However, it is a difficult optimization task and the tuning of parameter is not straightforward. Moreover, the equivalence of l_0-norm and l_1-norm minimization can only achieve under very strict assumption of the sparsity of signals (Candès and Wakin, 2008). Gradient based methods are discussed in paper (Figueiredo et al., 2007) and (Blumensath and Davies, 2008) to solve this problem. The l_p-norm ($0<p<1$) or the weak l_p-norm is a popular measure of sparsity used by the mathematical analysis community, so it is used to serve as a candidate function for l_0-norm (Candès et al., 2008). Although it is still non-convex, it is almost equivalent to l_0-norm and can be represented as a weighted l_1-norm form by the iterative-reweighed-least-squares (IRLS) method. However, this algorithm is very sensitive to the initialization of solution. Moreover, it is guaranteed to converge to a fixed-point that is not necessarily the optimal one.

Evolutionary algorithms (EAs) provide a general and global searching approach for solving combinational and NP-hard optimization tasks(De Jong, 2006), so in this paper we use EAs to solve the l_0-norm minimization problem discussed above. Genetic Algorithm (GA) is one of the effective EAs that simulate natural evolution (crossover, mutation and selection) over populations of candidate solution (Goldberg, 1989). However, GA is characteristic of slow convergence (Alberto and Carlos, 2003). The memetic algorithm (MA) (Alberto and Carlos, 2003; Badillo et al., 2011; Amaya et al., 2010; Krasnogor and Smith 2005) makes an improvement on GA by combining GA with a local searching operation, and proves to perform much better than GA in terms of the quality of solution and computational cost. In the MA, GA is used for coarse search, while the subsequent local improvement is then used to refine the GA. Its superior performance over GA has been found for various applications, such as combinatorial optimization problems (Tang et al., 2007), control design (Caponio et al., 2007), VLSI design (Tang and Yao, 2007), image segmentation (Jiao et al., 2010) and so on.

In order to tune multiple variables in the optimization problem discussed above, in our study a MA based alternate optimization strategy is employed to optimize the dictionaries and denoise the multiscale images. In the algorithm, two-dimensional individual is adopted to represent a dictionary. The individuals are used to perform a global search, followed by a local search operator, singular value decomposition (SVD), to further reduce the objective function. The dictionaries and sparse coefficients are alternately updated until the stop condition is satisfied. Some experiments are taken on some benchmark natural images to investigate the performance of our proposed method.

The rest of this paper is organized as follows. In Section 2, we addressed the classic image denoising problem, and depicted the memetic algorithm-enhanced multiscale dictionaries learning algorithm. In Section 3, some simulation experiments are taken to illustrate the efficiency and superiority of our proposed method to its counterparts. Finally some conclusions are drawn in Section 4.

2. Evolution-enhanced multiscale dictionaries learning

Considering the classic image denoising problem: an image is measured in the presence of an additive zero-mean white and homogeneous Gaussian noise, with standard deviation σ. Thus the measured image is $Y=X+n$, and the goal of image denoising is to recover the clean image X from the noisy image Y.

2.1. A. Multiscale overcomplete dictionaries learning

Consider a redundant transformation on the noisy image, and denote $\Re$ as the transformation dictionary. The noisy image can be written as,

$$\mathbf{Y}=\Re\boldsymbol{\beta}'=\Re\boldsymbol{\beta}+n \tag{1}$$

where $\boldsymbol{\beta}'$and $\boldsymbol{\beta}$ are the transformation coefficients of the noisy image Y and clean image X, respectively. When the transformation has the redundancy and multiscale property, $\Re$ is often determined by the frame theory, such as undecimated wavelets frame, or some frame composed by cascade orthogonal bases.

Considering the multiscale property of $\Re$, we can reformulate (1) as,

$$\mathbf{Y}=\Re\boldsymbol{\beta}'=[\mathbf{R}_1,\ldots,\mathbf{R}_N]\times[\boldsymbol{\beta}'_1,\ldots,\boldsymbol{\beta}'_N]^T=[\mathbf{R}_1,\ldots,\mathbf{R}_N]\times[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N]^T+n \tag{2}$$

where $\mathbf{R}_1,\ldots,\mathbf{R}_N$ are the N multiscale dictionaries and $\boldsymbol{\beta}'_1,\ldots,\boldsymbol{\beta}'_N$ are the corresponding coefficients. Inspired by the example-based denoising scheme (Chatterjee and Milanfar, 2009), we extract the patches from the multiscale coefficient images $\boldsymbol{\beta}'_j(j=1,\ldots,N)$ (the patches are processed in raster-scan order in $\boldsymbol{\beta}'_j$, from left to right and top to bottom). Then we use these patches to train a dictionary that is representative of all the image patches and used to recover β_j from $\boldsymbol{\beta}'_j$.

Let $\beta_{i,j}\in R^p$ denote the ith$(p)^{1/2}\times(p)^{1/2}$ local patch vector extracted from the multiscale coefficients matrix β_j at the spatial location i: $\beta_{i,j}=P_i\beta_j$, where P_i denotes a rectangular windowing operator and the overlapping is allowed. $\beta_{i,j}$ can represent the ith patch in β_j with its coefficients being ordered lexicographically as column vector. Assume each patch vector $\beta_{i,j}$ belongs to the *Sparseland* signal (Elad, 2010), i.e., $\beta_{i,j}$ can be represented sparsely under a redundant dictionary $D_j\in R^{p\times K}$ that contains K prototype signal-atoms for columns $\{\tilde{d}_{m,j}\}_{m=1}^{K}$, that is, $\beta_{i,j}=D_j\alpha_{i,j}$, $\alpha_{i,j}\in R^K$ and $\|\alpha_{i,j}\|_0\ll K$, the MAP estimator for denoising this coefficient patch is built by solving (Elad, 2010),

$$\{\hat{\alpha}_{i,j}\}=\begin{cases}\min\limits_{\{\alpha_{i,j},\mathbf{D}_j,\boldsymbol{\beta}_j,\mathbf{R}_i\}}\sum\limits_{j=1}^{N}\sum\limits_{i=1}^{Q_j}\|\alpha_{i,j}\|_0\\ \text{s.t.}\sum\limits_{j=1}^{N}\sum\limits_{i=1}^{Q_j}\|\mathbf{D}_j\alpha_{i,j}-\mathbf{P}_i\boldsymbol{\beta}_j\|_2^2\le\varepsilon;\\ \left\|\sum\limits_{j=1}^{N}\mathbf{R}_j\boldsymbol{\beta}_j-\mathbf{Y}\right\|_2^2\le\delta\end{cases} \tag{3}$$

where P_i is a patch extraction operator, Q_j is the number of patches extracted from β_j, and ε and δ are dictated by σ.

Assume the number of the example patches at each scale take the same value: $Q_1=\ldots=Q_N=Q$, and denote the sparse coefficients of patches in β_j under the jth multiscale dictionary $D_j\in R^{p\times K}$ as $\alpha_j=[\alpha_{1,j},\alpha_{2,j},\ldots\alpha_{Q,j}]\in R^{K\times Q}$. Let $\Phi=[P_1,P_2,\ldots,P_Q]$, the optimization problem can then be reduced to,

$$\begin{cases}\min\limits_{\{\boldsymbol{\alpha}_j,\mathbf{D}_j,\Re,\boldsymbol{\beta}\}}\sum\limits_{j=1}^{N}\|\boldsymbol{\alpha}_j\|_{0,1}\\ \text{s.t.}\sum\limits_{j=1}^{N}\|\mathbf{D}_j\boldsymbol{\alpha}_j-\boldsymbol{\Phi}\boldsymbol{\beta}_j\|_2^2\le\varepsilon\\ \|\Re\boldsymbol{\beta}-\mathbf{Y}\|_2^2\le\delta\end{cases} \tag{4}$$

where $\|\boldsymbol{\alpha}_j\|_{0,1}$ is the l_0/l_1—*norm* of the matrix α_j and represent the