Digital Signal Processing 47 (2015) 157–168

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Variational Bayesian image restoration with group-sparse modeling of wavelet coefficients

Ganchi Zhang *, Nick Kingsbury

Signal Processing Group, Dept. of Engineering, University of Cambridge, United Kingdom

ARTICLE INFO

Article history:
Available online 7 May 2015

Keywords:
Image restoration
Wavelet group-sparse modeling
Variational Bayesian inference
Majorization minimization
Dual-tree complex wavelets

ABSTRACT

In this work, we present a recent wavelet-based image restoration framework based on a group-sparse Gaussian scale mixture model. A hierarchical Bayesian estimation is derived using a combination of variational Bayesian inference and a subband-adaptive majorization–minimization method that simplifies computation of the posterior distribution. We show that both of these iterative methods can converge together without needing nested loops, and thus good solutions can be found rapidly in the non-convex search space. We also integrate our method, variational Bayesian with majorization minimization (VBMM), with tree-structured modeling of the wavelet coefficients. This extension achieves significant gains in performance over the coefficient-sparse version of the algorithm. The experimental results demonstrate that the proposed method and its tree-structured extensions are effective for various imaging applications such as image deconvolution, image superresolution and compressive sensing magnetic resonance imaging (MRI) reconstruction, and that they outperform more conventional sparsity-inducing methods based on the l_1-norm.

1. Introduction

Linear inverse problems appear often in many applications of image processing such as restoration, motion estimation, reconstruction and segmentation, where a noisy indirect observation $\mathbf{y}$, of an original image $\mathbf{x}$, is modeled as [1,2]

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{B}$ of size $M \times N$ is the matrix representation of a direct linear operator and $\mathbf{n}$ is usually additive Gaussian noise with variance ν^2.

In many scenarios, this inverse problem is highly ill-posed, i.e. the direct operator does not have an inverse or it is nearly singular so that its inverse is very sensitive to noise [3]. Thus it can only be solved satisfactorily by incorporating some regularization techniques, often using Bayesian inference with prior information [4]. In previous works, it is found that wavelet-based tools, such as the Discrete Wavelet Transform (DWT), are powerful for modeling this prior knowledge [4–6].

In the past two decades, the DWT has been exploited for a wide range of signal processing applications such as denoising, deconvolution, superresolution, compression and classification (see, e.g., [7–11]). The DWT provides an efficient implementation based on a filter bank structure utilizing decimation and two discrete filters, a low-pass and a high-pass filter [12]. Wavelet-based regularization methods are good for image restoration problems because wavelet coefficients tend to be sparse for most image types.

Although the DWT is compact, it suffers from shift dependency, lack of directionality, oscillation and aliasing [13]. These will significantly constrain the performance of a DWT-based signal processing system. To solve these shortcomings, the dual-tree complex wavelet transform (DT $\mathbb{C}$WT) first proposed by Kingsbury, is a recent simple and efficient redundant transform that has been widely used in solving diverse signal processing problems. The DT $\mathbb{C}$WT is better than the DWT for image restoration problems due to the fact that directional filters encourage greater sparsity and complex coefficients show more consistent persistence across scale. Other recent extensions of the DWT, such as curvelets [14] and contourlets [15], would also work in this context but few, if any, combine the efficiency and good performance of the dual-tree approach.

It is known that the wavelet coefficients of natural images display non-Gaussian statistics and their marginal distributions typically show a large peak at zero with long heavy tails [16,17]. To account for this non-Gaussian behavior, many univariate parametric models such as generalized Laplacian distributions [17] and Bessel K form density models [18] have been previously used to

* Corresponding author.
E-mail addresses: gz243@cam.ac.uk (G. Zhang), ngk@eng.cam.ac.uk (N. Kingsbury).

http://dx.doi.org/10.1016/j.dsp.2015.04.011
1051-2004/

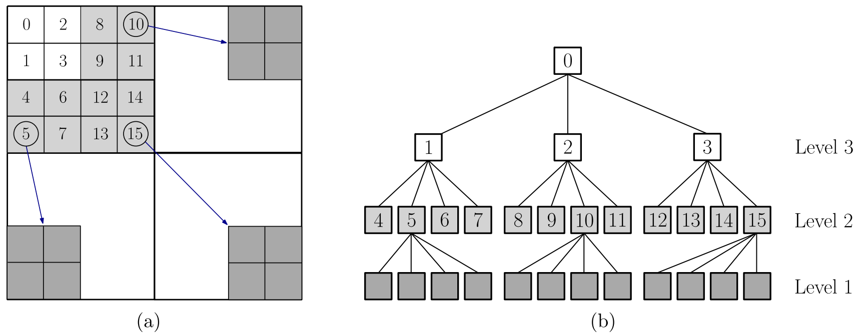

Fig. 1. (a) 8×8 image with 3-level 2D DWT decomposition. (b) Quadtree structure of wavelet coefficients.

model the wavelet coefficients. However, these models do not consider the persistence across scales of wavelet coefficients [19]. In fact, the energies of wavelet coefficients of natural images exhibit a strong characteristic signal-dependent structure. Fig. 1 depicts an example of quadtree structure that corresponds to an 8×8 image with 3-level 2D DWT decomposition. To well capture the statistical dependencies, bivariate shrinkage [20], Hidden Markov Tree models (HMM) [21,22] and Gaussian Scale Mixture Models (GSM) [16,23] have been widely applied to model wavelet coefficients whose energies are not randomly distributed. Among those methods, it is acknowledged that the GSM model can be used in the framework of sparse Bayesian learning (SBL) where the sparsity is obtained by reweighting the Gaussian prior [24,25]. Based on this connection, several researchers have shown that Bayesian methods are applicable for wavelet-based regularization problems [4,5,16].

Recently, Bayesian group-sparse (or block sparse) modeling has emerged where the sparsity is imposed on groups instead of individual components [27,28]. In [28], variational Bayesian (VB) inference is used for group-sparse modeling and has been shown to find sparse solutions effectively. These approaches can potentially be used in the wavelet domain since a pair of coefficients at a certain location and adjacent scales are typically both large or both small in amplitude [29]. Tree-structure existing in the wavelet domain allows group-sparse models to be easily constructed and used. One of the major contributions of our work is to investigate the use of Bayesian group-sparse modeling for wavelet-based regularization problems.

In [26], we proposed a hierarchical Bayesian modeling of wavelet coefficients derived from a group-sparse GSM model. Based on a combination of VB inference with a subband-adaptive majorization minimization (MM) method, the VBMM method in [26] effectively simplifies computation of the posterior distribution and finds good solutions in the non-convex search space. In addition, the VBMM method has also shown good potential with group-sparse modeling. In [30], we incorporate the VBMM method with a wavelet tree structure based on overlapped groups, which leads to an improved solution compared with unstructured coefficient-sparse modeling.

In this paper, we extend the ideas from [26] to generalize the VBMM method and discuss the theoretical foundations in some detail. Different from [26] and [30], we also include the results of image superresolution and MRI image reconstruction. The proposed method can handle very large data sets with a good performance and low computation cost. The paper is organized as follows. Section 2 describes our proposed VBMM image restoration framework. Section 3 discusses the tree-structured extensions of VBMM. Experimental results are shown in Section 4. Conclusions are provided in Section 5.

2. VBMM image restoration

In this section, we describe our proposed VBMM image restoration framework and its tree-structured extensions.

2.1. Model formulations

To obtain a wavelet-based formulation, we note that the image $\mathbf{x}$ can be represented by wavelet expansion as $\mathbf{x} = \mathbf{M}\mathbf{w}$ where $\mathbf{w}$ is an $N \times 1$ vector representing all wavelet coefficients, and $\mathbf{M}$ is the inverse wavelet transform whose columns are the wavelet basis functions. In the case of an orthogonal basis, $\mathbf{M}$ is a square orthogonal matrix, whereas for an over-complete dictionary (e.g. a tight frame), $\mathbf{M}$ has N columns and M rows, with $N > M$ [6]. The linear model in (1) then becomes

$$\mathbf{y} = \mathbf{B}\mathbf{M}\mathbf{w} + \mathbf{n} \tag{2}$$

and the resulting likelihood of the data assuming Gaussian noise $\mathbf{n}$ can be shown to be

$$p(\mathbf{y}|\mathbf{w}, \nu^2) = \left(2\pi\nu^2\right)^{-\frac{M}{2}} \exp\{-\frac{1}{2\nu^2}\|\mathbf{y} - \mathbf{B}\mathbf{M}\mathbf{w}\|^2\} \tag{3}$$

A GSM model is now employed to model the wavelet coefficients. Inspired from [28], we adopt a model which incorporates group sparsity such that $\mathbf{w}_i$, the ith group of $\mathbf{w}$, follows a zero mean Gaussian distribution with an (as yet) unknown variance of σ_i^2 per element. Therefore the conditional prior of $\mathbf{w}$ can be expressed as

$$p\,(\mathbf{w}|\mathbf{S}) = \prod_{i=1}^{G} \mathcal{N}\left(\mathbf{w}_i|0, \sigma_i^2\right) = \mathcal{N}\left(\mathbf{w}|0, \mathbf{S}^{-1}\right) \tag{4}$$

where $\mathbf{w}_i$ is a vector of coefficients comprising the ith group of size g_i, $\mathbf{S}$ is a diagonal matrix of size $N \times N$ formed from the vector $\mathbf{s}$ of size G whose ith entry is $s_i = 1/\sigma_i^2$, and G denotes the number of groups. The case $G = N$ corresponds to independent sparse modeling of the wavelet coefficients [28]; whereas the case, $G = N/2$ and $g_i = 2$ for all i, can be used to model the real and imaginary parts of G complex coefficients, each with a 2-D circularly symmetric pdf. To be consistent with the following algebra, $\mathbf{S}$ needs to be of size $N \times N$ and, when $N > G$, its diagonal must be an expanded form of $\mathbf{s}$ where each s_i appears g_i times for the elements of group g_i, and $N = \sum_{i=1}^{G} g_i$.

To proceed with Bayesian inference, the posterior distribution can be calculated via:

$$p\left(\mathbf{w}|\mathbf{y}, \mathbf{S}, \nu^2\right) = \frac{p\left(\mathbf{y}|\mathbf{w}, \nu^2\right) \times p\,(\mathbf{w}|\mathbf{S})}{p\left(\mathbf{y}|\mathbf{S}, \nu^2\right)} \tag{5}$$

Because both $p\left(\mathbf{y}|\mathbf{w}, \nu^2\right)$ and $p\,(\mathbf{w}|\mathbf{S})$ are Gaussian functions of $\mathbf{w}$, the posterior distribution can be rearranged into a Gaussian form as