

Context-Aware Sparse Decomposition for Image Denoising and Super-Resolution

Jie Ren, Jiaying Liu, *Member, IEEE*, and Zongming Guo, *Member, IEEE*

Abstract—Image prior models based on sparse and redundant representations are attracting more and more attention in the field of image restoration. The conventional sparsity-based methods enforce sparsity prior on small image patches independently. Unfortunately, these works neglected the contextual information between sparse representations of neighboring image patches. It limits the modeling capability of sparsity-based image prior, especially when the major structural information of the source image is lost in the following serious degradation process. In this paper, we utilize the contextual information of local patches (denoted as context-aware sparsity prior) to enhance the performance of sparsity-based restoration method. In addition, a unified framework based on the markov random fields model is proposed to tune the local prior into a global one to deal with arbitrary size images. An iterative numerical solution is presented to solve the joint problem of model parameters estimation and sparse recovery. Finally, the experimental results on image denoising and super-resolution demonstrate the effectiveness and robustness of the proposed context-aware method.

Index Terms—Context-aware, image denoising, image restoration (IR), Markov random fields (MRFs), sparse representation, sparsity pattern, super-resolution.

I. INTRODUCTION

IMAGE restoration (IR) aims at recovering an original image x (i.e., high-resolution, clean image) from its degraded counterpart y . The mathematical degradation model for the IR problem is typically formalized by

$$y = Hx + v \quad (1)$$

where H is a degradation matrix, and v is assumed as a zero-mean Gaussian noise $\mathcal{N}(0, \sigma^2)$. When H is an identity matrix, the IR problem in (1) is a denoising problem; when H is a downsampling matrix, it corresponds to an interpolation or super-resolution problem.

From a statistical point of view, image restoration is an inverse problem which is to estimate x given y , maximizing the probability $\Pr(x|y)$. Based on the Bayes' rule, *maximum a posterior* (MAP) estimator can be obtained by:

$$\hat{x} = \arg \max_x \Pr(x|y) = \arg \max_x \Pr(y|x) \Pr(x) \quad (2)$$

Manuscript received May 17, 2012; revised September 12, 2012; accepted November 5, 2012. Date of publication December 4, 2012; date of current version February 6, 2013. This work was supported in part by the National Natural Science Foundation of China under Contract 61101078, the Doctoral Fund of Ministry of Education of China under Contract 20110001120117, and the National Key Technology R&D Program of China under Grant 2012BAH18B03. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wai-Kuen Cham.

The authors are with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (corresponding author: liujiaying@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2012.2231690

where $\Pr(y|x)$ represents the data penalization or likelihood term, while $\Pr(x)$ models the prior information that we prefer the target original image x to be. Due to the ill-posedness of IR problem, one of the most important issues is to model *a priori* knowledge of natural images to regularize the IR inverse problem.

Modeling the prior knowledge of natural images is a challenging task. Many previous work have contributed to this issue from various fields, including probability theory, statistics, partial differential equation, and transform-domain methods. Recently, an emerging group of techniques that relies on sparse and redundant representations of image signals has attracted a lot of attention [1]. The sparsity-based image prior has been proved to be effective in various applications, such as image denoising [2], deblurring [3], super-resolution [4], inpainting [5], [6], and classification [7].

Generally, due to the high dimensionality of image signals, sparsity-based prior focuses on small patches of natural images. To deal with arbitrary size image, the whole image is divided into small image patches. In conventional sparsity-based methods [2], [4], each patch is processed independently. The final output is obtained by stitching and averaging the patches in the overlapped regions. Although simple overlapping and averaging operation has achieved good performance, we argue that the smoothness constraint in the traditional dispose is not effective enough to regularize the IR problem. Especially when the observation is downsampled or noisy, major structural information of the source image may be lost. Therefore, the contextual information between neighboring patches can provide more hints for recovering the details.

In this paper, we propose a context-aware sparsity prior to explicitly model the structural correlations of neighboring patches. In contrast to the previous work that dealt with similar issue in the specific wavelet transform domain [8], we analyze and address this problem in a more general and flexible framework of sparse representations. The main contributions of our work are summarized as follows:

- 1) Context-aware sparsity prior is proposed to model the correlations between sparsity patterns of neighboring patches. The proposed sparsity prior explicitly models the contextual information of local image structure. Therefore, it improves the capability of sparsity-based regularization for the IR problem.
- 2) MRF-based global restoration framework is proposed to tune the local sparsity prior into a global one in order to deal with natural images of arbitrary size.
- 3) An iterative numerical solution to the MRF-based global optimization problem is presented to approximate the

exact estimation which is computationally intractable in practice. The proposed algorithm adopts an adaptive signal recovery scheme, which updates model parameters update and recover the sparsity patterns alternately.

The rest of this paper is organized as follows. Section II briefly reviews the previous work that are related to this paper. In Section III, we introduce the generative signal model and briefly review some typical prior models in literature. Section IV gives an analysis of the context correlations of sparse representations, which is the motivation of the proposed work. Based on the analysis, we present the context-aware sparsity prior and tune it into a MRF-based global prior. In Section V, numerical solution to the MRF-based global restoration problem is proposed. The effectiveness and robustness of the proposed method are demonstrated by experimental results and analysis in Section VI. Finally, Section VII concludes this paper.

II. RELATED WORK

A. Sparse and Redundant Representations

The sparsity concept of natural images comes from the early work on transform-domain techniques, such as discrete cosine transform (DCT) and discrete wavelet transform (DWT). When applying a transform to natural images, a few coefficients represent the principal components of image structure. In contrast, noisy or high-frequency oscillating part distributes over the whole coefficient fields which leads to many zero or small transform coefficients. Based on the above observation, the well-known *Shrinkage* techniques are proposed for image denoising [9], [10]. Simple scalar thresholding operation is applied on each transform coefficient to eliminate small ones and leave large ones.

Sparse and redundant representations of signals generalize the concept of transform-domain sparsity. It states that a high-quality image x can be approximately represented by a vector α over a dictionary $\Phi \in R^{n \times m}$ (each column in Φ is referred to as an atom Φ_i), such that

$$x \approx \Phi\alpha, \quad \text{s.t. } \|\alpha\|_0 \leq T \quad (3)$$

where T is a predefined threshold and l_0 -norm $\|\alpha\|_0$ counts the number of nonzero elements in α . Based on the sparsity prior model, the problem of recovering the image x turns into an optimal estimation of representation coefficients α by solving the following MAP estimation:

$$\hat{\alpha} = \arg \max_{\alpha} \Pr(y|\alpha) \Pr(\alpha) \quad (4)$$

where $\Pr(\alpha) \propto \exp(-\lambda\|\alpha\|_0)$ is inclined to the sparsity of signal representation. Owing to the computational complexity of the l_0 regularized estimation problem, approximation algorithms are commonly employed, such as greedy pursuit algorithms like the Orthogonal Matching Pursuit (OMP) [11], [12] and convex relaxation algorithms like the Basis Pursuit (BP) [13].

In conventional sparse recovery algorithms such as OMP and BP, independency between the dictionary atoms is implicitly assumed. Although this assumption reduces the computational complexity, it limits the representation

capability to model signals in the real world. For example, in the wavelet transform domain of natural images, the locations of large coefficients are strongly correlated. Recent work [14]–[21] show that structured sparsity models that consider the probabilistic dependencies between atoms are more adaptive to natural image signals. This series of work can further promote the performance of sparse recovery algorithms. Inspired by these pioneer work, we are also interested in modeling the statistical dependencies between dictionary atoms. Different from the previous work [15], [20], [21] which mainly focused on the atom correlations within the local patch, we are especially interested in exploring the correlations of atoms between neighboring patches. It is accomplished by modeling the contextual information of sparse representations of neighboring image patches.

Another important issue in sparse and redundant representations is to select an appropriate dictionary [22]. An over-complete dictionary can be chosen as a pre-specified set of functions, *e.g.*, Redundant DCT, Curvelet [23], Contourlet [24], Shearlets [25]. Although the pre-specified dictionaries lead to fast transforms, they are limited to sparsely describe arbitrary and new family of signals of interest. Therefore, dictionary learning approaches [26]–[29] are proposed to design the dictionary by adapting its content to fit a given set of signal examples. One of the most efficient methods is K-SVD [27] which uses either OMP or BP as part of its iterative procedure for dictionary learning. In this work, we choose the Redundant DCT and adaptive dictionary trained by K-SVD to evaluate the robustness of the proposed method.

B. MRF-Based Global Prior

Traditionally, dictionary learning and sparse coding methods for sparse representations focus on small image patches because of the high dimensionality of the image space. However, in practice, we need to deal with arbitrary natural images rather than small size patches. One heuristic way is to divide the image into small non-overlapped patches and apply any patch-wise sparse coding technique for each patch separately. Unfortunately, blocking artifacts appear when the independently processed non-overlapped patches are stitched together into a whole image.

An intuitive improvement to alleviate the blocking artifacts is to overlap the patches and average the pixels near patch boundaries [30]–[32]. Elad and Aharon [2] presented a systematic global MAP estimator for the whole image prior, and derived a numerical solution which is similar to [32]. The key idea of defining such a global image prior is to combine the local sparsity prior with MRFs model [6]. It enforces sparsity on each single patch while implicitly imposing the smoothness constraint on overlapped region between patches. However, when the observation is noisy and downsampled, major structural information of the source image may be lost. The smoothness constraint is not effective enough to regularize the IR problem. Therefore, it is reasonable to explore contextual information of local image structures, which is one of the main concerns of this work.

III. BACKGROUND

A. Generative Signal Model

First, we introduce the generative statistical model in [21] which is a fundamental of our work and analysis. We consider a measured signal y that is linearly modeled as

$$y = H\Phi\alpha + v \quad (5)$$

where $\Phi = [\phi_1, \dots, \phi_m]$ is a dictionary, ϕ_i is a dictionary atom, and m is the number of atoms, i.e., dictionary size. α is a sparse representation over the dictionary, and v is additive white Gaussian noise with variance σ^2 . Note that (5) can be simplified to $y = \tilde{\Phi}\alpha + v$, where $\tilde{\Phi} = H\Phi = [\tilde{\phi}_1, \dots, \tilde{\phi}_m] \in \mathbb{R}^{n \times m}$, and $\tilde{\phi}_i = H\phi_i$.

We define the sparsity pattern of α as $S \in \{-1, 1\}^m$, where $S_i = 1$ implies that $\alpha_i \neq 0$ whereas $S_i = -1$ implies that $\alpha_i = 0$. Given the sparsity pattern S , the positions of nonzero coefficients in α are also fixed. We denote the nonzero coefficients in α as α_S , and the corresponding atoms in $\tilde{\Phi}$ which participate in the representation α_S are grouped into a sub-dictionary denoted by $\tilde{\Phi}_S$. Following [15], we consider a Gaussian distribution with zero mean and variance $\sigma_{\alpha,i}^2$ for each nonzero coefficient α_i and assume that the nonzero coefficients are independent of each other. The conditional distribution of α_S given the sparsity pattern S is a multivariate Gaussian distribution with zero mean and covariance Σ_S ,

$$\Pr(\alpha_S|S) = \frac{1}{\det(2\pi \Sigma_S)^{1/2}} \exp\left(-\frac{1}{2}\alpha_S^T \Sigma_S^{-1} \alpha_S\right) \quad (6)$$

where Σ_S is a $k \times k$ diagonal matrix in which the diagonal elements are the corresponding variances $\sigma_{\alpha,i}^2$ of the nonzero coefficients α_i , and k is the total number of the nonzero coefficients in α . According to the well-known Gaussian assumption of noise characteristics, the conditional distribution of the signal y given its sparse representation α_S and sparsity pattern S can be written as

$$\Pr(y|\alpha_S, S) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|y - \tilde{\Phi}_S \alpha_S\|_2^2}{2\sigma^2}\right). \quad (7)$$

The conditional distribution of signal y given its sparsity pattern S is the marginal probability distribution over all possible values of α_S , i.e.,

$$\Pr(y|S) = \int_{\mathbb{R}^k} \Pr(y|\alpha_S, S) \Pr(\alpha_S|S) d\alpha_S \quad (8)$$

which is a multivariate Gaussian distribution with zero mean and conditional covariance $\Sigma_{(y|S)}$ [33],

$$\Pr(y|S) = \frac{1}{\det(2\pi \Sigma_{(y|S)})^{1/2}} \exp\left(-\frac{1}{2}y^T \Sigma_{(y|S)}^{-1} y\right) \quad (9)$$

where

$$\Sigma_{(y|S)} = \tilde{\Phi}_S \Sigma_S \tilde{\Phi}_S^T + \sigma^2 \mathbf{I}_n, \quad \Sigma_{(y|S)}^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I}_n - \tilde{\Phi}_S \mathbf{Q}_S^{-1} \tilde{\Phi}_S^T \right)$$

with $\mathbf{Q}_S = \tilde{\Phi}_S^T \tilde{\Phi}_S + \sigma^2 \Sigma_S^{-1}$.

B. Review of Prior Models in the Literature

In this work, we are especially interested in the estimation of the underlying ‘‘true’’ sparsity patterns S of an original signal x , given its degradation measurement y . The MAP estimator is obtained by maximizing the posterior probability of sparsity pattern S ,

$$\hat{S} = \arg \max_S \Pr(S|y) = \arg \max_S \Pr(y|S) \Pr(S). \quad (10)$$

We denote the recovery of sparsity pattern in (10) as a **Sparse Decomposition** problem in this paper. In (10), the prior probability $\Pr(S)$ is chosen to be some sparsity penalty. For example, the prior model can be written as

$$\Pr(S) = C \cdot \exp(-\lambda \cdot \|S + \mathbf{1}\|_0) \quad (11)$$

where C is the normalization constant, λ is a shape control parameter and $\mathbf{1}$ is the all-ones vector. This form of prior corresponds to the sparsity prior implied in typical deterministic sparse coding algorithms, e.g., OMP and BP. Furthermore, the independent adaptive model in [34] assigns a different prior probability P_i for each atom $\tilde{\phi}_i$ to be selected in the sparse representation. Therefore, the joint probability for S is given by

$$\Pr(S) = \prod_i g_i, \quad g_i = \begin{cases} P_i, & S_i = 1, \\ 1 - P_i, & S_i = -1. \end{cases} \quad (12)$$

P_i is estimated from the data by computing the frequency of occurrence of atom α_i participating in all representations. An alternative to the independent adaptive model (12) is formalized by

$$\Pr(S) = \frac{1}{Z(b)} \exp(b^T S) = \frac{1}{Z(b)} \prod_{i=1}^m \exp(b_i S_i) \quad (13)$$

where $b = [b_1, b_2, \dots, b_m]^T$ is a vector of model parameters, $Z(b)$ is a partition function of parameters b that normalizes the distribution. A bias b_i is associated with the dictionary atom $\tilde{\phi}_i$ where $b_i < 0$ favors $S_i = -1$, which in turn makes the atom $\tilde{\phi}_i$ mostly unselected. The equivalence of (12) and (13) is apparent when we let $P_i = \exp(b_i)/(2 \cosh(b_i))$, where $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ is the hyperbolic cosine function.

All the above models are based on the same assumption that there is no interaction between dictionary atoms. That is, in the sparse coding stage, dictionary atoms are selected into the final representation independently. However, natural image signals exhibit significant connections between atoms of the dictionary used for synthesis. Recent work on structured sparsity models as in [15], [20], [21] suggested using a more general graphical model framework, specifically the Boltzmann machine (BM) model, to consider the interaction effects between dictionary atoms. The BM model is a special case of the exponential probability distribution family with energy function $E(S)$,

$$\Pr(S) = \frac{1}{Z(W, b)} \exp(-E(S))$$

$$E(S) = -b^T S - \frac{1}{2} S^T W S \quad (14)$$

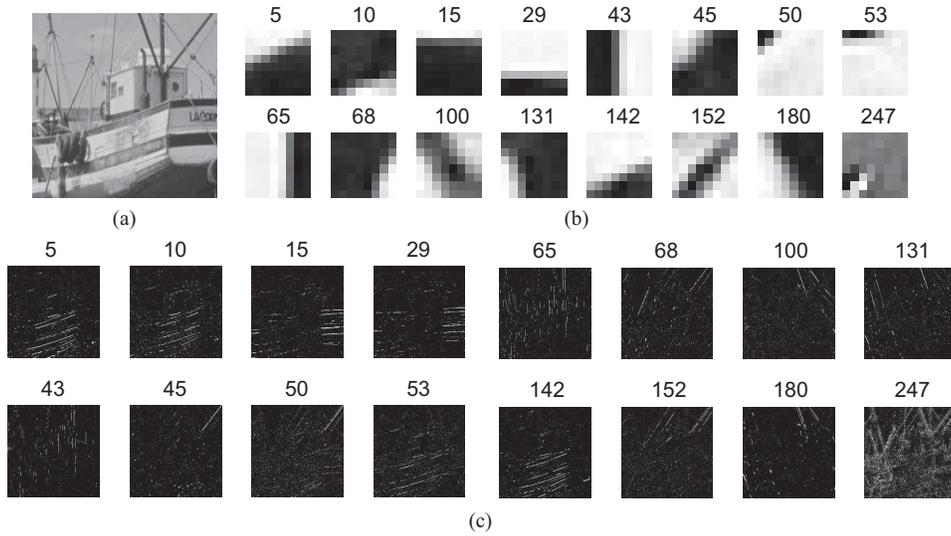


Fig. 1. Illustration of the spatial distributions of dictionary atom coefficients on image *Boats*. (a) *Boats* image. (b) Learned dictionary atoms (16 selected out of 256). (c) Spatial distributions of corresponding atom coefficients.

where $Z(W, b)$ is a partition function of the Boltzmann parameters W, b . When $W = \mathbf{0}$, the BM model reduces to the independent model in (13). The interaction matrix W accounts for the dependencies between dictionary atoms. Specifically, $W(i, j) > 0$ suggests that the atoms $\tilde{\phi}_i$ and $\tilde{\phi}_j$ have an “excitatory” interaction, i.e., S_i and S_j tend to be the same value; for an “inhibitory” interaction ($W(i, j) < 0$), they tend to be opposite value. $W(i, j) = 0$ indicates the independency between atoms.

Note that once the optimal sparsity pattern \hat{S} has been recovered, one can further reconstruct a source signal x if necessary by performing the MAP estimation (i.e., oracle estimator in [35]),

$$\begin{aligned} \hat{\alpha}_S &= \arg \max_{\alpha_S} \Pr(\alpha_S | y, \hat{S}) = \mathbf{Q}_S^{-1} \tilde{\Phi}_S^T y \\ x &= \tilde{\Phi}_S \hat{\alpha}_S = \tilde{\Phi}_S \mathbf{Q}_S^{-1} \tilde{\Phi}_S^T y. \end{aligned} \quad (15)$$

Instead of focusing on the atom interactions within local patches, in this work, we pay more attention to model the atom interactions between neighboring patches. In the next section, we give an analysis of the contextual correlations of sparse representations as the motivation of our work, and then a new local sparsity prior and an extended MRF-based global prior are presented in the following subsections.

IV. CONTEXT-AWARE SPARSE DECOMPOSITION

A. Context Correlation Analysis of Sparse Representations

Sparse coding of image patches extracted from the whole image can be seen as filtering of the image with a set of filters. Similar to the wavelet transform, casual observation indicates that the sparse coefficients (or large amplitudes) of every patch are sparsely distributed throughout the whole image and often tend to occur in clusters (*e.g.*, at edges and within textures as illustrated in Fig. 1). Meanwhile, clusters of different coefficient distributions seem not to be fully independent each other, such as the distribution clusters of coefficient index 5 and 10, 15 and 29, 53 and 142. It implies that different

structural filters (or atoms of a dictionary) are highly correlated in the spatial neighborhood.

The main reason for the above phenomena is attributed to the regularity of natural image structures. In general, local structures of natural images have the properties of visual continuity and consistency. We find that these properties may not necessarily appear in the image domain but commonly in the principal components domain, i.e., the dictionary atoms domain. Therefore, it is necessary to decompose the image into a few of dictionary atoms, and analyze the image regularity in the dictionary atoms domain.

We proceed with the following experiments to motivate the necessary of modeling the spatially probabilistic dependencies between dictionary atoms. The dictionary is adaptively learned from the data itself by K-SVD algorithm [27]. We extract a set of patches of size 8-by-8 from noise-free and high-resolution natural images. The extracted patches are overlapped with a few pixels in adjacent locations. For each patch, sparse representations over an adaptive learned dictionary of size 64-by-256 are obtained using the OMP algorithm. Specifically, the model error is set to $\sigma = 1$, so that the OMP algorithm stops when the residual error falls below $\epsilon = 1.15 \cdot \sqrt{64} \cdot \sigma$. Then the sparsity pattern of each patch is obtained from the computed representations. Let S and $S_{\diamond t}$ ($t = 1, \dots, T$) represent the sparsity patterns of center position patch and its neighbor in the t -th orientation, respectively (the spatial configuration is shown in Fig. 2). We take all pairs of sparsity patterns as samples for S and $S_{\diamond t}$. The i -th element of S , and the j -th element of $S_{\diamond t}$ are represented as S_i and $S_{j \diamond t}$, $i, j = 1, \dots, m$, respectively. Then, the empirical marginal distributions for each S_i , $\Pr(S_i = 1)$, and for all pairs of atoms at neighboring spatial position with different orientations, $\Pr(S_i = 1, S_{j \diamond t} = 1)$, are computed. The empirical conditional probability $\Pr(S_i = 1 | S_{j \diamond t} = 1)$ is computed by $\Pr(S_i = 1 | S_{j \diamond t} = 1) = \frac{\Pr(S_i = 1, S_{j \diamond t} = 1)}{\Pr(S_{j \diamond t} = 1)}$.

To examine the dependencies between pairs of atoms at neighboring spatial position, we compute the following

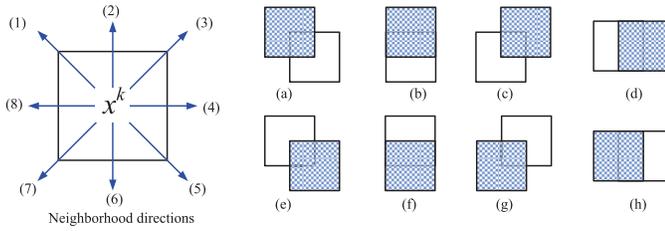


Fig. 2. Local neighborhood system of patch x^k . We set $J = 8$ for spatial configuration with eight different orientations. (a) Left-up. (b) Up. (c) Right-up. (d) Right. (e) Right-down. (f) Down. (g) Left-down. (h) Left.

three-dimension dependency matrix value:

$$Q_{i,j}^{(t)} = \left| \log_{10} \left(\frac{\Pr(S_i = 1 | S_{j \diamond t} = 1)}{\Pr(S_i = 1)} + \delta \right) \right|. \quad (16)$$

Therefore, if S_i and $S_{j \diamond t}$ are independent, then $\Pr(S_i = 1 | S_{j \diamond t} = 1)$ is approximately equal to $\Pr(S_i = 1)$, and the value of (16) is near zero. Otherwise, $\Pr(S_i = 1 | S_{j \diamond t} = 1)$ is near zero or one, then the value of (16) is approaching to 1. In this experiment, we set $\delta = 0.1$, so that for $\Pr(S_i = 1 | S_{j \diamond t} = 1) = 0$, the value of $Q_{i,j}^{(t)}$ is 1.

The simulation results on the image Boats are shown in Fig. 3. For better visualization, the three-dimensional matrix Q is split into eight orientations, and each slice of Q in the corresponding orientation is displayed. In these figures, the black color corresponding to near-zero values is dominant. It implies that most of the atom pairs at neighboring position are nearly independent. However, there are some pairs exhibiting significant dependencies, particularly near the main diagonal. The correlations between sparsity patterns of neighboring patches in different orientations reveal another regularity of natural images.

B. Context-Aware Sparsity Prior

The significant dependencies exhibited among the atom pairs can be utilized as another level of image regularity beyond the sparsity prior. An important issue is to model the spatial correlation between atoms adaptively, i.e., we need to decide which atom pairs should have strong correlations and make the others have weak or no correlations.

Motivated by the BM model, we propose a new prior for the sparsity patterns of sparse representations. The proposed sparsity prior explicitly models the atom interactions between neighboring patches. It contains two components including the context-aware part and the sparsity part. Therefore, we denote the new prior model as context-aware sparsity prior for the combination. More specifically, given the orientated neighboring sparsity patterns $\{S_{\diamond t}\}_{t=1}^J$, we define the context-aware energy $E_c(S)$ by

$$E_c(S) = - \sum_{t=1}^J S^T W_{\diamond t} S_{\diamond t} \quad (17)$$

where the matrix $W_{\diamond t}$ captures the interaction strength between dictionary atoms in the t -th orientation. The interaction strength of each element in $W_{\diamond t}$ relies on the co-occurrence probability between dictionary atoms in the corresponding orientation, which should be adaptive to the image content

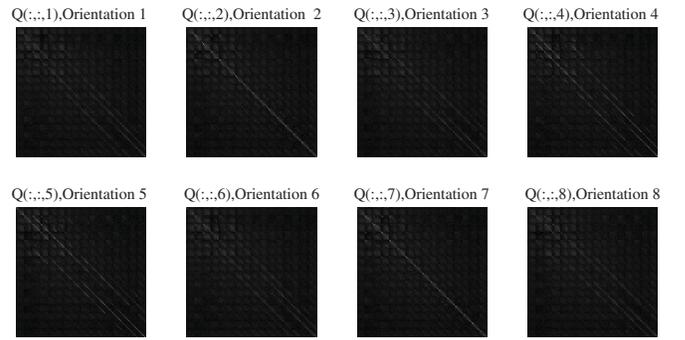


Fig. 3. Visualization of the computed dependency matrix Q on image Boats. The 3-D matrix Q is split into eight orientations, and each slice of Q corresponds to one orientation.

as well. In Section V-A, we present a maximum likelihood estimator to obtain the optimal parameters $W_{\diamond t}$ in an adaptive scheme. Meanwhile, we also include the sparsity penalty energy $E_s(S)$ as in (13),

$$E_s(S) = -S^T b. \quad (18)$$

The total energy $E_{\text{total}}(S)$ for each sparsity pattern is the sum of two parts, i.e., $E_{\text{total}} = E_c(S) + E_s(S)$. Then, the prior probability is formalized by using E_{total} ,

$$\Pr(S) \propto \exp(-E_{\text{total}}) \propto \exp\left(S^T \left(\sum_{t=1}^J W_{\diamond t} S_{\diamond t} + b\right)\right). \quad (19)$$

Let $\tilde{W} = [W_{\diamond 1}, \dots, W_{\diamond J}]$, and $\tilde{S} = [(S_{\diamond 1})^T, \dots, (S_{\diamond J})^T]^T$, then (19) can be expressed in a clearer form,

$$\Pr(S) = \frac{1}{Z(\tilde{W}, b)} \exp\left(S^T (\tilde{W} \tilde{S} + b)\right) \quad (20)$$

where \tilde{W} , b are model parameters, and $Z(\tilde{W}, b)$ is the partition function for normalization. Compared to the BM model, the proposed prior model places more emphasis on the dependencies of atoms in the spatial context. In fact, we can combine the proposed model with the BM model for further improvement of the modeling capability, which is one of the interests of future work.

C. MRF-Based Global Restoration

The prior model proposed in the previous subsection is defined in a patch-wise scheme. It is enforced over the local neighborhood range of each patch. In fact, the neighboring sparsity patterns \tilde{S} are always unknown when addressing the sparsity pattern recovery for one single patch. Meanwhile, when dealing with an arbitrary size image, it is necessary to extend the local prior to a global one as in [2], [6].

MRFs theory provides a convenient and consistent way to model context-dependent entities through characterizing mutual influences among entities using conditional MRFs distributions. It has been proved to be a robust and accurate model for natural image [36]. Therefore, it is natural to incorporate the context-aware sparsity prior into the MRFs framework.

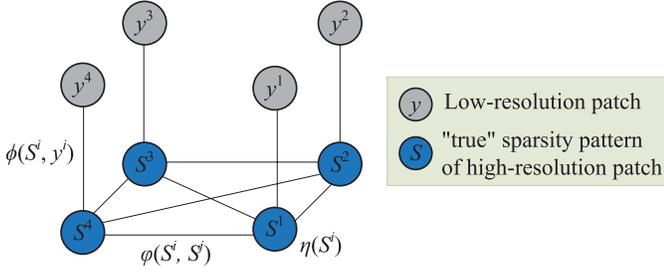


Fig. 4. Eight-connected MRFs. Nodes y and S are low-resolution patch and “true” sparsity pattern, respectively.

For an input degraded image Y of arbitrary size, we first break it into overlapped small patches $\{y^k\}_{k=1}^K$. Each patch y^k has a corresponding high-quality patch x^k , and the “true” sparsity pattern of x^k is denoted as S^k . $\mathcal{S} = \{S^k\}_{k=1}^K$ represents the whole set of sparsity patterns. We introduce an 8-connected MRFs to model the relationships among the degraded patches and their corresponding high-quality patches, as illustrated in Fig. 4. Based on the MRFs model, we define three types of potential functions corresponding to the likelihood term $\phi(S^k, y^k)$, sparsity term $\eta(S^k)$ and context-aware term $\psi(S^k, S^p)$,

$$\begin{aligned}\phi(S^k, y^k) &\propto \Pr(y^k | S^k) \\ \eta(S^k) &\propto \exp\left((S^k)^T b\right) \\ \psi(S^k, S^p) &\propto \exp\left((S^k)^T W_{\text{ot}} S_{\text{ot}}^k\right)\end{aligned}$$

which use the fact that patch x^p is adjacent to x^k in the t -th orientation. Once the potential functions are determined, the MRFs with homogeneous potentials could be written as

$$\Pr(\mathcal{S}, Y) \propto \prod_k \eta(S^k) \phi(S^k, y^k) \prod_{k,p} \psi(S^k, S^p). \quad (21)$$

Note that $\eta(S^k)$ and $\psi(S^k, S^p)$ with 8 orientations can be reassembled to the prior probability $\Pr(S^k)$ in (20), and $\phi(S^k, y^k)$ corresponds to the likelihood probability $\Pr(y^k | S^k)$. Therefore, the probability for the whole MRFs is written as

$$\Pr(\mathcal{S}, Y) \propto \prod_k \Pr(y^k | S^k) \Pr(S^k). \quad (22)$$

The complete set of sparsity patterns \mathcal{S} in the MRFs can be optimally estimated by maximizing the joint probability of MRFs,

$$\max \Pr(\mathcal{S}, Y) = \max \sum_{k=1}^K \left(\ln \Pr(y^k | S^k) + \ln \Pr(S^k) \right). \quad (23)$$

V. NUMERICAL SOLUTION

In practice, for a given set of degraded patches $\{y^k\}_{k=1}^K$ extracted from an input image Y , we would like to estimate both the model parameters and the set of sparsity patterns of each patch, $\mathcal{S} = \{S^k\}_{k=1}^K$. The model parameters include \tilde{W} , b , and $\{\sigma_{\alpha,i}^2\}_{i=1}^m$. Instead of addressing both together, we suggest a block-coordinate maximization algorithm that starts with $\tilde{W} = 0$, $b_i = \text{const}$, and $\sigma_{\alpha,i}^2 = \text{const}$. In fact, under this

Algorithm 1: Block-Coordinate Maximization Algorithm for Joint Estimation

Input: Noisy observations $\{y^k\}_{k=1}^K$, dictionary $\tilde{\Phi}$, noise variance σ^2 .

Output: Recovery of sparsity patterns $\hat{\mathcal{S}} = \{\hat{S}^k\}_{k=1}^K$ and model parameters $\theta = [\tilde{W}, b, v = \{\sigma_{\alpha,i}^2\}_{i=1}^m]$.

$l = 0$;

Initialize $\mathcal{S}^{(0)}$ with OMP results;

while $l < \text{maxIter}$ **do**

$l = l + 1$;

Estimation of the model parameters

$$\theta^{(l)} = [\tilde{W}^{(l)}, b^{(l)}, v^{(l)}];$$

Pursuing the sparsity patterns $\mathcal{S}^{(l)}$ using the updated parameters;

end

return $\hat{\mathcal{S}}, \tilde{W}, b, v$.

condition the sparse recovery of \mathcal{S} is close to the results by independently using the plain OMP algorithm on each patch. In our implementation, we use OMP algorithm to initialize the sparsity pattern recovery. Model parameters are updated by using the current sparsity pattern recovery results. Once the parameters are updated, improved recovery of sparsity patterns can be obtained. Each iteration contains these two stages and the algorithm stops when it converges or reaches the maximum iteration number. The pseudocode of the proposed block-coordinate maximization algorithm is summarized in Algorithm 1.

A. Model Parameters Estimation

Given $\mathcal{X} = \{y^k, S^k, \alpha^k, \tilde{S}^k\}_{k=1}^K$ as examples sampled from the model, we suggest using the Maximum Likelihood Estimation (MLE) for learning the model parameters $\theta = [\tilde{W}, b, \{\sigma_{\alpha,i}^2\}_{i=1}^m] \in \Theta$. Mathematically, we have

$$\begin{aligned}\hat{\theta}_{ML} &= \arg \max_{\theta} \Pr(\mathcal{X} | \theta) \\ &= \arg \max_{\theta} \sum_{i=1}^m \mathcal{L}(\sigma_{\alpha,i}^2) + \mathcal{L}(\tilde{W}, b),\end{aligned} \quad (24)$$

where

$$\begin{aligned}\mathcal{L}(\sigma_{\alpha,i}^2) &= \frac{1}{2} \sum_{k=1}^K f_i^k, \\ \mathcal{L}(\tilde{W}, b) &= \frac{1}{2} \sum_{k=1}^K (S^k)^T (\tilde{W} \tilde{S}^k + b) - K \ln Z(\tilde{W}, b),\end{aligned} \quad (25)$$

are log-likelihood functions for the model parameters and

$$f_i^k = \begin{cases} \frac{(\alpha_i^k)^2}{\sigma_{\alpha,i}^2} + \ln(\sigma_{\alpha,i}^2), & S_i^k = 1, \\ 0, & S_i^k = -1. \end{cases} \quad (26)$$

For the estimation of variances, a closed-form estimator is obtained by:

$$\hat{\sigma}_{\alpha,i}^2 = \frac{\sum_{k=1}^K (\alpha_i^k)^2 \cdot q_i^k}{\sum_{k=1}^K q_i^k} \quad (27)$$

where

$$q_i^k = \begin{cases} 1, & S_i^k = 1, \\ 0, & S_i^k = -1. \end{cases}$$

However, ML estimation of \tilde{W} and b is computationally intensive due to the exponential complexity in m associated with the partition function $Z(\tilde{W}, b)$. In [15], a gradient-based optimization algorithm with Gibbs sampling and mean-field techniques is used. Faktor *et al.* [20], [21] proposed a more efficient algorithm using the MPL estimation [37] and sequential subspace optimization (SESOP) method [38]. In our work, we adopt the same procedure as in [21]. Details can be found in that paper.

B. Recovery of Sparsity Patterns

Once the parameters are known, one way to compute the global optimal configuration for the MRFs model in (23) is to provide a set of possible candidates for each node, then approximately solve it by the Belief Propagation (BP) algorithm. However, since the number of possible configurations of each node is exponential to the number of the dictionary atoms (i.e., there are 2^m possible candidates for S), it is computationally intractable in practice. Instead, we present an approximated numerical solution that iteratively recovers the sparsity pattern of each patch, as in the Gauss-Seidel iterative method.

In the proposed algorithm, all the patches are processed in raster-scan order in an image, i.e., from left to right and top to bottom. When processing the current center patch, all sparsity patterns of the neighboring patches \tilde{S} are utilizing the latest updated value and kept fixed during the recovery of sparsity pattern for center patch. Due to the overlapping of extracted patches, the updated sparsity pattern of the current patch is immediately used in the processing of next neighboring patch. The procedure is performed repeatedly to propagate the contextual information among all the nodes.

The above simplification for solving the whole set of sparsity patterns of the MRFs can also be viewed as a block-coordinate method, in which when updating one single sparsity pattern, the others are known and fixed. Under this configuration, the recovery of individual sparsity pattern is rewritten as

$$\begin{aligned} \hat{S} &= \arg \max_S \Pr(S|y) = \arg \max_S \{\ln \Pr(y|S) + \ln \Pr(S)\} \\ &= \arg \max_S \left(\frac{1}{2\sigma^2} y^T \tilde{\Phi}_S \mathbf{Q}_S^{-1} \tilde{\Phi}_S^T y - \frac{1}{2} \ln \det(\mathbf{Q}_S) \right. \\ &\quad \left. - \frac{1}{2} \ln \left(\prod_{S_i=1} \frac{\sigma_{\alpha,i}^2}{\sigma^2} \right) + S^T (\tilde{W}\tilde{S} + b) \right). \end{aligned} \quad (28)$$

In this configuration, the exact MAP estimator requires an exhaustive search over all 2^m possible configurations. Therefore, we propose a greedy algorithm as an approximate MAP estimation for solving (28). The greedy algorithm starts with an initialization with $S_i = -1, \forall i$, and then iteratively changes the value of entry S_i to 1 that makes the posterior probability in (28) with the biggest growth comparing to all

Algorithm 2: MRF-based Sparsity Patterns Recovery Algorithm

Input: Noisy observations $\{y^k\}_{k=1}^K$, dictionary $\tilde{\Phi}$, noise variance σ^2 , model parameters $\theta = [\tilde{W}, b, v = \{\sigma_{\alpha,i}^2\}_{i=1}^m]$, initialization $S^{(0)}$, \maxPass .

Output: Recovery of sparsity patterns $\hat{S} = \{\hat{S}^k\}_{k=1}^K$.
 $p = 0$;

while $p < \maxPass$ **do**

$p = p + 1$;

for every patch y^k in raster-scan order **do**

 Collect the sparsity patterns of neighboring patches, \tilde{S}^k ;

 Let $Z^{(0)} = (S^k)^{(0)}$;

 Let $\forall i, Z_i^{(0)} = -1$;

$l = 1$;

repeat

for $\forall i, Z_i^{(l-1)} = -1$ **do**

for $j = 1, 2, \dots, m$ **do**

if $j \neq i$ **then** $Z_j^{(l)} = Z_j^{(l-1)}$;

else $Z_j^{(l)} = 1$;

end

 Evaluate $P(i) = \Pr(Z^{(l)}|y)$ using (28);

end

$i_* = \arg \max_i \{P(i)\}$;

for $j = 1, 2, \dots, m$ **do**

if $j \neq i_*$ **then** $Z_j^{(l)} = Z_j^{(l-1)}$;

else $Z_j^{(l)} = 1$;

end

$l = l + 1$;

until $\Pr(Z^{(l)}|y) < \Pr(Z^{(l-1)}|y)$;

$\hat{S}^k = Z^{(l-1)}$;

end

return $\hat{S} = \{\hat{S}^k\}_{k=1}^K$.

other candidates. The iteration stops until the the posterior probability in (28) reaches a local optimal value.

The pseudocode of the MRF-based recovery algorithm is summarized in Algorithm 2.

C. Fast and Efficient Implementation

A straightforward implementation of the method presented in this section is highly demanding in time-space complexity. In order to achieve a practical and efficient algorithm, we introduce some constraints and exploit certain optimization strategy as follows:

- 1) The full number of parameters in \tilde{W} is very large. For example, if the patch size is $n = 8 \times 8$, and the redundancy ratio of a dictionary is set to be 4, the dictionary size is $m = n \times 4 = 256$. When we choose to use the 8-connected MRFs to model the spatial connections, the number of parameters in \tilde{W} is $8 \times m^2 = 524288$. To reduce the number of parameters while maintaining the performance of the proposed model, we impose a

TABLE I
SUMMARY OF DENOISING RESULTS ON DIFFERENT KINDS OF DICTIONARIES (PSNR: dB). THE BEST RESULT FOR EACH PAIR OF COMPARING METHODS FOR DIFFERENT NOISE LEVELS IS HIGHLIGHTED

σ	Image	Redundant DCT			Global K-SVD			Adaptive K-SVD		
		OMP	CASD	Δ PSNR	OMP	CASD	Δ PSNR	OMP	CASD	Δ PSNR
5	Barbara	34.33	34.74	0.42	34.17	34.34	0.18	34.74	34.88	0.13
	Boats	34.62	35.05	0.43	34.88	35.14	0.27	35.11	35.22	0.12
	House	35.62	35.92	0.29	35.83	36.02	0.19	36.13	36.27	0.13
	Lena	34.84	35.14	0.31	35.20	35.32	0.12	35.12	35.28	0.16
	Average	34.85	35.21	0.31	35.02	35.21	0.19	35.28	35.41	0.14
10	Barbara	29.89	30.41	0.53	29.72	30.00	0.29	30.56	30.78	0.22
	Boats	30.30	30.88	0.58	30.79	31.17	0.37	31.07	31.28	0.21
	House	31.94	32.35	0.41	32.42	32.73	0.30	32.57	32.80	0.23
	Lena	30.49	30.85	0.36	31.04	31.27	0.22	30.89	31.07	0.18
	Average	30.66	31.12	0.47	30.99	31.29	0.30	31.27	31.48	0.21
15	Barbara	27.38	27.98	0.60	27.26	27.61	0.34	28.27	28.59	0.31
	Boats	27.92	28.59	0.68	28.52	28.97	0.45	28.82	29.12	0.31
	House	29.87	30.41	0.55	30.53	30.94	0.41	30.69	31.09	0.41
	Lena	28.10	28.57	0.46	28.74	29.04	0.29	28.69	28.92	0.23
	Average	28.32	28.89	0.57	28.76	29.14	0.37	29.12	29.43	0.32
20	Barbara	25.69	26.35	0.65	25.63	26.01	0.39	26.62	27.02	0.40
	Boats	26.29	27.05	0.76	26.99	27.51	0.53	27.21	27.62	0.42
	House	28.38	29.07	0.69	29.11	29.66	0.54	29.22	29.73	0.51
	Lena	26.50	27.06	0.56	27.21	27.57	0.36	27.15	27.49	0.34
	Average	26.72	27.38	0.67	27.24	27.69	0.46	27.55	27.97	0.42
25	Barbara	24.44	25.16	0.72	24.43	24.85	0.42	25.34	25.81	0.47
	Boats	25.09	25.94	0.85	25.82	26.41	0.59	25.93	26.46	0.53
	House	27.18	28.05	0.88	27.97	28.67	0.70	28.02	28.77	0.75
	Lena	25.29	25.98	0.69	26.04	26.51	0.47	25.98	26.45	0.47
	Average	25.50	26.28	0.79	26.07	26.61	0.55	26.32	26.87	0.56
50	Barbara	21.01	21.88	0.86	21.13	21.74	0.62	21.37	21.96	0.59
	Boats	21.58	22.54	0.96	22.15	22.87	0.72	21.85	22.52	0.67
	House	23.33	24.48	1.15	23.96	24.96	1.00	23.71	24.63	0.92
	Lena	21.74	22.69	0.94	22.41	23.14	0.73	22.11	22.82	0.70
	Average	21.92	22.90	0.98	22.41	23.18	0.77	22.26	22.98	0.72

constraint on each $W_{\sigma t}$, $t = 1, 2, \dots, 8$. That is only the elements on the main diagonal are considered to be updated while all other elements are set to be 0. It means that we eliminate the correlations between different atoms $\tilde{\Phi}_i$ and $\tilde{\Phi}_j$, $i \neq j$. We make the above constraint based on the observation of the experiment in Section IV-A. As shown in Fig. 3, the strong correlations of natural images exist mostly on the main diagonal. When imposing this constraint on the proposed model, \tilde{W} is constructed as follows:

$$\tilde{W} = [\text{diag}(W_{\sigma 1}), \dots, \text{diag}(W_{\sigma 8})]$$

where diag is the operator that erases all the elements of a matrix to zero except the diagonal ones. Under this constraint, the number of parameters in \tilde{W} is reduced to $8 \times m = 2048$.

- 2) In Algorithm 1, the model parameters are updated after the every recovery stage of the sparsity patterns. However, the model parameters estimation algorithm is time consuming, and we find it change slightly in the iterations. Therefore, to reduce the computational

complexity of the proposed algorithm, model parameters are only updated several times during the iteration of sparsity patterns recovery.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method is evaluated with two applications: image denoising and super-resolution. As described before, for image denoising, the degradation operator H is an identity matrix. For super-resolution, the degradation operator H is a downsampling operation matrix. Therefore, $\tilde{\Phi} = H\Phi$ corresponds to the low-resolution dictionary. Instead of explicitly constructing the degradation operator H , we directly learn a pair of low- and high-resolution dictionaries as proposed in [4], [39]. In particular, we adopt the dictionary learning algorithm proposed by Zeyde *et al.* [39] that splits joint learning into separable low-resolution and high-resolution dictionary learning to improve the recovery accuracy. We collect a set of high quality natural images¹ as the training dataset for dictionary learning. For learning a pair of low and high-resolution

¹Downloaded from website Available at: <http://www.ifp.illinois.edu/~jyang29/>.

TABLE II
SUMMARY OF DENOISING RESULTS ON DIFFERENT DICTIONARY SIZES $m = r \cdot 8^2$, $r = 3-5$ (PSNR: dB).
THE BEST RESULT FOR EACH PAIR OF COMPARING METHODS IS HIGHLIGHTED

σ	Image	$r = 3$			$r = 4$			$r = 5$		
		OMP	CASD	Δ PSNR	OMP	CASD	Δ PSNR	OMP	CASD	Δ PSNR
5	<i>Barbara</i>	34.76	34.78	0.02	34.74	34.88	0.13	34.73	34.69	-0.04
	<i>Boats</i>	35.14	35.14	0.00	35.11	35.22	0.12	35.12	35.09	-0.04
	<i>House</i>	36.11	36.18	0.07	36.13	36.27	0.13	36.11	36.15	0.04
	<i>Lena</i>	35.14	35.20	0.06	35.12	35.28	0.16	35.12	35.14	0.02
	Average	35.29	35.33	0.04	35.28	35.41	0.13	35.27	35.27	0.00
10	<i>Barbara</i>	30.57	30.69	0.12	30.56	30.78	0.22	30.56	30.66	0.10
	<i>Boats</i>	31.06	31.17	0.11	31.07	31.28	0.21	31.07	31.17	0.10
	<i>House</i>	32.63	32.80	0.18	32.57	32.80	0.23	32.55	32.76	0.20
	<i>Lena</i>	30.90	31.00	0.10	30.89	31.07	0.18	30.88	30.95	0.07
	Average	31.29	31.41	0.12	31.27	31.48	0.21	31.26	31.38	0.12
15	<i>Barbara</i>	28.28	28.51	0.23	28.27	28.59	0.31	28.26	28.47	0.21
	<i>Boats</i>	28.83	29.05	0.22	28.82	29.12	0.31	28.80	29.02	0.22
	<i>House</i>	30.69	31.04	0.36	30.69	31.09	0.41	30.62	30.96	0.34
	<i>Lena</i>	28.71	28.87	0.17	28.69	28.92	0.23	28.69	28.86	0.18
	Average	29.12	29.37	0.24	29.12	29.43	0.32	29.09	29.33	0.24
20	<i>Barbara</i>	26.65	26.97	0.32	26.62	27.02	0.40	26.61	26.90	0.30
	<i>Boats</i>	27.23	27.56	0.33	27.21	27.62	0.42	27.16	27.51	0.35
	<i>House</i>	29.24	29.74	0.50	29.22	29.73	0.51	29.16	29.66	0.49
	<i>Lena</i>	27.23	27.49	0.25	27.15	27.49	0.34	27.15	27.45	0.30
	Average	27.59	27.94	0.35	27.55	27.97	0.42	27.52	27.88	0.36
25	<i>Barbara</i>	25.41	25.79	0.38	25.34	25.81	0.47	25.33	25.71	0.38
	<i>Boats</i>	25.97	26.40	0.44	25.93	26.46	0.53	25.91	26.39	0.48
	<i>House</i>	28.05	28.76	0.71	28.02	28.77	0.75	27.98	28.71	0.73
	<i>Lena</i>	26.00	26.41	0.41	25.98	26.45	0.47	25.95	26.40	0.45
	Average	26.36	26.84	0.48	26.32	26.87	0.55	26.29	26.80	0.51
50	<i>Barbara</i>	21.45	22.08	0.62	21.37	21.96	0.59	21.28	21.87	0.59
	<i>Boats</i>	21.96	22.68	0.72	21.85	22.52	0.67	21.77	22.47	0.70
	<i>House</i>	23.80	24.78	0.99	23.71	24.63	0.92	23.58	24.55	0.97
	<i>Lena</i>	22.17	22.92	0.75	22.11	22.82	0.70	22.04	22.77	0.73
	Average	22.35	23.11	0.77	22.26	22.98	0.72	22.17	22.91	0.75

ictionaries, the high-resolution images are downsampled by a factor of 3 to simulate the low-resolution images.

We test the proposed method on several standard test images which are commonly used in previous work on denoising and super-resolution. Four test images are shown in Fig. 5. To evaluate the performance of image denoising, we add the additive Gaussian white noise with 6 levels of known standard deviations σ into the original clean images. For image super-resolution, we first downsample the high-quality images by bicubic interpolation with a factor of 3, and add unit variance Gaussian noise to the downsampled images to obtain the noisy downsampled low-resolution images.

In our implementation, for denoising, the test images are divided into 8×8 patches with 7-pixel patch overlapping, which means that the overlaps between neighboring patches in Left, Up-Left, Up directions are $[8 \times 7]$, $[7 \times 7]$ and $[7 \times 8]$ (as illustrated in Fig. 2(h), (a), (b)), respectively. For $3 \times$ super-resolution, the patch sizes are set to be 9×9 and 3×3 in high- and low-resolution domain, respectively. The patch

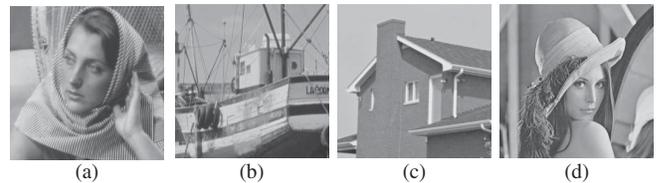


Fig. 5. Four examples of test images. (a) *Barbara*. (b) *Boats*. (c) *House*. (d) *Lena*.

overlapping in low-resolution domain is set to be 2-pixel. Therefore, it is 6-pixel patch overlapping in the high-resolution domain.

A. Image Denoising

1) *Denoising Over Redundant Dictionaries*: The proposed method does not rely on any assumption about the dictionary structure. To verify the effectiveness and robustness of the proposed method over different types of redundant

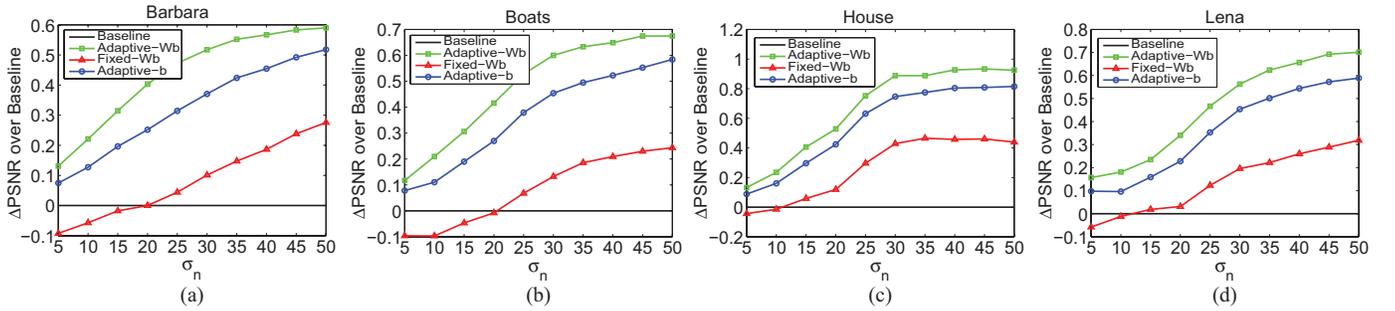


Fig. 6. PSNR curves of different models on the four test images with various noise levels. (a) *Barbara*. (b) *Boats*. (c) *House*. (d) *Lena*.

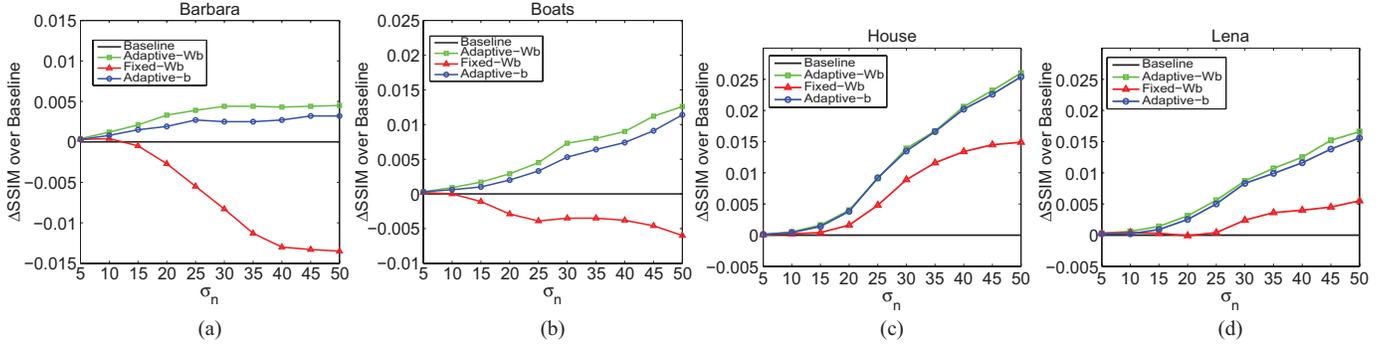


Fig. 7. SSIM curves of different models on the four test images with various noise levels. (a) *Barbara*. (b) *Boats*. (c) *House*. (d) *Lena*.

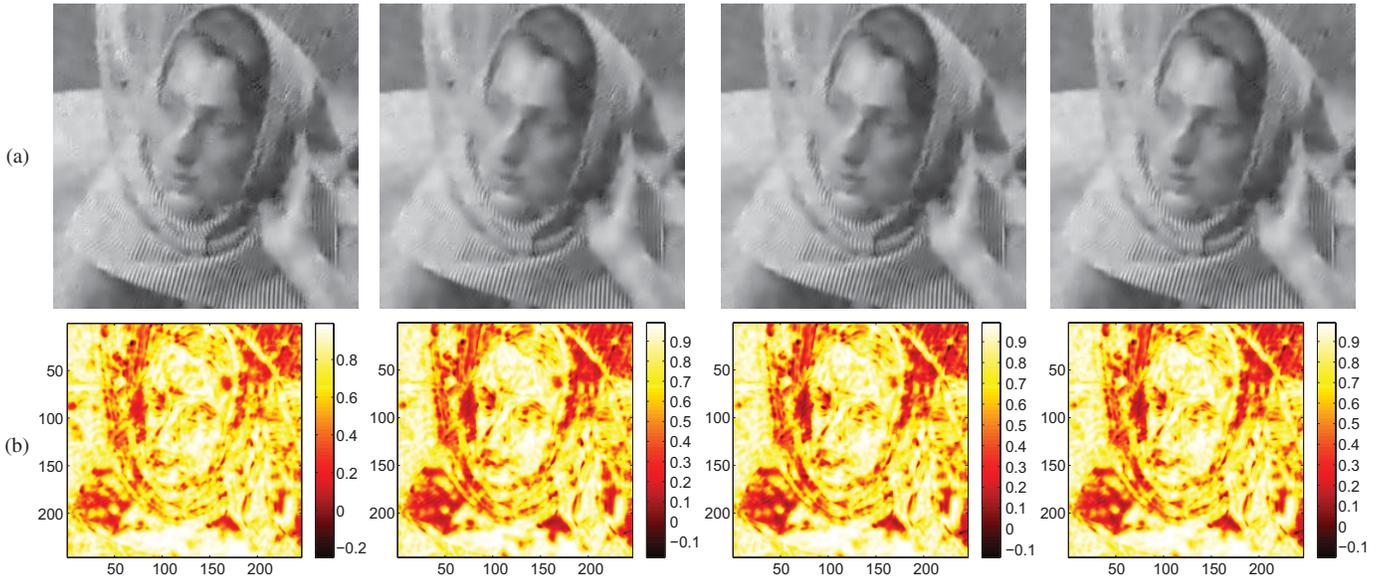


Fig. 8. Denoising results and corresponding SSIM maps of different prior models ($\sigma = 50$). (a) Denoised images. (b) SSIM maps. From left to right: plain OMP, fixed-Wb, adaptive-b, and adaptive-Wb model.

dictionaries, we conduct a set of experiments to compare the denoising performance of the proposed method (CASD) with the basic sparse coding algorithm, OMP. Three types of dictionaries are considered, including the pre-specified Redundant DCT dictionary and two dictionaries trained by K-SVD algorithm [27]. Specifically, these two dictionaries include a globally trained dictionary (denoted as Global K-SVD) and an adaptively trained dictionary (denoted as Adaptive K-SVD) learned from a general set of natural images and each input

image, respectively. The redundant ratio of the dictionary is set to be $r = 4$. Therefore, the dictionary size is $m = r \cdot 8^2 = 256$ for each dictionary.

The result is summarized in Table I. The proposed method outperforms the baseline in denoising with 6 levels of standard deviations ($\sigma = 5, 10, 15, 20, 25, 50$). It is robust to the dictionary type in terms of peak signal-to-noise ratio (PSNR) before patch averaging. Nevertheless, the effectiveness of the proposed context-aware prior is different. For Redundant DCT,

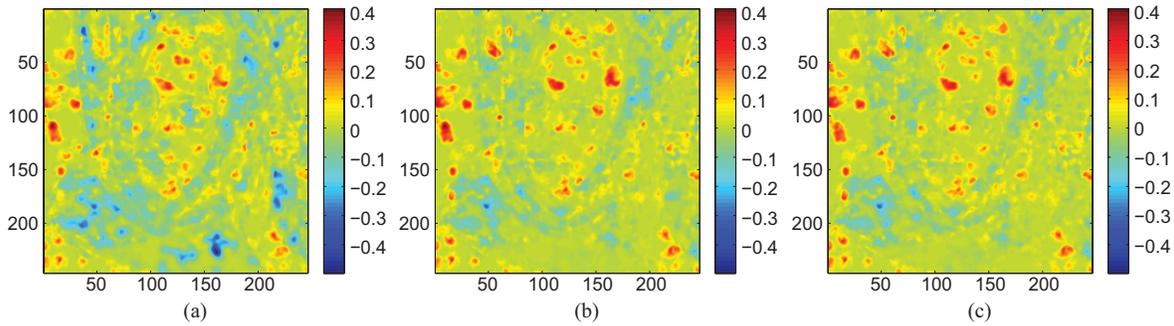


Fig. 9. SSIM gains of different prior models over the OMP method ($\sigma = 50$). (a) Fixed-Wb. (b) Adaptive-b. (c) Adaptive-Wb.

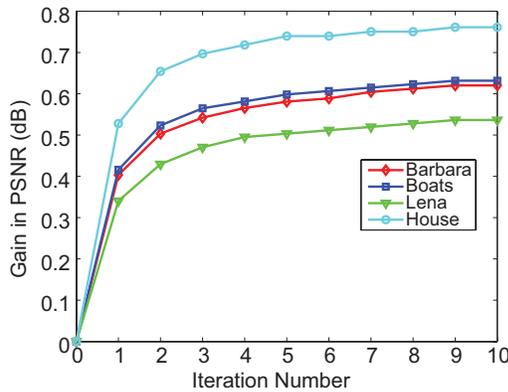


Fig. 10. Improvement in the denoising results after each iteration of the proposed numerical algorithm ($\sigma = 20$).

the average gains in PSNR are from 0.31 dB to 0.98 dB for 6 levels of noise, while for Global K-SVD and Adaptive K-SVD, the average gains are 0.19 ~ 0.77 dB and 0.14 ~ 0.72 dB, respectively.

2) *Effect of Redundant Ratio of Adaptive Dictionary:* To verify the robustness of the proposed method to dictionary size, we conduct the following experiment on Adaptive K-SVD dictionary. The dictionary size is $m = r \cdot 8^2$ in which the redundant ratio r is set to be 3–5, respectively. Over these adaptive dictionaries of different redundant ratios, the denoising results of the OMP method and the proposed CASD method are summarized in Table II (in terms of PSNR). The average PSNR gains of the proposed method with 6 standard noise deviations are 0.04 ~ 0.77 dB, 0.13 ~ 0.72 dB, 0.0 ~ 0.75 dB for redundant ratio $r = 3$ –5, respectively. Therefore, it is concluded that the proposed method is robust to the dictionary size. Since the proposed method is more effective when $r = 4$ in the low noise level, we use $r = 4$ redundant dictionary for the rest of the denoising experiments.

3) *Comparisons of Prior Models:* The proposed prior in (20) consists of two parts of model parameters \tilde{W}, b . We denote the proposed prior model as Adaptive-Wb. When the interaction matrix \tilde{W} is zero, the proposed prior model reduces to the independent adaptive model in (13), denoted as Adaptive-b. Meanwhile, Fixed-Wb represents the model with ($\tilde{W} = \mathbf{0}, b = \mathbf{0}$). It is a statistical approximation to the deterministic baseline algorithm, OMP. We analyze the contribution

TABLE III
SUMMARY OF DENOISING PSNR RESULTS ON FOUR METHODS IN DECIBELS. IN EACH CELL. TOP LEFT: RESULTS OF FoE [6]. TOP RIGHT: BM3D [41]. BOTTOM LEFT: K-SVD [2]. BOTTOM RIGHT: THE PROPOSED

σ	Barbara		Boats		House		Lena	
5	37.42	38.34	37.93	38.71	38.29	39.72	38.46	39.02
	38.16	38.24	38.41	38.45	39.34	39.38	38.64	38.67
10	32.97	34.75	33.65	34.95	35.11	36.64	34.21	35.17
	34.28	34.38	34.62	34.69	35.92	35.99	34.64	34.66
15	30.21	32.70	31.39	32.81	33.52	34.91	31.83	32.97
	32.10	32.22	32.39	32.46	34.2	34.31	32.45	32.46
20	28.19	31.20	29.79	31.29	32.24	33.75	29.97	31.45
	30.52	30.64	30.81	30.90	32.95	33.10	30.92	30.94
25	26.82	30.01	28.54	30.14	31.27	32.84	29.03	30.28
	29.23	29.33	29.58	29.69	31.94	32.18	29.77	29.81
50	23.39	26.33	24.61	26.47	27.50	29.71	24.96	26.77
	25.13	25.18	25.6	25.75	28.11	28.44	26.10	26.25

TABLE IV
COMPARISONS OF METHODS ON $3 \times$ SUPER-RESOLUTION IN TERMS OF PSNR (dB)

Method	Barb.	Boats	House	Lena	C.Man	Pepper	Avg
Bicubic	24.1	27.4	29.8	27.2	23.8	27.0	26.6
ScSR [4]	24.2	28.2	31.0	28.3	24.5	28.1	27.4
Proposed	24.3	28.7	31.5	28.7	24.7	28.5	27.7

of each component of the proposed context-aware prior model. As illustrated in Fig. 6, Adaptive-Wb and Adaptive-b models consistently outperform the baseline for the four test images in terms of PSNR before patch averaging. The gains over the baseline are approximately linear with the noise standard deviation when $\sigma \geq 15$. Specifically, Adaptive-b achieves most of the improvements over the baseline. However, the proposed context-aware model Adaptive-Wb further improves the recovery accuracy of denoising. For the Fixed-Wb model, it is comparable to the baseline when $\sigma \leq 20$ and superior to the baseline in the high noise levels ($\sigma \geq 20$).

In Fig. 7, the SSIM [40] results of the denoised images corresponding to different prior models and various noise levels are illustrated. Due to the patch averaging processing, the gains of the three prior models over the baseline are reduced. However, the superiorities of these prior models are



Fig. 11. Visual comparisons of denoised results with state-of-the-art methods. From left to right: original image, noisy image ($\sigma = 15$), FoE [6], BM3D [41], K-SVD [2], and the proposed method.

preserved. Note that there are exceptions for the prior model Fixed-Wb on *Barbara* and *Boats* images. Fig. 8 shows the final denoised images and their corresponding SSIM maps [40] in noise level $\sigma = 50$. Specifically, the SSIM gains of these prior models over the OMP method are illustrated in Fig. 9. There are more serious structure distortion in the stripe texture region in the denoised image of Fixed-Wb model than that of the OMP method.

4) *Iteration Number*: In the numerical solution presented in Section V, an iterative scheme is proposed to solve the joint estimation problem. Fig. 10 demonstrates the improvement over the initial recovery result in terms of PSNR (before patch averaging). As illustrated in the Fig. 10, all curves start at zero, going towards positive values. The initial result is marked as Iteration Number 0. After the first iteration, there is a significant performance gain in PSNR for each image ($0.3 \sim 0.5$ dB for different images). The following iterations give additional but slight improvements over the first iteration (about 0.2 dB after 10 iterations). Therefore, we use the first iteration result as a trade-off between recovery quality and computational complexity.

5) *Comparisons to the State-of-the-Art Methods*: Finally, we compare the proposed denoising method over Adaptive K-SVD dictionary with the typical state-of-the-art methods, including FoE [6], K-SVD [2] and BM3D [41]. Table III summarizes these denoising results for the three methods and the proposed method. As demonstrated in Table III, the average PSNR results of FoE method and K-SVD method are 31.30 dB

and 32.32 dB, computed on all test images results with six noise levels. The average result of the proposed method is 32.42 dB, showing average improvements of 1.12 dB and 0.1 dB in favor of FoE and K-SVD, respectively. BM3D method performs the best in average result of 32.54 dB which exploits the nonlocal similarity in natural images beyond the sparsity prior model. Integrating the self-similarity property in natural images into the proposed method is one of our future work. The visual comparison of the proposed method with these state-of-the-art methods in noise level $\sigma = 15$ is illustrated in Fig. 11. The proposed method significantly alleviates the visual artifacts of K-SVD method and reveals more texture details (especially in the brick texture region).

B. Image Super-Resolution

In this subsection, we present the experimental results of $3\times$ single-image super-resolution with noise level $\sigma = 1$. We compare the proposed method with bicubic method and the sparsity-based method (ScSR) in [4]. The dictionary size for both ScSR [4] and the proposed method are set to be 1024.

For objective evaluation, Table IV shows the PSNR results of the three methods. From Table IV, the sparsity-based method [4] is superior to the bicubic method about 0.8 dB in average and the proposed method further improves the performance of sparsity-based baseline about 0.3 dB. We also illustrate the visual quality of SR results for subjective assessment in Fig. 12. The figures show that the bicubic



Fig. 12. Visual comparisons of super-resolution results of different methods. From left to right: original image, low-resolution image, ScSR [4], and the proposed method.

method constantly produces the lowest visual quality with blurring and jaggling artifacts along the edge regions. The sparsity-based method [4] produces SR images with sharp edges. However, there are some artifacts around the edges. It is mainly because of the noise in the low-resolution input image and the separate reconstruction of each patch without considering the structural correlations of neighboring patches. The proposed method alleviates the artifacts by incorporating the structural correlations of dictionary atoms into the sparse coding algorithm. Richer texture can also be super-resolved by the proposed method, *e.g.*, the hat of Lena image.

VII. CONCLUSION

In this paper, we propose a context-aware prior model and a MRF-based global extension to improve the sparsity-based prior model. To exploit the contextual information of neighboring image patches, we explicitly model the structural correlation of the sparsity patterns for overlapped patches. Specifically, the probabilistic dependencies between dictionary atoms in adjacent spatial locations are modeled as in the BM model. MRF-based modeling is utilized to tune the local prior into a global one for dealing with arbitrary size images. In order to solve the joint estimation of model parameters and

sparsity recovery, iterative numerical solution is presented to approximate the global optimization result. We conduct several experiments on image denoising and super-resolution to evaluate the effectiveness and robustness of the proposed method. Experimental results show that the proposed method stably improves the performance of sparsity-based prior model.

REFERENCES

- [1] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.
- [2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [3] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [4] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [5] M. Elad, J.-L. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, 2005.
- [6] S. Roth and M. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 860–867.

- [7] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.
- [8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [9] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [10] E. Simoncelli and E. Adelson, "Noise removal via bayesian wavelet coring," in *Proc. Int. Conf. Image Process.*, Sep. 1996, pp. 379–382.
- [11] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 1993, pp. 40–44.
- [12] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.
- [14] M. Duarte, M. Wakin, and R. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden Markov tree model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2008, pp. 5137–5140.
- [15] P. J. Garrigues and B. A. Olshausen, "Learning horizontal connections in a sparse coding model of natural images," in *Proc. Adv. Neural Inform. Process. Syst.*, 2008, pp. 505–512.
- [16] V. Cevher, M. F. Duarte, C. Hedge, and R. G. Baraniuk, "Sparse signal recovery using Markov random fields," in *Proc. Adv. Neural Inform. Process. Syst.*, 2009, pp. 257–264.
- [17] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 417–424.
- [18] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," in *Proc. IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sep. 2009.
- [19] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [20] T. Fator, Y. Eldar, and M. Elad, "Denoising of image patches via sparse representations with learned statistical dependencies," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 5820–5823.
- [21] T. Peleg, Y. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2286–2303, May 2012.
- [22] R. Rubinstein, A. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
- [23] E. J. Candes and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Ann. Statist.*, vol. 30, no. 3, pp. 784–842, 2002.
- [24] M. Do and M. Vetterli, "Contourlets: A directional multiresolution image representation," in *Proc. Int. Conf. Image Process.*, 2002, pp. 357–360.
- [25] G. Easley, D. Labate, and W.-Q. Lim, "Optimally sparse image representations using shearlets," in *Proc. 14th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2006, pp. 974–978.
- [26] K. Egan, S. Aase, and J. Hakon-Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Aug. 1999, pp. 2443–2446.
- [27] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [28] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 801–808.
- [29] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 689–696.
- [30] O. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: Theory," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 539–554, Mar. 2006.
- [31] O. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II: Adaptive algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 555–571, Mar. 2006.
- [32] O. G. Guleryuz, "Weighted averaging for denoising with overcomplete dictionaries," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 3020–3034, Dec. 2007.
- [33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1993.
- [34] J. S. Turek, I. Yavneh, and M. Elad, "On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3526–3535, Aug. 2011.
- [35] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct. 2009.
- [36] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, S. Singh, Ed., New York: Berlin, Springer-Verlag, 2009.
- [37] A. Hyvärinen, "Consistency of pseudolikelihood estimation of fully visible boltzmann machines," *Neural Comput.*, vol. 18, no. 10, pp. 2283–2292, 2006.
- [38] G. Narkiss and M. Zibulevsky, "Sequential subspace optimization method for large-scale unconstrained problems," EE Dept., Technion-Israel Institute of Technology, Haifa, Israel, Tech. Rep. 32000, 2005.
- [39] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Conf. Curves Surfaces*, 2010, pp. 711–730.
- [40] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [41] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.



Jie Ren was born in Shenyang, China, in 1984. He received the B.S. (Hons.) degree in computer science from Beijing Normal University, Beijing, China, in 2007. He is currently pursuing the Ph.D. degree in computer science with the Institute of Computer Science and Technology, Peking University, Beijing.

His current research interests include image and video processing, super-resolution, denoising, and sparse representation.



Jiaying Liu (S'09–M'10) received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, and the Ph.D. degree with the Best Graduate Honor in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively.

She was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. Her current research

interests include image processing, sparse signal representation, and video compression.



Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively.

He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication.

Dr. Guo is the Executive Member of the China-Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State

Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008, and the Government Allowance granted by the State Council in 2009.