

Image Restoration via Simultaneous Sparse Coding: Where Structured Sparsity Meets Gaussian Scale Mixture

Weisheng Dong · Guangming Shi · Yi Ma · Xin Li

Received: 30 January 2014 / Accepted: 3 February 2015 / Published online: 18 February 2015
© Springer Science+Business Media New York 2015

Abstract In image processing, sparse coding has been known to be relevant to both variational and Bayesian approaches. The regularization parameter in variational image restoration is intrinsically connected with the shape parameter of sparse coefficients' distribution in Bayesian methods. How to set those parameters in a principled yet spatially adaptive fashion turns out to be a challenging problem especially for the class of nonlocal image models. In this work, we propose a structured sparse coding framework to address this issue—more specifically, a nonlocal extension of Gaussian scale mixture (GSM) model is developed using simultaneous sparse coding (SSC) and its applications into image restoration are explored. It is shown that the variances of sparse coefficients (the field of scalar multipliers of Gaussians)—if treated as a latent variable—can be jointly estimated along with the unknown sparse coefficients via the method of alternating optimization. When applied to image restoration, our experimental results have shown that the proposed SSC–GSM technique can both preserve the sharpness

of edges and suppress undesirable artifacts. Thanks to its capability of achieving a better spatial adaptation, SSC–GSM based image restoration often delivers reconstructed images with higher subjective/objective qualities than other competing approaches.

Keywords Simultaneous sparse coding · Gaussian scale mixture · Structured sparsity · Alternative minimization · Variational image restoration

1 Introduction

1.1 Background and Motivation

Sparse representation of signals/images has been widely studied by signal and image processing communities in the past decades. Historically, sparsity dated back to the idea of variable selection studied by statisticians in late 1970s (Hocking 1976) and coring operator invented by RCA researchers in early 1980s (Carlson et al. 1985). The birth of wavelet (Daubechies 1988) (a.k.a. filter bank theory (Vetterli 1986) or multi-resolution analysis (Mallat 1989) in late 1980s rapidly sparked the interest in sparse representation, which has found successful applications into image coding (Shapiro 1993; Said and Pearlman 1996; Taubman and Marcellin 2001) and denoising (Donoho and Johnstone 1994; Mihcak and Ramchandran 1999; Chang et al. 2000). Under the framework of sparse coding, a lot of research have been centered at two related issues: basis functions (or dictionary) and statistical modeling of sparse coefficients. Exemplar studies of the former are the construction of directional multiresolution representation (e.g., contourlet (Do and Vetterli 2005)) and over-complete dictionary learning from training data (e.g., K-SVD (Aharon et al. 2012; Elad

Communicated by Julien Mairal, Francis Bach, Michael Elad.

W. Dong (✉) · G. Shi
School of Electronic Engineering, Xidian University,
Xi'an 710071, China
e-mail: wsdong@mail.xidian.edu.cn

G. Shi
e-mail: gmshi@xidian.edu.cn

Y. Ma
School of Information Science and Technology,
ShanghaiTech University, Shanghai, China
e-mail: mayi@shanghaitech.edu.cn

X. Li
Lane Department of CSEE, West Virginia University,
Morgantown, WV 26506-6109, USA
e-mail: xin.li@ieee.org

and Aharon 2012), multiscale dictionary learning (Mairal et al. 2008), online dictionary learning (Mairal et al. 2009a) and the non-parametric Bayesian dictionary learning (Zhou et al. 2009)); the latter includes the use of Gaussian mixture models (Zoran and Weiss 2011; Yu et al. 2012), variational Bayesian models (Ji et al. 2008; Wipf et al. 2011), universal models (Ramirez and Sapiro 2012), and centralized Laplacian model (Dong et al. 2013b) for sparse coefficients.

More recently, a class of nonlocal image restoration techniques (Buades et al. 2005; Dabov et al. 2007; Mairal et al. 2009b; Katkovnik et al. 2010; Dong et al. 2011, 2013a, b) have attracted increasingly more attention. The key motivation behind lies in the observation that many important image structures in natural images including edges and textures can be characterized by the abundance of self-repeating patterns. Such observation has led to the formulation of simultaneous sparse coding (SSC) (Mairal et al. 2009b). Our own recent works (Dong et al. 2011, 2013a) have continued to demonstrate the potential of exploiting SSC in image restoration. However, a fundamental question remains open: *how to achieve (local) spatial adaptation within the framework of nonlocal image restoration?* This question is related to the issue of regularization parameters in a variational setting or shape parameters in a Bayesian one; but the issue becomes even more thorny when it is tangled with nonlocal regularization/sparsity. To the best of our knowledge, how to tune those parameters in a principled manner remains an open problem (e.g. please refer to (Ramirez and Sapiro 2012) and its references for a survey of recent advances).

In this work, we propose a new image model named SSC–GSM that connects Gaussian scale mixture (GSM) with SSC. Our idea is to model each sparse coefficient as a Gaussian distribution with a positive scaling variable and impose a sparse distribution prior (i.e., the Jeffrey prior (Box and Tiao 2011) used in this work) over the positive scaling variables. We show that the maximum a posteriori (MAP) estimation of both sparse coefficients and scaling variables can be efficiently calculated by the method of alternating minimization. By characterizing the set of sparse coefficients of similar patches with the *same* prior distribution (i.e., the same non-zero means and positive scaling variables), we can effectively exploit both local and nonlocal dependencies among the sparse coefficients, which have been shown important for image restoration applications (Dabov et al. 2007; Mairal et al. 2009b; Dong et al. 2011). Our experimental results have shown that SSC–GSM based image restoration can deliver images whose subjective and objective qualities are often higher than other competing methods. Visual quality improvements are attributed to better preservation of edge sharpness and suppression of undesirable artifacts in the restored images.

1.2 Relationship to Other Competing Approaches

The connection between sparse coding and Bayesian inference has been previously studied in sparse Bayesian learning (Tipping 2001; Wipf and Rao 2004, 2007) and more recently in Bayesian compressive sensing (Ji et al. 2008), latent variable Bayesian models for promoting sparsity (Wipf et al. 2011). Despite offering a generic theoretical foundation as well as promising results, the Bayesian inference techniques along this line of research often involve potentially expensive sampling (e.g., approximated solutions for some choice of prior are achieved in Wipf et al. (2011)). By contrast, our SSC–GSM approach is conceptually much simpler and admits analytical solutions involving iterative shrinkage/filtering operators only. The other works closely related to the proposed SSC–GSM are group sparse coding with a Laplacian scale mixture (LSM) (Garrigues and Olshausen 2010) and field-of-GSM (Lyu and Simoncelli 2009). In (Garrigues and Olshausen 2010), a LSM model with Gamma distribution imposed over the scaling variables was used to model the sparse coefficients. Approximated estimates of the scale variables were obtained using the Expectation-Maximization (EM) algorithm. Note that the scale variables derived in LSM (Garrigues and Olshausen 2010) is very similar to the weights derived in the reweighted l_1 -norm minimization (Candes et al. 2008). In contrast to those approaches, a GSM model with nonzero means and a noninformative sparse prior imposed over scaling variables are used to model the sparse coefficients here. Instead of using the EM algorithm for an approximated solution, our SSC–GSM offers an efficient inference of both scaling variables and sparse coefficients via alternating optimization method. In Lyu and Simoncelli (2009), a field of FSM (FoGSM) model was constructed using the product of two independent homogeneous Gaussian Markov random fields (hGMRFs) to exploit the dependencies among adjacent blocks. Despite similar motivations to exploit the dependencies between the scaling variables, the techniques used in Lyu and Simoncelli (2009) is significantly different and requires a lot more computations than our SSC–GSM formulation.

The proposed work is related to non-parametric Bayesian dictionary learning (Zhou et al. 2012) or solving inverse problems with piecewise linear estimators (Yu et al. 2012) but ours is motivated by the connection between non-local SSC strategy and local parametric GSM model. When compared with previous works on image denoising (e.g., K-SVD denoising (Elad and Aharon 2012), spatially adaptive singular-value thresholding (Dong et al. 2013a) and Expected Patch Log Likelihood (EPLL) (Zoran and Weiss 2011)), SSC–GSM targets at a more general framework of combining adaptive sparse inference with dictionary learning. SSC–GSM based image deblurring

has also been experimentally shown superior to existing patch-based methods (e.g., Iterative Decoupled Deblurring BM3D (IDD-BM3D) (Danielyan et al. 2012) and Nonlocal centralized sparse representation (NCSR) (Dong et al. 2013b)) and the gain in terms of ISNR is as much as 1.5dB over IDD-BM3D (Danielyan et al. 2012) for some test images (e.g., butterfly image - please refer to Fig. 5).

The remainder of this paper is organized as follows. In Sect. 2, we formulate the sparse coding problem with GSM model and generalize it into the new SSC–GSM model. In Sect. 3, we elaborate on the details of how to solve SSC–GSM by alternative minimization and emphasize the analytical solutions for both subproblems. In Sect. 4, we study the application of SSC–GSM into image restoration and discuss efficient implementation of SSC–GSM based image restoration algorithms. In Sect. 5, we report our experimental results in image denoising, image deblurring and image super-resolution as supporting evidence for the effectiveness of SSC–GSM model. In Sect. 6, we make some conclusions about the relationship of sparse coding to image restoration as well as perspectives about the future directions.

2 Simultaneous Sparse Coding via Gaussian Scale Mixture Modeling

2.1 Sparse Coding with Gaussian Scale Mixture Model

The basic idea behind sparse coding is to represent a signal $\mathbf{x} \in \mathbb{R}^n$ (n is the size of image patches) as the linear combination of basis vectors (dictionary elements) $\mathbf{D}\boldsymbol{\alpha}$ where $\mathbf{D} \in \mathbb{R}^{n \times K}$, $n \leq K$ is the dictionary and the coefficients $\boldsymbol{\alpha} \in \mathbb{R}^K$ satisfies some sparsity constraint. In view of the challenge with l_0 -optimization, it has been suggested that the original nonconvex optimization is replaced by its l_1 -counterpart:

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \tag{1}$$

which is convex and easier to solve. Solving the l_1 -norm minimization problem corresponds to the MAP inference of $\boldsymbol{\alpha}$ with an identically independent distributed (i.i.d) Laplacian prior $P(\alpha_i) = \frac{1}{2\theta_i} e^{-\frac{|\alpha_i|}{\theta_i}}$, wherein θ_i denotes the standard derivation of α_i . It is easy to verify that the regularization parameter should be set as $\lambda_i = 2\sigma_n^2/\theta_i$ when the i.i.d Laplacian prior is used, where σ_n^2 denotes the variance of approximation errors. In practice, the variances θ_i 's of each α_i are unknown and may not be easy to accurately estimated from the observation \mathbf{x} considering that real signal/image are non-stationary and may be degraded by noise and blur.

In this paper we propose to model sparse coefficients $\boldsymbol{\alpha}$ with a GSM (Andrews and Mallows 1974) model. The GSM model decomposes coefficient vector $\boldsymbol{\alpha}$ into the point-wise product of a Gaussian vector $\boldsymbol{\beta}$ and a hidden scalar multiplier $\boldsymbol{\theta}$ -i.e., $\alpha_i = \theta_i \beta_i$, where θ_i is the positive scaling variable with probability $P(\theta_i)$. Conditioned on θ_i , a coefficient α_i is Gaussian with the standard derivation of θ_i . Assuming that θ_i are i.i.d and independent of β_i , the GSM prior of $\boldsymbol{\alpha}$ can be expressed as

$$P(\boldsymbol{\alpha}) = \prod_i P(\alpha_i), \quad P(\alpha_i) = \int_0^\infty P(\alpha_i|\theta_i)P(\theta_i)d\theta_i. \tag{2}$$

As a family of probabilistic distributions, the GSM model can contain many kurtotic distributions (e.g., the Laplacian, Generalized Gaussian, and student's t-distribution) given an appropriate $P(\theta_i)$.

Note that for most of choices of $P(\theta_i)$ there is no analytical expression of $P(\alpha_i)$ and thus it is difficult to compute the MAP estimates of α_i . However, such difficulty can be overcome by joint estimation of (α_i, θ_i) . For a given observation $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{n}$, where $\mathbf{n} \sim N(0, \sigma_n^2)$ denotes the additive Gaussian noise, we can formulate the following MAP estimator

$$\begin{aligned} (\boldsymbol{\alpha}, \boldsymbol{\theta}) &= \operatorname{argmax} \log P(\mathbf{x}|\boldsymbol{\alpha}, \boldsymbol{\theta})P(\boldsymbol{\alpha}, \boldsymbol{\theta}) \\ &= \operatorname{argmax} \log P(\mathbf{x}|\boldsymbol{\alpha}) + \log P(\boldsymbol{\alpha}|\boldsymbol{\theta}) + \log P(\boldsymbol{\theta}), \end{aligned} \tag{3}$$

where $P(\mathbf{x}|\boldsymbol{\alpha})$ is the likelihood term characterized by Gaussian function with variance σ_n^2 . The prior term $P(\boldsymbol{\alpha}|\boldsymbol{\theta})$ can be expressed as

$$P(\boldsymbol{\alpha}|\boldsymbol{\theta}) = \prod_i P(\alpha_i|\theta_i) = \prod_i \frac{1}{\theta_i \sqrt{2\pi}} \exp\left(-\frac{(\alpha_i - \mu_i)^2}{2\theta_i^2}\right). \tag{4}$$

Instead of assuming the mean $\mu_i = 0$, we propose to use a biased-mean μ_i for α_i here inspired by our previous work NCSR (Dong et al. 2013b) (its estimation will be elaborated later).

The adoption of GSM model allows us to generalize the sparsity from statistical modeling of sparse coefficients $\boldsymbol{\alpha}$ to the specification of sparse prior $P(\boldsymbol{\theta})$. It has been suggested in the literature that noninformative prior (Box and Tiao 2011) $P(\theta_i) \approx \frac{1}{\theta_i}$ —a.k.a. Jeffrey's prior—is often the favorable choice. Therefore, we have also adopted this option in this work, which translates Eq. (3) into

$$\begin{aligned} (\boldsymbol{\alpha}, \boldsymbol{\theta}) &= \operatorname{argmin}_{\boldsymbol{\alpha}, \boldsymbol{\theta}} \frac{1}{2\sigma_n^2} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \sum_i \log(\theta_i \sqrt{2\pi}) \\ &\quad + \sum_i \frac{(\alpha_i - \mu_i)^2}{2\theta_i^2} + \sum_i \log \theta_i, \end{aligned} \tag{5}$$

where we have used $P(\boldsymbol{\theta}) = \sum_i P(\theta_i)$. Noting that Jeffrey's prior is unstable as $\theta_i \rightarrow 0$; so we replace $\log \theta_i$ by $\log(\theta_i + \epsilon)$

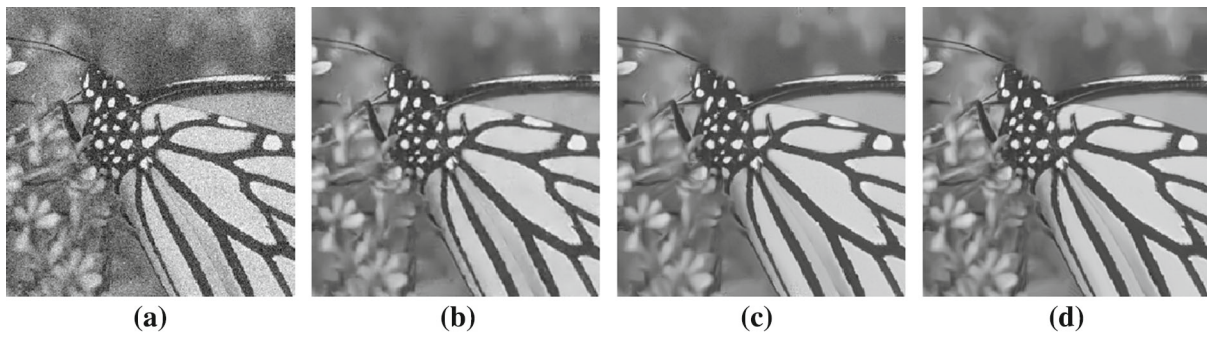


Fig. 1 Denoising performance comparison between the variants of the proposed method. **(a)** Noisy image ($\sigma_n = 20$); **(b)** KSVD (Elad and Aharon 2012) (PSNR=29.90 dB); **(c)** the proposed method (without

nonlocal extension) uses the model of Eq. (7) (PSNR=30.18 dB); **(d)** the proposed SSC-GSM method (PSNR=30.84 dB)

Table 1 Parameters setting for each experiment

| | Denoising | Deblurring | | Super-resolution | |
|--------|-----------|--------------|-------------|------------------|-------|
| | | Unifor. blur | Gauss. blur | Noiseless | Noisy |
| η | 0.35 | 0.11 | 0.12 | 0.05 | 1.25 |

where ϵ is a small positive number for numerical stability and rewrite $\sum_i \log(\theta_i + \epsilon)$ into $\log(\theta + \epsilon)$ for notational simplicity. The above equation can then be further translated into the following sparse coding problem

$$(\alpha, \theta) = \underset{\alpha, \theta}{\operatorname{argmin}} \|x - D\alpha\|_2^2 + 4\sigma_n^2 \log(\theta + \epsilon) + \sigma_n^2 \sum_i \frac{(\alpha_i - \mu_i)^2}{\theta_i^2}. \tag{6}$$

Note that the matrix form of original GSM model is $\alpha = \Lambda\beta$ and $\mu = \Lambda\gamma$ where $\Lambda = \operatorname{diag}(\theta_i) \in \mathbb{R}^{K \times K}$ is a diagonal matrix characterizing the variance field for a chosen image patch. Accordingly, the sparse coding problem in Eq. (6) can be translated from (α, μ) domain to (β, γ) domain as follows

$$(\beta, \theta) = \underset{\beta, \theta}{\operatorname{argmin}} \|x - D\Lambda\beta\|_2^2 + 4\sigma_n^2 \log(\theta + \epsilon) + \sigma_n^2 \|\beta - \gamma\|_2^2. \tag{7}$$

In other words, the sparse coding formulation of GSM model boils down to the joint estimation of β and θ . But unlike (Portilla et al. 2003) that treats the multiplier as a hidden variable and cancel it out through integration (i.e., the derivation of Bayes Least-Square estimate), we explicitly use the field of Gaussian scalar multiplier to characterize the variability and dependencies among local variances. Such sparse coding formulation of GSM model is appealing because it allows us to further exploit the power of GSM by connecting it with structured sparsity as we will detail next.

2.2 Exploiting Structured Sparsity for the Estimation of the Field of Scalar multipliers

A key observation behind our approach is that for a collection of similar patches, their corresponding sparse coefficients α 's should be characterized by the *same prior* i.e., the probability density function with the same θ and μ . Therefore, if one considers the SSC of GSM models for a collection of m similar patches, the structured/group sparsity based extension of Eq. (7) can be written as

$$(\mathbf{B}, \theta) = \underset{\mathbf{B}, \theta}{\operatorname{argmin}} \|X - D\Lambda\mathbf{B}\|_F^2 + 4\sigma_n^2 \log(\theta + \epsilon) + \sigma_n^2 \|\mathbf{B} - \Gamma\|_F^2, \tag{8}$$

where $X = [x_1, \dots, x_m]$ denotes the collection of m similar patches¹, $\mathbf{A} = \Lambda\mathbf{B}$ is the group representation of GSM model for sparse coefficients and their corresponding first-order and second-order statistics are characterized by $\Gamma = [\gamma_1, \dots, \gamma_m] \in \mathbb{R}^{K \times m}$ and $\mathbf{B} = [\beta_1, \dots, \beta_m] \in \mathbb{R}^{K \times m}$ respectively, wherein $\gamma_j = \mu, j = 1, 2, \dots, m$. Given a collection of m similar patches, we have adopted the nonlocal means approach (Buades et al. 2005) for estimating μ

$$\mu = \sum_{j=1}^m w_j \alpha_j, \tag{9}$$

where $w_j \sim \exp(-\|x - x_j\|_2^2/h)$ is the weighting coefficient based on patch similarity. It follows from $\mu = \Lambda\gamma$ that

$$\gamma = \sum_{j=1}^m w_j \Lambda^{-1} \alpha_j = \sum_{j=1}^m w_j \beta_j. \tag{10}$$

¹ Throughout this paper, we will use subscript/superscript to denote column/row vectors of a matrix respectively.

Table 2 The PSNR (dB) results by different denoising methods

| σ_w | 5 | 10 | 15 | 20 | 50 | 100 | | | | | | |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------|--------------|
| Lena | 38.86 | 38.68 | 36.07 | 35.83 | 34.43 | 34.14 | 33.20 | 32.88 | 29.07 | 28.95 | 25.37 | 25.96 |
| | 38.70 | 38.85 | 35.81 | 35.96 | 34.09 | 34.23 | 32.92 | 33.08 | 28.42 | 29.05 | 25.66 | 25.91 |
| Monarch | 38.69 | 38.53 | 34.74 | 34.48 | 32.46 | 32.15 | 30.92 | 30.58 | 26.28 | 25.59 | 22.31 | 21.82 |
| | 38.49 | 38.74 | 34.57 | 34.82 | 32.34 | 32.52 | 30.69 | 30.84 | 25.68 | 26.02 | 22.05 | 22.52 |
| Barbara | 38.38 | 38.44 | 35.07 | 34.95 | 33.27 | 32.96 | 31.97 | 31.53 | 27.51 | 27.13 | 23.05 | 23.56 |
| | 38.36 | 38.65 | 34.98 | 35.27 | 33.02 | 33.32 | 31.72 | 32.06 | 27.10 | 27.60 | 23.30 | 24.05 |
| Boat | 37.50 | 37.34 | 34.10 | 33.99 | 32.29 | 32.17 | 31.02 | 30.87 | 26.89 | 26.76 | 23.71 | 23.94 |
| | 37.35 | 37.42 | 33.90 | 33.95 | 32.03 | 32.11 | 30.74 | 30.82 | 26.60 | 26.79 | 23.64 | 23.90 |
| C. Man | 38.54 | 38.24 | 34.52 | 34.14 | 32.31 | 31.96 | 30.86 | 30.54 | 26.59 | 26.36 | 22.91 | 23.14 |
| | 38.17 | 38.39 | 34.12 | 34.28 | 31.99 | 32.03 | 30.48 | 30.50 | 26.16 | 26.29 | 22.89 | 23.23 |
| Couple | 37.60 | 37.41 | 34.13 | 33.96 | 32.20 | 32.06 | 30.83 | 30.70 | 26.48 | 26.31 | 23.19 | 23.34 |
| | 37.44 | 37.51 | 33.94 | 33.94 | 31.95 | 31.98 | 30.56 | 30.63 | 26.21 | 26.41 | 23.22 | 23.36 |
| F. Print | 36.67 | 36.71 | 32.65 | 32.57 | 30.46 | 30.31 | 28.97 | 28.78 | 24.53 | 24.21 | 21.07 | 21.18 |
| | 36.81 | 36.84 | 32.70 | 32.63 | 30.46 | 30.36 | 28.99 | 28.87 | 24.53 | 24.50 | 21.29 | 21.54 |
| Hill | 37.31 | 37.16 | 33.84 | 33.68 | 32.06 | 31.89 | 30.85 | 30.71 | 27.13 | 26.99 | 24.10 | 24.30 |
| | 37.17 | 37.23 | 33.69 | 33.70 | 31.86 | 31.89 | 30.61 | 30.69 | 26.86 | 27.05 | 24.13 | 24.24 |
| House | 40.13 | 40.00 | 37.06 | 37.05 | 35.31 | 35.32 | 34.03 | 34.16 | 29.53 | 29.90 | 25.20 | 25.63 |
| | 39.91 | 40.02 | 36.80 | 36.79 | 35.11 | 35.03 | 33.97 | 34.00 | 29.63 | 30.36 | 25.65 | 26.70 |
| Man | 37.99 | 37.84 | 34.18 | 34.03 | 32.12 | 31.98 | 30.73 | 30.60 | 26.84 | 26.72 | 23.86 | 24.00 |
| | 37.78 | 37.91 | 33.96 | 34.06 | 31.89 | 31.99 | 30.52 | 30.60 | 26.60 | 26.76 | 23.97 | 24.02 |
| Peppers | 38.30 | 38.15 | 34.94 | 34.80 | 33.01 | 32.87 | 31.61 | 31.47 | 26.94 | 26.87 | 23.05 | 23.14 |
| | 38.06 | 38.22 | 34.66 | 34.83 | 32.70 | 32.87 | 31.26 | 31.41 | 26.53 | 26.82 | 22.64 | 23.34 |
| Straw | 35.81 | 35.92 | 31.46 | 31.39 | 29.13 | 28.95 | 27.52 | 27.36 | 22.79 | 22.67 | 19.42 | 19.50 |
| | 35.87 | 36.04 | 31.50 | 31.56 | 29.13 | 29.16 | 27.50 | 27.51 | 22.48 | 22.84 | 19.23 | 19.52 |
| Average | 37.98 | 37.87 | 34.40 | 34.24 | 32.42 | 32.23 | 31.04 | 30.85 | 26.71 | 26.54 | 23.10 | 23.29 |
| | 37.84 | 37.98 | 34.22 | 34.32 | 32.21 | 32.29 | 30.83 | 30.92 | 26.44 | 26.71 | 23.14 | 23.53 |

In each cell, the results of the four denoising methods are reported in the following order: top left-BM3D-SAPCA (Katkovnik et al. 2010); top right-LSSC (Mairal et al. 2009b); bottom left-NCSR (Dong et al. 2013b); bottom right-proposed SSC-GSM. The highest PSNR values among four are highlighted in bold in each cell

A practical issue of Eq. (10) is that the original patches x_j and x , as well as the sparse codes α_j and α are not available, and thus we cannot directly compute γ using the Eq. (10). To avoid such difficulty, we can treat γ as another optimization variable and jointly estimate it with the sparse coefficients as

$$(\mathbf{B}, \theta, \gamma) = \underset{\mathbf{B}, \theta, \gamma}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{DAB}\|_F^2 + 4\sigma_n^2 \log(\theta + \epsilon) + \sigma_n^2 \|\mathbf{B} - \Gamma\|_F^2, \text{ s. t. } \gamma = \mathbf{B}\mathbf{w}, \quad (11)$$

where the weights $\mathbf{w} = [w_1, \dots, w_m]^T$ are pre-computed using the initial estimate of the image. Using the alternative directional multiplier method (ADMM) (Boyd et al. 2011), Eq. (11) can be approximately solved. However, the computational complexity for solving the sub-problem of γ is high. Alternatively, we can overcome such difficulty by iteratively estimating γ from the current estimates of the sparse coefficients without any sacrifice of the performance. Let

$\beta_j = \hat{\beta}_j + e_j$, wherein e_j denotes the estimation error of β_j and is assumed to be Gaussian and zero-mean. Then, Eq. (10) can be re-expressed as

$$\gamma = \sum_{j=1}^m w_j \hat{\beta}_j + \sum_{j=1}^m e_j = \hat{\gamma} + n_w, \quad (12)$$

where n_w denotes the estimation error of γ . As e_j is assumed to be zero-mean Gaussian, n_w would be small. Thus, γ can be readily obtained from the estimates of representation coefficients β_j . In practice, we recursively compute γ using the previous estimates of β_j after each iteration.

We call such new formulation in Eq. (8) *Simultaneous Sparse Coding for Gaussian Scalar Mixture* (SSC-GSM) and propose to develop computationally efficient solution to this problem in the next section. Note that here the formulation of SSC-GSM in Eq. (8) is for a given dictionary \mathbf{D} . However, the dictionary \mathbf{D} can also be optimized for a fixed pair of (\mathbf{B}, θ) such that both dictionary learning and statisti-

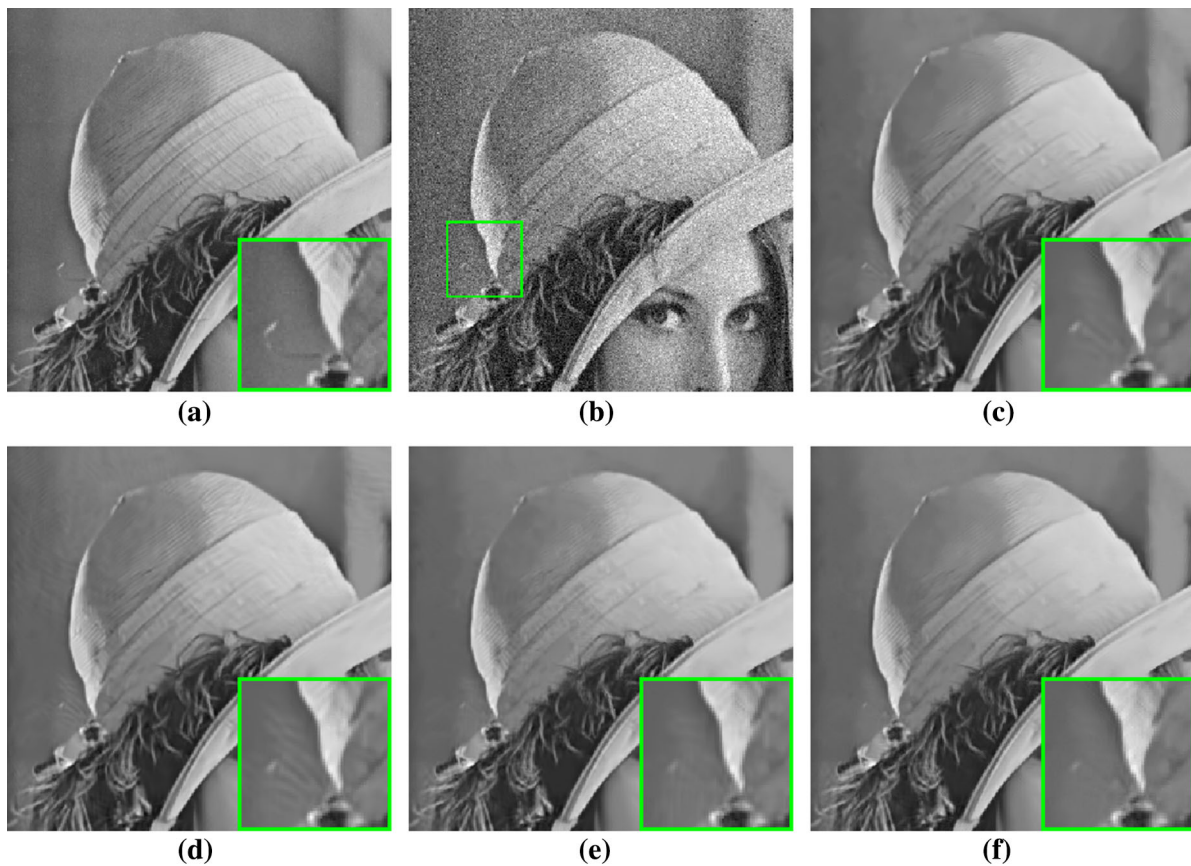


Fig. 2 Denoising performance comparison on the *Lena* image with moderate noise corruption. (a) Original image; (b) Noisy image ($\sigma_n = 20$); denoised images by (c) BM3D-SAPCA (Katkovnik et al. 2010) (PSNR=33.20dB, SSIM=0.8803); (d) LSSC (Mairal et al.

2009b) (PSNR=32.88dB, SSIM=0.8742); (e) NCSR (Dong et al. 2013b) (PSNR=32.92dB, SSIM=0.8760); (f) Proposed SSC-GSM (PSNR=33.08, SSIM=0.8787)

cal modeling of sparse coefficients can be unified within the framework of Eq. (8). Figure 1 shows the denoising results by the two proposed methods that use the model of Eqs. (7) and (8), respectively. Both the two proposed methods use the same dictionary D and γ . From Fig. 1 we can see that the proposed method without nonlocal extension outperforms the KSVD method (Elad and Aharon 2012). With the nonlocal extension, the denoising performance of the proposed method is significantly improved.

3 Solving Simultaneous Sparse Coding via Alternating Minimization

In this section, we will show how to solve the optimization problem in Eq. (8) by alternatively updating the estimates of B and θ . The key observation lies in that the two subproblems - minimization of B for a fixed θ and minimization of θ for a fixed B - both can be efficiently solved. Specifically, both subproblems admits closed-form solutions when the dictionary is orthogonal.

3.1 Solving θ for a Fixed B

For a fixed B , the first subproblem simply becomes

$$\theta = \operatorname{argmin}_{\theta} \|X - D\Lambda B\|_F^2 + 4\sigma_n^2 \log(\theta + \epsilon), \quad (13)$$

which can be rewritten as

$$\begin{aligned} \theta &= \operatorname{argmin}_{\theta} \|X - \sum_{i=1}^K d_i \beta^i \theta_i\|_F^2 + 4\sigma_n^2 \log(\theta + \epsilon) \\ &= \operatorname{argmin}_{\theta} \|\tilde{x} - \tilde{D}\theta\|_2^2 + 4\sigma_n^2 \log(\theta + \epsilon), \end{aligned} \quad (14)$$

where the long vector $\tilde{x} \in \mathbb{R}^{mn}$ denotes the vectorization of the matrix X , the matrix $\tilde{D} = [\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_K] \in \mathbb{R}^{mn \times K}$ whose each column \tilde{d}_j denotes the vectorization of the rank-one matrix $d_j \beta^j$, and $\beta^i \in \mathbb{R}^m$ denotes the i -th row of matrix B . For optimizing the nonconvex log penalty in Eq. (14), the principled difference of convex functions (DC) programming approach can be used for a local minimum (Gasso et al. 2009). It has also been shown in Candes et al. (2008) that the log penalty can be linearly approximated and thus a local min-

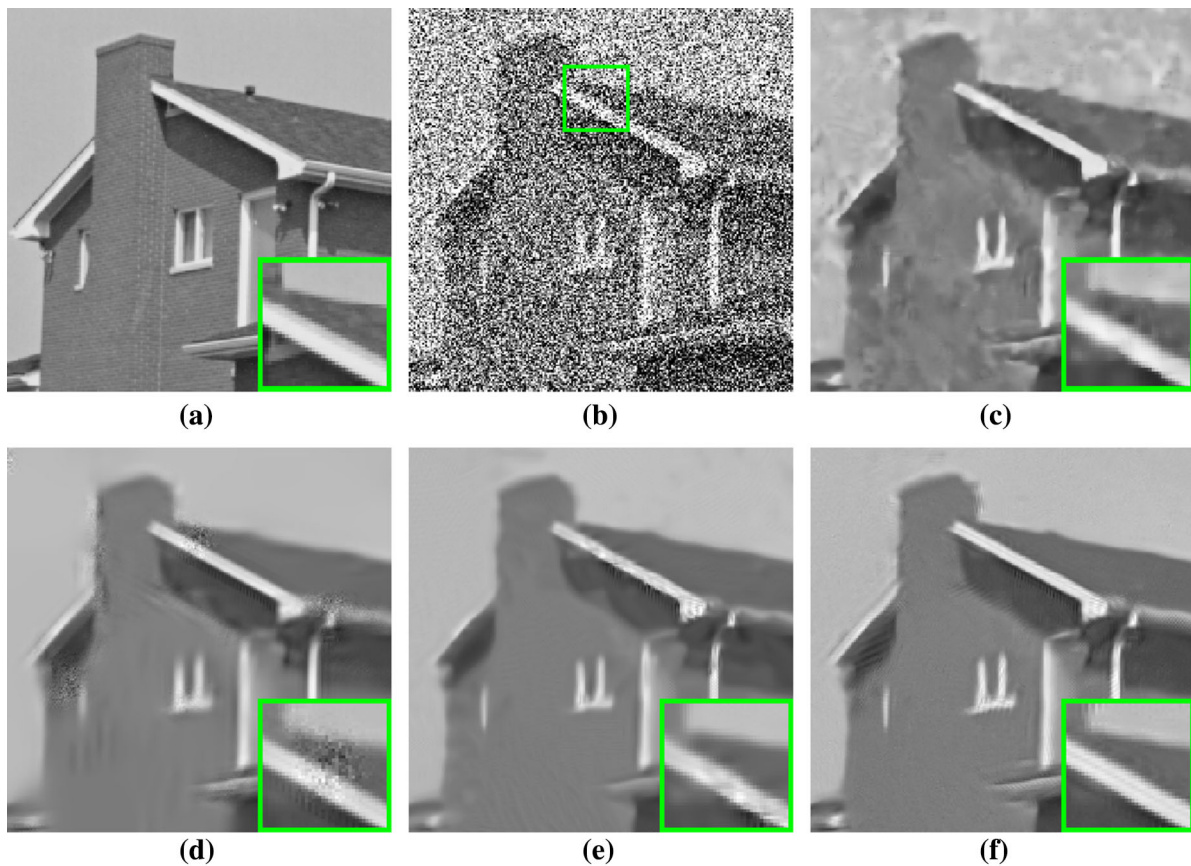


Fig. 3 Denoising performance comparison on the *House* image with strong noise corruption. (a) Original image; (b) Noisy image ($\sigma_n = 100$); denoised images by (c) BM3D-SAPCA (Katkovnik et al. 2010) (PSNR=35.20dB, SSIM=0.6767); (d) LSSC (Mairal et al. 2009b) (PSNR=25.63 dB, SSIM=0.7389); (e) NCSR (Dong et al. 2013b) (PSNR=25.65 dB, SSIM=0.7434); (f) Proposed SSC-GSM (PSNR=26.70, SSIM=0.7430)

imum of the nonconvex objective function can be obtained by iteratively solving a weighted ℓ_1 -penalized optimization problem.

However, the optimization of Eq. (13) can be much simplified when the dictionary D is orthogonal (e.g., DCT or PCA basis). In the case of orthogonal dictionary, Eq. (13) can be rewritten as

$$\theta = \underset{\theta}{\operatorname{argmin}} \|A - \Lambda B\|_F^2 + 4\sigma_n^2 \log(\theta + \epsilon), \tag{15}$$

where we have used $X = DA$. For expression convenience, we can rewrite Eq. (15) as

$$\theta = \underset{\theta}{\operatorname{argmin}} \sum_i a_i \theta_i^2 + b_i \theta_i + c \log \theta_i + \epsilon, \tag{16}$$

where $a_i = \|\beta^i\|_2^2$, $b_i = -2\alpha^i (\beta^i)^T$ and $c = 4\sigma_n^2$. Hence, Eq. (16) can be decomposed into a sequence of scalar minimization problem, i.e.,

$$\theta_i = \underset{\theta_i}{\operatorname{argmin}} a_i \theta_i^2 + b_i \theta_i + c \log(\theta_i + \epsilon), \tag{17}$$

which can be solved by taking $\frac{df(\theta_i)}{d\theta_i} = 0$, where $f(\theta_i)$ denotes the right hand side of Eq. (17). We derive

$$g(\theta_i) = \frac{df(\theta_i)}{d\theta_i} = 2a_i \theta_i + b_i + \frac{c}{\theta_i + \epsilon}. \tag{18}$$

By solving $g(\theta_i) = 0$, we obtain the following two stationary points of $f(\theta_i)$, i.e.,

$$\theta_{i,1} = -\frac{b_i}{4a_i} + \sqrt{\frac{b_i^2}{16} - \frac{c}{2a_i}}, \quad \theta_{i,2} = -\frac{b_i}{4a_i} - \sqrt{\frac{b_i^2}{16} - \frac{c}{2a_i}}, \tag{19}$$

when $b_i^2/(16a_i^2) - c/(2a_i) \geq 0$. Then, the global minimizer of $f(\theta_i)$ can be obtained by comparing $f(0)$, $f(\theta_{i,1})$ and $f(\theta_{i,2})$.

When $b_i^2/(16a_i^2) - c/(2a_i) < 0$, there does not exist any stationary points in the range of $[0, \infty)$. As ϵ is a small positive constant, $g(0) = b_i + c/\epsilon$ is always positive. Thus, $f(0)$ is the global minimizer for this case. In summary, the solution to Eq. (17) can be written as

Table 3 PSNR(dB) and SSIM results of the deblurred images

| Images | 9×9 uniform blur, $\sigma_n = \sqrt{2}$ | | | | | | | | | | | |
|--|--|---------------|---------------|-----------------|---------------|---------------|----------------|----------------|---------------|---------------|----------------|--|
| | <i>Butterfly</i> | <i>Boats</i> | <i>C. Man</i> | <i>Starfish</i> | <i>Parrot</i> | <i>Lena</i> | <i>Barbara</i> | <i>Peppers</i> | <i>Leaves</i> | <i>House</i> | <i>Average</i> | |
| FISTA (Beck and Teboulle 2009) | 28.37 | 29.04 | 26.82 | 27.75 | 29.11 | 28.33 | 25.75 | 28.43 | 26.49 | 31.99 | 28.21 | |
| | 0.9058 | 0.8355 | 0.8278 | 0.8200 | 0.8750 | 0.8274 | 0.7440 | 0.8134 | 0.9023 | 0.8490 | 0.8400 | |
| IDD-BM3D (Danielyan et al. 2012) | 29.21 | 31.20 | 28.56 | 29.48 | 31.06 | 29.70 | 27.98 | 29.62 | 29.38 | 34.44 | 30.06 | |
| | 0.9216 | 0.8820 | 0.8580 | 0.8640 | 0.9041 | 0.8654 | 0.8225 | 0.8422 | 0.9418 | 0.8786 | 0.8780 | |
| NCSR (Dong et al. 2013b) | 29.68 | 31.08 | 28.62 | 30.28 | 31.95 | 29.96 | 28.10 | 29.66 | 29.98 | 34.31 | 30.36 | |
| | 0.9273 | 0.8810 | 0.8574 | 0.8807 | 0.9103 | 0.8676 | 0.8255 | 0.8402 | 0.9485 | 0.8755 | 0.8814 | |
| Proposed SSC-GSM | 30.45 | 31.36 | 28.83 | 30.58 | 32.05 | 30.11 | 28.78 | 29.79 | 30.83 | 34.31 | 30.71 | |
| | 0.9377 | 0.8918 | 0.8669 | 0.8862 | 0.9145 | 0.8783 | 0.8465 | 0.8491 | 0.9582 | 0.8748 | 0.8904 | |
| Gaussian blur with standard deviation 1.6, $\sigma_n = \sqrt{2}$ | | | | | | | | | | | | |
| FISTA (Beck and Teboulle 2009) | 30.36 | 29.36 | 26.80 | 29.65 | 31.23 | 29.47 | 25.03 | 29.42 | 29.33 | 31.50 | 29.22 | |
| | 0.9374 | 0.8509 | 0.8241 | 0.8878 | 0.9066 | 0.8537 | 0.7377 | 0.8349 | 0.9480 | 0.8254 | 0.8606 | |
| IDD-BM3D (Danielyan et al. 2012) | 30.73 | 31.68 | 28.17 | 31.66 | 32.89 | 31.45 | 27.19 | 29.99 | 31.40 | 34.08 | 30.92 | |
| | 0.9469 | 0.9036 | 0.8705 | 0.9156 | 0.9319 | 0.9103 | 0.8231 | 0.8806 | 0.9639 | 0.8820 | 0.9029 | |
| NCSR (Dong et al. 2013b) | 30.84 | 31.49 | 28.34 | 32.27 | 33.39 | 31.26 | 27.91 | 30.16 | 31.57 | 33.63 | 31.09 | |
| | 0.9476 | 0.8968 | 0.8591 | 0.9229 | 0.9354 | 0.9009 | 0.8304 | 0.8704 | 0.9648 | 0.8696 | 0.8998 | |
| Proposed SSC-GSM | 31.12 | 31.78 | 28.40 | 32.26 | 33.30 | 31.52 | 28.42 | 30.18 | 32.02 | 34.65 | 31.37 | |
| | 0.9522 | 0.9054 | 0.8719 | 0.9245 | 0.9377 | 0.9109 | 0.8462 | 0.8770 | 0.9693 | 0.8834 | 0.9079 | |

$$\theta_i = \begin{cases} 0, & \text{if } b_i^2/(16a_i^2) - c/(2a_i) < 0, \\ v_i, & \text{otherwise} \end{cases} \quad (20)$$

where

$$v_i = \underset{\theta_i}{\operatorname{argmin}} \{f(0), f(\theta_{i,1}), f(\theta_{i,2})\}. \quad (21)$$

3.2 Solving \mathbf{B} for a Fixed $\boldsymbol{\theta}$

The second subproblem is in fact even easier to solve. It takes the following form

$$\mathbf{B} = \underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\mathbf{B}\|_F^2 + \sigma_n^2 \|\mathbf{B} - \Gamma\|_F^2. \quad (22)$$

Since both terms are l_2 , the closed-form solution to Eq. (22) is essentially the classical Wiener filtering

$$\mathbf{B} = (\hat{\mathbf{D}}^T \hat{\mathbf{D}} + \sigma_n^2 \mathbf{I})^{-1} (\hat{\mathbf{D}}^T \mathbf{X} + \Gamma), \quad (23)$$

where $\hat{\mathbf{D}} = \mathbf{D}\mathbf{A}$. Note that when \mathbf{D} is orthogonal, Eq. (23) can be further simplified into

$$\mathbf{B} = (\mathbf{A}^T \mathbf{A} + \sigma_n^2 \mathbf{I})^{-1} (\mathbf{A}^T \mathbf{X} + \Gamma), \quad (24)$$

where $\mathbf{A}^T \mathbf{A} + \sigma_n^2 \mathbf{I}$ is a diagonal matrix and therefore its inverse can be easily computed.

By alternatively solving both subproblems of Eqs. (13) and (22) for the estimates of \mathbf{A} and \mathbf{B} , the image data matrix \mathbf{X} can then be reconstructed as

$$\hat{\mathbf{X}} = \mathbf{D}\hat{\mathbf{A}}\hat{\mathbf{B}}, \quad (25)$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ denotes the final estimates of \mathbf{A} and \mathbf{B} .

4 Application of Bayesian Structured Sparse Coding into Image Restoration

In the previous sections, we have seen how to solve SSC-GSM problem for a single image data matrix \mathbf{X} (a collection of image patches similar to a chosen exemplar). In this section, we generalize such formulation to whole-image reconstruction and study the applications of SSC-GSM into image restoration including image denoising, image deblurring and image superresolution. The standard image degradation model is used here: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^M$ denotes the original and degraded images respectively, $\mathbf{H} \in \mathbb{R}^{N \times M}$ is the degradation matrix and \mathbf{w} is additive white Gaussian noise observing $N(0, \sigma_n^2)$. The whole-image reconstruction problem can be formulated as

$$\begin{aligned} (\mathbf{x}, \{\mathbf{B}_l\}, \{\boldsymbol{\theta}_l\}) = \underset{\mathbf{x}, \{\mathbf{B}_l\}, \{\boldsymbol{\theta}_l\}}{\operatorname{argmin}} & \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \\ & + \sum_{l=1}^L \{\eta \|\tilde{\mathbf{R}}\mathbf{x} - \mathbf{D}\mathbf{A}_l \mathbf{B}_l\|_F^2 \\ & + \sigma_n^2 \|\mathbf{B} - \Gamma\|_F^2 + 4\sigma_n^2 \log(\boldsymbol{\theta}_l + \epsilon)\}, \end{aligned} \quad (26)$$

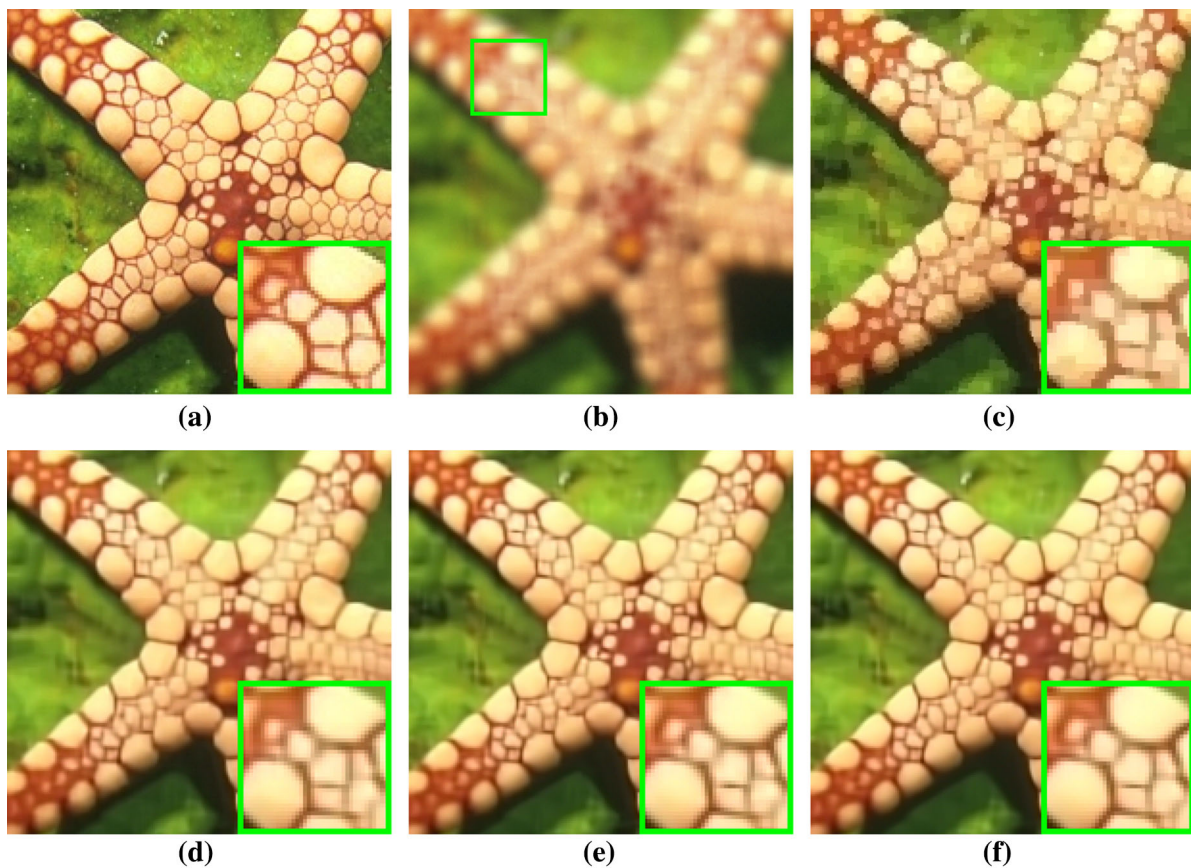


Fig. 4 Deblurring performance comparison on the *Starfish* image. (a) Original image; (b) Noisy and blurred image (9×9 uniform blur, $\sigma_n = \sqrt{2}$); deblurred images by (c) FISTA (Beck and Teboulle 2009) (PSNR=27.75 dB, SSIM=0.8200); (d) IDD-BM3D (Danielyan

et al. 2012) (PSNR=29.48 dB, SSIM=0.8640); (e) NCSR (Dong et al. 2013b) (PSNR=30.28 dB, SSIM=0.8807); (f) Proposed SSC-GSM (PSNR=30.58 dB, SSIM=0.8862)

where $\tilde{\mathbf{R}}_l \mathbf{x} \doteq [\mathbf{R}_{l_1} \mathbf{x}, \mathbf{R}_{l_2} \mathbf{x}, \dots, \mathbf{R}_{l_m} \mathbf{x}] \in \mathbb{R}^{n \times m}$ denotes the data matrix formed by a group of image patches similar to the l -th exemplar patch \mathbf{x}_l (including \mathbf{x}_l itself), $\mathbf{R}_l \in \mathbb{R}^{n \times N}$ denotes a matrix extracting the l -th patch \mathbf{x}_l from \mathbf{x} , and L is the total number of exemplars extracted from the reconstructed image \mathbf{x} . For a given exemplar patch \mathbf{x}_l , we search similar patches by performing k -nearest-neighbor search within a large local window (e.g., 40×40). As the original image is not available, we use the current estimate of original image for patch matching, i.e.,

$$S = \{l_j \mid \|\hat{\mathbf{x}}_l - \hat{\mathbf{x}}_{l_j}\|_2^2 < T\}, \tag{27}$$

where T denotes the pre-selected threshold and S denotes the collection of positions of those similar patches. Alternatively, we can form the sample set S by selecting the patches that are within the first m ($m = 40$ in our implementation) closest to $\hat{\mathbf{x}}_l$. Invoking the principle of alternative optimization again, we propose to solve the whole-image reconstruction problem in Eq. (26) by alternating the solutions to the following two subproblems.

4.1 Solving \mathbf{x} for a Fixed $\{\mathbf{B}_l\}, \{\theta_l\}$

Let $\hat{\mathbf{X}}_l = \mathbf{D}\mathbf{A}_l\mathbf{B}_l$. When $\{\mathbf{B}_l\}$ and $\{\theta_l\}$ are fixed, so is $\{\hat{\mathbf{X}}_l\}$. Therefore, Eq. (26) reduces to the following l_2 -optimization problem

$$\mathbf{x} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \sum_{l=1}^L \eta \|\mathbf{R}\mathbf{x}_l - \hat{\mathbf{X}}_l\|_F^2, \tag{28}$$

which admits the following closed-form solution

$$\mathbf{x} = \left(\mathbf{H}^T \mathbf{H} + \eta \sum_{l=1}^L \tilde{\mathbf{R}}_l^T \tilde{\mathbf{R}}_l \right)^{-1} \left(\mathbf{H}^T \mathbf{y} + \eta \sum_{l=1}^L \tilde{\mathbf{R}}_l^T \hat{\mathbf{X}}_l \right), \tag{29}$$

where $\tilde{\mathbf{R}}_l^T \tilde{\mathbf{R}}_l \doteq \sum_{j=1}^m \mathbf{R}_j^T \mathbf{R}_j$, $\tilde{\mathbf{R}}_l^T \hat{\mathbf{X}}_l \doteq \sum_{j=1}^m \mathbf{R}_j^T \hat{\mathbf{x}}_{l_j}$ and $\hat{\mathbf{x}}_{l_j}$ denotes the j -th column of matrix $\hat{\mathbf{X}}_l$. Note that for image denoising application where $\mathbf{H} = \mathbf{I}$ —the matrix to be inverted in Eq. (29)—is diagonal, and its inverse can be computed easily. Similar to the K-SVD approach, Eq. (29) can be computed by weighted averaging each reconstructed patches sets $\hat{\mathbf{X}}_l$. For image deblurring and super-resolution applications,

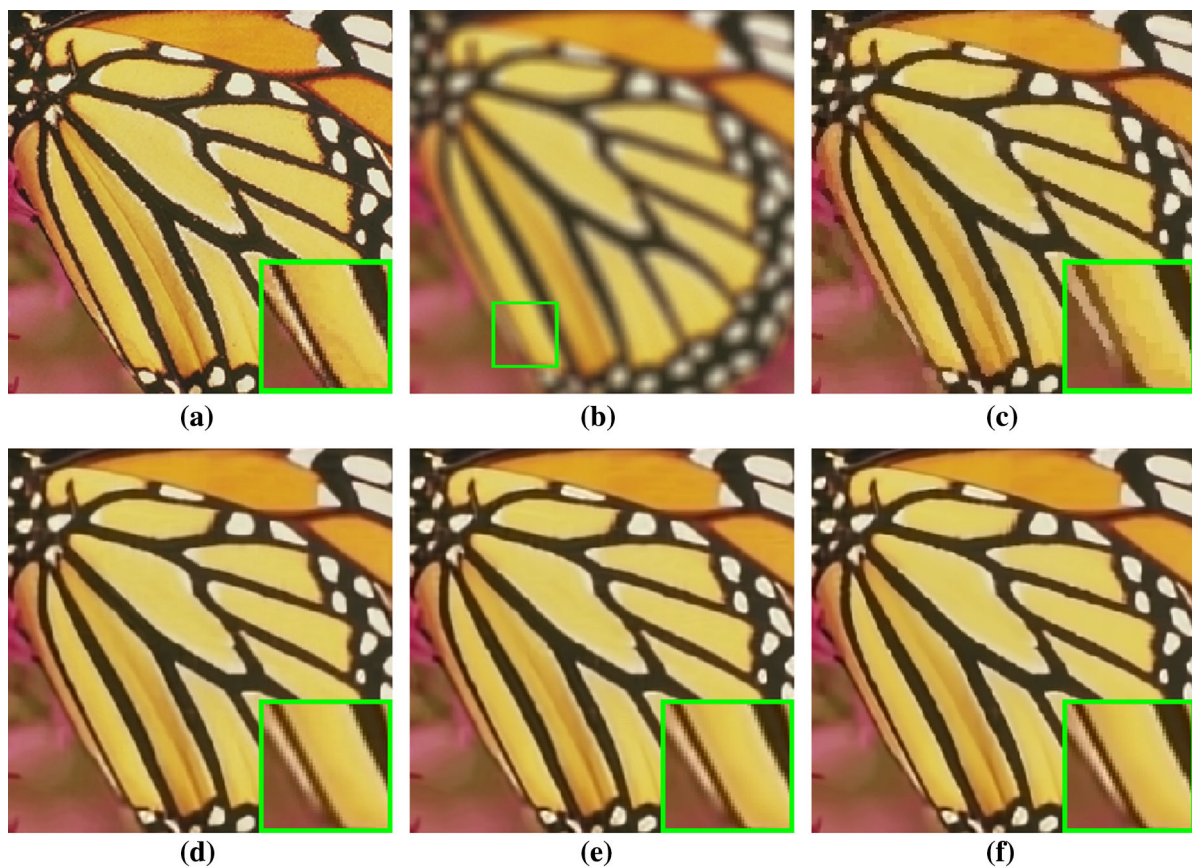


Fig. 5 Deblurring performance comparison on the *Butterfly* image. (a) Original image; (b) Noisy and blurred image (9×9 uniform blur, $\sigma_n = \sqrt{2}$); deblurred images by (c) FISTA (Beck and Teboulle 2009) (PSNR=28.37 dB, SSIM=0.9058); (d) IDD-BM3D (Danielyan

et al. 2012) (PSNR=29.21 dB, SSIM=0.9216); (e) NCSR (Dong et al. 2013b) (PSNR=29.68 dB, SSIM=0.9273); (f) Proposed SSC-GSM (PSNR=30.45 dB, SSIM=0.9377)

Eq. (29) can be computed by using a conjugate gradient (CG) algorithm.

4.2 Solving $\{\mathbf{B}_l\}$, $\{\theta_l\}$ for a Fixed \mathbf{x}

When \mathbf{x} is fixed, the first term in Eq. (26) goes away and the subproblem boils down to a sequence of patch-level SSC-GSM problems formulated for each exemplar - i.e.,

$$(\mathbf{B}_l, \theta_l) = \underset{\mathbf{B}_l, \theta_l}{\operatorname{argmin}} \|\mathbf{X}_l - \mathbf{D}\mathbf{\Lambda}_l\mathbf{B}_l\|_F^2 + \frac{\sigma_n^2}{\eta} \|\mathbf{B} - \Gamma\|_F^2 + \frac{4\sigma_n^2}{\eta} \log(\theta_l + \epsilon), \quad (30)$$

where we use $\mathbf{X}_l = \tilde{\mathbf{R}}_l\mathbf{x}$. This is exactly the problem we have studied in the previous section.

One important issue of the SSC-GSM-based image restoration is the selection of the dictionary. To adapt to the local image structures, instead of learning an over-complete dictionary for each dataset \mathbf{X}_l as in Mairal et al. (2009b),

we learn the principle component analysis (PCA) based dictionary for each dataset here (similar to NCSR (Dong et al. 2013b)). The use of the orthogonal dictionary much simplifies the Bayesian inference of SSC-GSM. Putting things together, a complete image restoration based on SSC-GSM can be summarized as follows.

In **Algorithm 1** we update \mathbf{D}_l in every k_0 to save computational complexity. We also found that **Algorithm 1** empirically converges even when the inner loop executes only one iteration (i.e., $J = 1$). We note that the above algorithm can lead to a variety of implementations depending the choice of degradation matrix \mathbf{H} . When \mathbf{H} is an identity matrix, **Algorithm 1** is an image denoising algorithm using iterative regularization technique (Xu and Osher 2007). When \mathbf{H} is a blur matrix or reduced blur matrix, Eq. (26) becomes the standard formulation of non-blind image deblurring or image super-resolution problem. The capability of capturing rapidly-changing statistics in natural images - e.g., through the use of GSM - can make patch-based nonlocal image models even more powerful.

Algorithm 1 SSC–GSM based Image Restoration**• Initialization:**

(a) set the initial estimate as $\hat{x} = y$ for image denoising and deblurring; or initialize \hat{x} by bicubic interpolation for image super-resolution;

(b) Set parameters η ;

(c) Obtain data matrices $\{X_l\}$'s from \hat{x} (through kNN search) for each exemplar and compute the PCA basis $\{D_l\}$ for each X_l .

• Outer loop (solve Eq. (26) by alternative optimization): Iterate on $k = 1, 2, \dots, k_{max}$

(a) Image-to-patch transformation: obtain data matrices $\{X_l\}$'s for each exemplar;

(b) Estimate biased means y using Eq. (10) for each X_l ;

(c) **Inner loop** (solve Eq. (30) for each data X_l): iterate on $J = 1, 2, \dots, J$;

(I) update θ_l for fixed B_l using Eq. (20);

(II) update B_l for fixed θ_l using Eq. (24);

End for

(d) Reconstruct X_l 's from θ_l and B_l using Eq. (25);

(e) If $\text{mod}(k, k_0) = 0$, update the PCA basis $\{D_l\}$ for each X_l ;

(f) Patch-to-image transformation: obtain reconstructed $\hat{x}^{(k+1)}$ from $\{X_l\}$'s by solving Eq. (29);

End for

• Output: $\hat{x}^{(k+1)}$.

5 Experimental Results

In this section, we report our experimental results of applying SSC–GSM based image restoration into image denoising, image deblurring and super-resolution. The experimental setup of this work is similar to that in our previous work on NCSR (Dong et al. 2013b). The basic parameter setting of SSC–GSM is as follows: patch size— 6×6 , number of similar blocks— $K = 44$; $k_{max} = 14$, $k_0 = 1$ for image denoising, and $k_{max} = 450$, $k_0 = 40$ for image deblurring and super-resolution. The regularization parameter η is empirically set. Its values are shown in Table 1. To evaluate the quality of restored images, both PSNR and SSIM (Wang et al. 2004) metrics are used. However, due to limited page space, we can only show part of the experimental results in this paper. More detailed comparisons and complete experimental results are available at the following website: http://see.xidian.edu.cn/faculty/wsdong/SSC_GSM.htm.

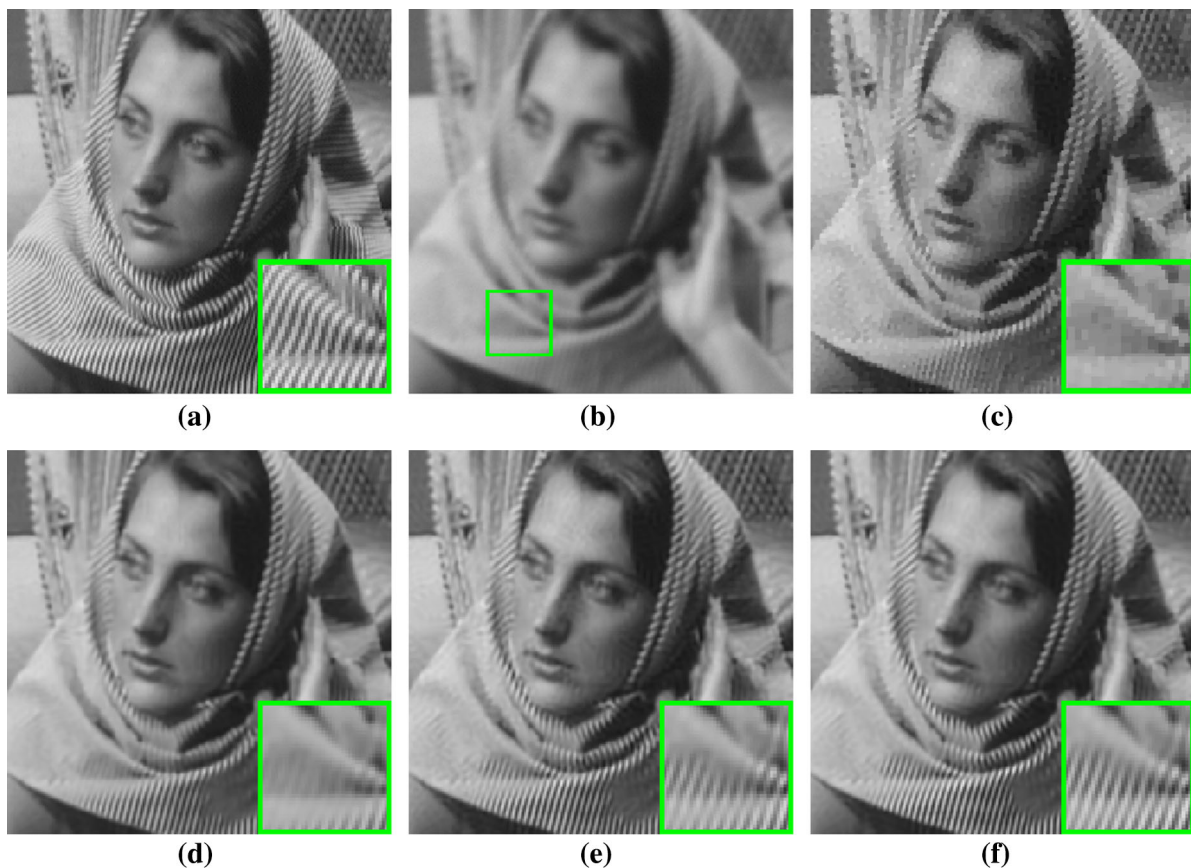


Fig. 6 Deblurring performance comparison on the *Barbara* image. (a) Original image; (b) Noisy and blurred image (Gaussian blur, $\sigma_n = \sqrt{2}$); deblurred images by (c) FISTA (Beck and Teboulle 2009) (PSNR=25.03 dB, SSIM=0.7377); (d) IDD-BM3D (Danielyan

et al. 2012) (PSNR=27.19 dB, SSIM=0.8231); (e) NCSR (Dong et al. 2013b) (PSNR=27.91 dB, SSIM=0.8304); (f) Proposed SSC–GSM (PSNR=28.42 dB, SSIM=0.8462)

Table 4 PSNR(dB) and SSIM results(luminance components) of the reconstructed HR images

| Images | Noiseless | | | | | | | | | |
|------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | Butterfly | Parrot | Plants | Hat | flower | Raccoon | Bike | Pathenon | Girl | Average |
| TV (Marquina and Osher 2008) | 26.56 | 27.85 | 0.8797 | 29.20 | 27.51 | 27.54 | 23.66 | 26.00 | 31.24 | 27.88 |
| | 0.9000 | 0.8900 | 0.8909 | 0.8483 | 0.8148 | 0.7070 | 0.7582 | 0.7232 | 0.7880 | 0.8121 |
| Sparsity (Yang et al. 2010) | 24.70 | 28.70 | 31.55 | 29.63 | 27.87 | 28.51 | 23.23 | 26.27 | 32.87 | 28.15 |
| | 0.8170 | 0.8823 | 0.8715 | 0.8288 | 0.7963 | 0.7273 | 0.7212 | 0.7025 | 0.8017 | 0.7943 |
| NCSR (Dong et al. 2013b) | 28.10 | 30.50 | 34.00 | 31.27 | 29.50 | 29.28 | 24.74 | 27.19 | 33.65 | 29.80 |
| | 0.9156 | 0.9144 | 0.9180 | 0.8699 | 0.8558 | 0.7706 | 0.8027 | 0.7506 | 0.8273 | 0.8472 |
| Proposed SSC–GSM | 28.45 | 30.65 | 34.33 | 31.51 | 29.73 | 29.38 | 24.77 | 27.37 | 33.65 | 29.97 |
| | 0.9272 | 0.9190 | 0.9236 | 0.8753 | 0.8638 | 0.7669 | 0.8062 | 0.7556 | 0.8236 | 0.8512 |
| | Noisy | | | | | | | | | |
| TV (Marquina and Osher 2008) | 25.49 | 27.01 | 29.70 | 28.13 | 26.57 | 26.74 | 23.11 | 25.35 | 29.86 | 26.88 |
| | 0.8477 | 0.8139 | 0.8047 | 0.7701 | 0.7557 | 0.6632 | 0.7131 | 0.6697 | 0.7291 | 0.7519 |
| Sparsity (Yang et al. 2010) | 23.61 | 27.15 | 29.57 | 28.31 | 26.60 | 27.22 | 22.45 | 25.40 | 30.71 | 26.78 |
| | 0.7532 | 0.7738 | 0.7700 | 0.7212 | 0.7052 | 0.6422 | 0.6477 | 0.6205 | 0.7051 | 0.7043 |
| NCSR (Dong et al. 2013b) | 26.86 | 29.51 | 31.73 | 29.94 | 28.08 | 28.03 | 23.80 | 26.38 | 32.03 | 28.48 |
| | 0.8878 | 0.8768 | 0.8594 | 0.8238 | 0.7934 | 0.6812 | 0.7369 | 0.6992 | 0.7637 | 0.7914 |
| Proposed SSC–GSM | 27.00 | 29.59 | 31.93 | 30.21 | 28.03 | 28.02 | 23.82 | 26.56 | 32.00 | 28.57 |
| | 0.8978 | 0.8853 | 0.8632 | 0.8354 | 0.7966 | 0.6747 | 0.7405 | 0.7066 | 0.7600 | 0.7956 |

5.1 Image Denoising

We have compared SSC–GSM based image denoising method against three current state-of-the-art methods including BM3D Image Denoising with Shape-Adaptive PCA (BM3D-SAPCA) (Katkovnik et al. 2010) (it is an enhanced version of BM3D denoising (Dabov et al. 2007) in which local spatial adaptation is achieved by shape-adaptive PCA), learned simultaneous sparse coding (LSSC) (Mairal et al. 2009b) and nonlocally centralized sparse representation (NCSR) denoising (Dong et al. 2013b). As can be seen from Table 2, the proposed SSC–GSM has achieved highly competitive denoising performance to other leading algorithms. For the collection of 12 test images, BM3D-SAPCA and SSC–GSM are mostly the best two performing methods - on the average, SSC–GSM falls behind BM3D-SAPCA by less than 0.2dB for three out of six noise levels but deliver at least comparable for the other three. We note that the complexity of BM3D-SAPCA is much higher than that of the original BM3D; by contrast, our pure Matlab implementation of SSC–GSM algorithm (without any C-coded optimization) still runs reasonably fast. It takes around 20s to denoise a 256×256 image on a PC with an Intel i7-2600 processor at 3.4GHz.

Figures. 2 and 3 include the visual comparison of denoising results for two typical images (*lena* and *house*) at moderate ($\sigma_w = 20$) and heavy ($\sigma_w = 100$) noise lev-

els respectively. It can be observed from Fig. 2 that BM3D-SAPCA and SSC–GSM seem to deliver the best visual quality at the moderate noise level; by contrast, restored images by LSSC and NCSR both suffer from noticeable artifacts especially around the smooth areas close to the hat. When the noise contamination is severe, the superiority of SSC–GSM to other competing approaches is easier to justify—as can be seen from Fig. 3, SSC–GSM achieves the most visually pleasant restoration of the house image especially when one inspects the zoomed portions of roof regions closely.

5.2 Image Deblurring

We have also compared SSC–GSM based image deblurring and three other competing approaches in the literature: constrained total variation image deblurring (denoted by FISTA), Iterative Decoupled Deblurring BM3D (IDD-BM3D) (Danielyan et al. 2012) and nonlocally centralized sparse representation (NCSR) denoising (Dong et al. 2013b). Note that the IDD-BM3D and NCSR are two recently developed state-of-the-art non-blind image deblurring approaches. In our comparative study, two commonly-used blur kernel i.e., 9×9 uniform and 2D Gaussian with standard deviation of 1.6; blurred images are further corrupted by additive white Gaussian noise with variance of $\sigma_n = \sqrt{2}$. Table 3 includes the PSNR/SSIM comparison results for a collection of 11

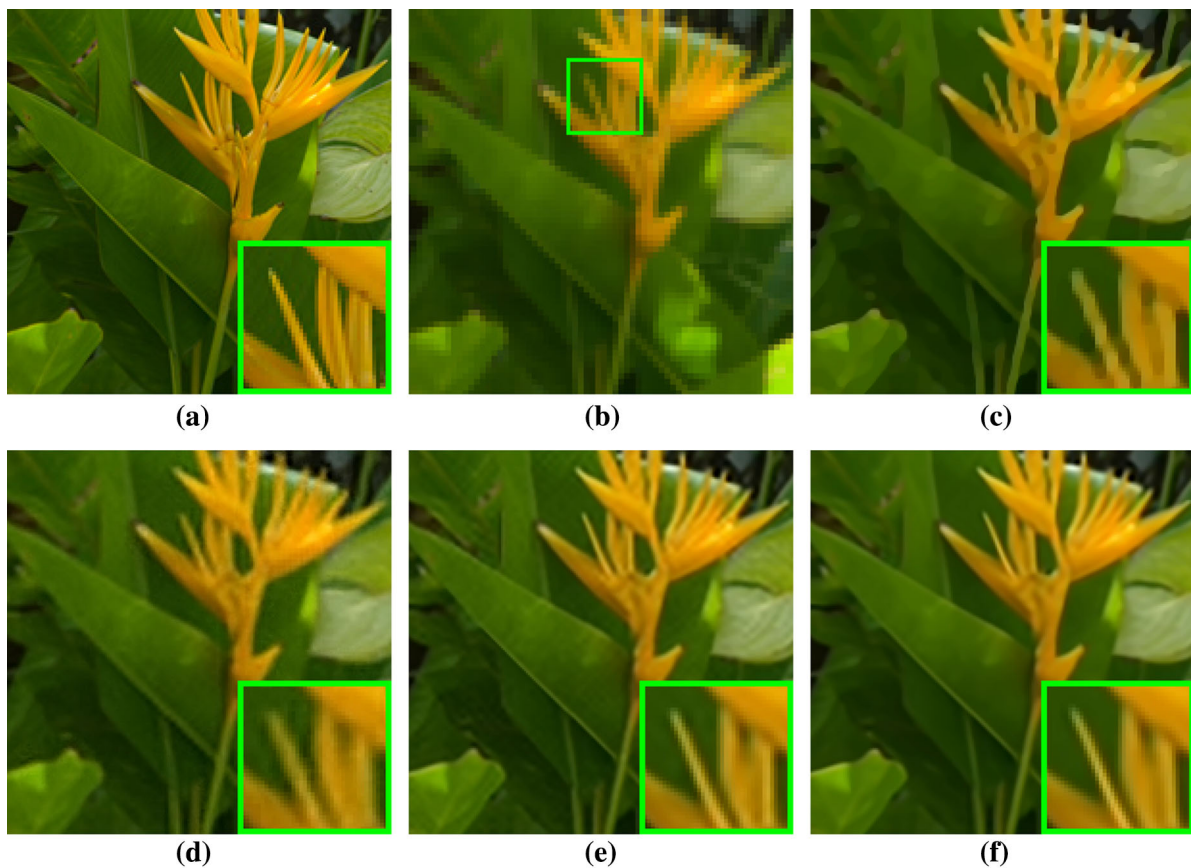


Fig. 7 Image super-resolution performance comparison on the *Plant* image (scaling factor 3, $\sigma_n = 0$). (a) Original image; (b) Low-resolution image; reconstructed images by (c) TV (Marquina and Osher 2008) (PSNR=31.34 dB, SSIM=0.8797); (d) Sparsity-based (Yang

et al. 2010) (PSNR=31.55 dB, SSIM=0.8964); (e) NCSR (Dong et al. 2013b) (PSNR=34.00 dB, SSIM=0.9369); (f) Proposed SSC-GSM (PSNR=34.33 dB, SSIM=0.9236)

images among four competing methods. It can be observed that SSC-GSM clearly outperforms all other three for 10 out of 11 images (the only exception is the *house* image for which IDD-BM3D slightly outperforms SSC-GSM by 0.13 dB). The gains are mostly impressive for *butterfly* and *barbara* images which contain abundant strong edges or textures. One possible explanation is that SSC-GSM is capable of striking a better tradeoff between exploiting local and non-local dependencies within those images.

Figures 4, 5 and 6 show the visual comparison of deblurring results for three test images: *starfish*, *butterfly* and *barbara* respectively. For *starfish*, it can be observed that IDD-BM3D and NCSR achieve deblurred images with similar quality (both noticeably better than FISTA); restored image by SSC-GSM is arguably the most preferred when compared against the original one (even though the PSNR gain is impressive). For *butterfly* and *barbara*, visual quality improvements achieved by SSC-GSM are readily observable—SSC-GSM is capable of both preserve the sharpness of edges and suppress undesirable artifacts. Such

experimental findings clearly suggest that the SSC-GSM model is a stronger prior for the class of photographic images containing strong edges/textures.

5.3 Image Superresolution

In our study on image super-resolution, simulated LR images are acquired from first applying a 7×7 uniform blur to the HR image, then down-sampling the blurred image by a factor of 3 along each dimension, and finally adding white Gaussian noise with $\sigma_n^2 = 25$ to the LR images. For color images, we work with the luminance channel only; simple bicubic interpolation method is applied to the upsampling of chrominance channels. Table 4 includes the PSNR/SSIM comparison for a set of 9 test images among four competing approaches. It can be seen that SSC-GSM outperforms others in most situations. Visual quality comparison as shown in Figs. 7 and 8 also justifies the superiority of SSC-GSM to other SR techniques.

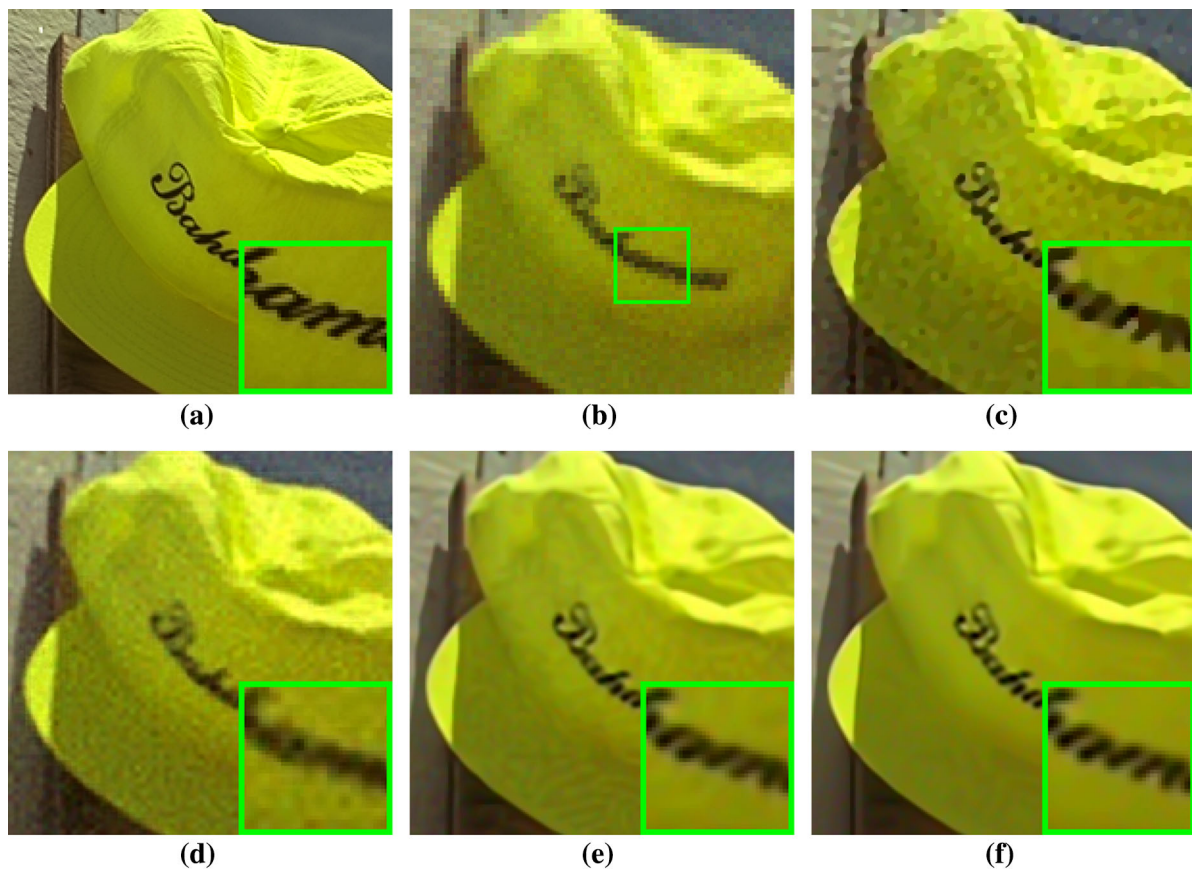


Fig. 8 Image super-resolution performance comparison on the *Hat* image (scaling factor 3, $\sigma_n = 5$). (a) Original image; (b) Low-resolution image; reconstructed images by (c) TV (Marquina and Osher 2008) (PSNR=28.13 dB, SSIM=0.7701); (d) Sparsity-based (Yang et al. 2010) (PSNR=28.31 dB, SSIM=0.7212); (e) NCSR (Dong et al. 2013b) (PSNR=29.94 dB, SSIM=0.8238); (f) Proposed SSC-GSM (PSNR=30.21 dB, SSIM=0.8354)

Table 5 Running time (sec) and the number of iterations (in parenthesis) of the test methods on a 256×256 test image on Intel Core i7-3770 CPU

| Denoising | | | |
|--------------------------------|-----------------------------|------------------------------------|----------------|
| LSSC (Mairal et al. 2009b) | NCSR (Dong et al. 2013b) | BM3D-SAPCA (Katkovnik et al. 2010) | SSC-GSM |
| – | 179.0 (27) | 127.7 (–) | 19.0 (4) |
| Deblurring | | | |
| FISTA (Beck and Teboulle 2009) | NCSR (Dong et al. 2013b) | BM3D-IDD (Danielyan et al. 2012) | SSC-GSM |
| 5.0 (120) | 139.6 (720) | 93.8 (200) | 501.0 (520) |
| Superresolution | | | |
| TV (Marquina and Osher 2008) | Sparsity (Yang et al. 2010) | NCSR (Dong et al. 2013b) | SSC-GSM |
| 25.4 (–) | 55.6 (–) | 264.1 (760) | 573.2 (400) |

5.4 Running Time

The proposed SSC–GSM algorithm was implemented under Matlab. The running time of the proposed method with comparison to other competing methods is reported in Table 5. The LSSC method is implemented in other platform and thus we don't report its running time. For image denoising, the proposed SSC–GSM method is about 6–9 times faster than the NCSR and BM3D-SAPCA methods. For image deblurring and superresolution, the proposed SSC–GSM method is much slower than other competing methods, as it requires more than four hundreds iterations. Since the patch grouping and the SSC for each exemplar patch can be implemented in parallel, the proposed SSC–GSM method can be much speeded up by using parallel computation techniques (e.g., GPU). Another way to accelerate the proposed method is to improve the convergence speed of **Algorithm 1**, which we remain it as future work.

6 Conclusions

In this paper, we proposed a new image model named SSC–GSM that connects SSC with GSM and explore its applications into image restoration. The proposed SSC–GSM model attempts to characterize both the biased-mean (like in NCSR) and spatially-varying variance (like in GSM) of sparse coefficients. It is shown that the formulated SSC–GSM problem, thanks to the power of alternating direction method of multipliers - can be decomposed into two subproblems both of which admit closed-form solutions when orthogonal basis is used. When applied to image restoration, SSC–GSM leads to computationally efficient algorithms involving iterative shrinkage/filtering only.

Extensive experimental results have shown that SSC–GSM can both preserve the sharpness of edges and suppress undesirable artifacts more effectively than other competing approaches. This work clearly shows the importance of spatial adaptation regardless the underlying image model is local or nonlocal; in fact, local variations and nonlocal invariance are two sides of the same coin - one has to take both of them into account during the art of image modeling.

In addition to image restoration, SSC–GSM can also be further studied along the line of dictionary learning. In our current implementation, we use PCA basis for its facilitating the derivation of analytical solutions. For non-unitary dictionary, we can solve the SSC–GSM problem by reducing it to iterative reweighted l_1 -minimization problem (Candes et al. 2008). It is also possible to incorporate dictionary D into the optimization problem formulated in Eq. (5); and from this perspective, we can view SSC–GSM as a generalization of K-SVD algorithm. Joint optimization of dictionary and

sparse coefficients is a more difficult problem and deserves more study. Finally, it is interesting to explore the relationship of SSC–GSM to the ideas in Bayesian nonparametrics (Polson and Scott 2010; Zhou et al. 2012) as well as the idea of integrating over hidden variables like BLS-GSM (Portilla et al. 2003).

Acknowledgments The authors would like to thank Zhouchen Lin of Peking University for helpful discussion. They would also like to thank the three anonymous reviewers for their valuable comments and constructive suggestions that have much improved the presentation of this paper. This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329402, in part by the Natural Science Foundation (NSF) of China under Grant 61471281, Grant 61227004 and Grant 61390512, in part by the Program for New Scientific and Technological Star of Shaanxi Province under Grant 2014KJXX-46, in part by the Fundamental Research Funds of the Central Universities of China under Grant BDY081424 and Grant K5051399020, and in part by NSF under Award ECCS-0968730.

References

- Aharon, M., Elad, M., & Bruckstein, A. (2012). The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, 54(11), 4311–4322.
- Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1), 99–102.
- Beck, A., & Teboulle, M. (2009). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11), 2419–2434.
- Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). New York: Wiley.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1–122.
- Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. *CVPR*, 2, 60–65.
- Candes, E., Wakin, M., & Boyd, S. (2008). Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5), 877–905.
- Carlson, C., Adelson, E., & Anderson, C. (Jun 1985). System for coring an image-representing signal. US Patent 4,523,230.
- Chang, S. G., Yu, B., & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9), 1532–1546.
- Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8), 2080–2095.
- Danielyan, A., Katkovnik, V., & Egiazarian, K. (2012). Bm3d frames and variational image deblurring. *IEEE Transactions on Image Processing*, 21(4), 1715–1728.
- Daubechies, I. (1988). Orthonormal bases of compactly supported bases. *Communications On Pure and Applied Mathematics*, 41, 909–996.
- Do, M. N., & Vetterli, M. (2005). The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12), 2091–2106.

- Dong, W., Li, X., Zhang, L., & Shi, G. (2011). Sparsity-based image denoising via dictionary learning and structural clustering. In: *IEEE Conference on Computer Vision and Pattern Recognition*.
- Dong, W., Shi, G., & Li, X. (2013a). Nonlocal image restoration with bilateral variance estimation: a low-rank approach. *IEEE Transactions on Image Processing*, 22(2), 700–711.
- Dong, W., Zhang, L., Shi, G., & Li, X. (2013b). Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4), 1620–1630.
- Donoho, D., & Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425–455.
- Elad, M., & Aharon, M. (2012). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 21(9), 3850–3864.
- Garrigues, P., & Olshausen, B. A. (2010). Group sparse coding with a laplacian scale mixture prior. In: *Advances in neural information processing systems*, (pp. 676–684).
- Gasso, G., Rakotomamonjy, A., & Canu, S. (2009). Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Transactions on Signal Processing*, 57(12), 4686–4698.
- Hocking, R. R. (1976). A biometrics invited paper: The analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1–49.
- Ji, S., Xue, Y., & Carin, L. (2008). Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6), 2346–2356.
- Katkovnik, V., Foi, A., Egiazarian, K., & Astola, J. (2010). From local kernel to nonlocal multiple-model image denoising. *International Journal of Computer Vision*, 86(1), 1–32.
- Lyu, S., & Simoncelli, E. (2009). Modeling multiscale subbands of photographic images with fields of gaussian scale mixtures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 693–706.
- Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2009a). Online dictionary learning for sparse coding. In: *2009 IEEE 26th International Conference on Machine Learning*, (pp. 689–696).
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., & Zisserman, A. (2009b). Non-local sparse models for image restoration. In: *2009 IEEE 12th International Conference on Computer Vision*, (pp. 2272–2279).
- Mairal, J., Sapiro, G., & Elad, M. (2008). Learning multiscale sparse representation for image and video restoration. *SIAM Multiscale Modeling and Simulation*, 7(1), 214–241.
- Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $l^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315, 69–87.
- Marquina, A., & Osher, S. J. (2008). Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37(3), 367–382.
- Mihcak, I. K. M.K., & Ramchandran, K. (1999). Local statistical modeling of wavelet image coefficients and its application to denoising. In: *IEEE International Conference on Acoust. Speech Signal Processing*, (pp. 3253–3256).
- Polson, N. G., & Scott, J. G. (2010). Shrink globally, act locally: sparse bayesian regularization and prediction. *Bayesian Statistics*, 9, 501–538.
- Portilla, J., Strela, V., Wainwright, M., & Simoncelli, E. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12, 1338–1351.
- Ramirez, I., & Sapiro, G. (2012). Universal regularizers for robust sparse coding and modeling. *IEEE Transactions on Image Processing*, 21(9), 3850–3864.
- Said, A., & Pearlman, W. A. (1996). A new fast and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6, 243–250.
- Shapiro, J. M. (1993). Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 41(12), 3445–3462.
- Taubman, D., & Marcellin, M. (2001). *JPEG2000: Image compression fundamentals, standards, and practice*. Norwell: Kluwer.
- Tipping, M. (2001). Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1, 211–244.
- Vetterli, M. (1986). Filter banks allowing perfect reconstruction. *Signal Processing*, 10(3), 219–244.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- Wipf, D. P., & Rao, B. D. (2004). Sparse bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8), 2153–2164.
- Wipf, D. P., & Rao, B. D. (2007). An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing*, 55(7), 3704–3716.
- Wipf, D. P., Rao, B. D., & Nagarajan, S. (2011). Latent variable bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9), 6236–6255.
- Xu, J., & Osher, S. (2007). Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. *IEEE Transactions on Image Processing*, 16(2), 534–544.
- Yang, J., Wright, J., Huang, T., & Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11), 2861–2873.
- Yu, G., Sapiro, G., & Mallat, S. (2012). Solving inverse problems with piecewise linear estimators: from gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5), 2481–2499.
- Zhou, M., Chen, H., Paisley, J., Ren, L., Sapiro, G., & Carin, L. (2009). Non-parametric bayesian dictionary learning for sparse image representation. In: *Advances in neural information processing systems*, (pp. 2295–2303).
- Zhou, M., Chen, H., Paisley, J., Ren, L., Li, L., Xing, Z., et al. (2012). Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Transactions on Image Processing*, 21(1), 130–144.
- Zoran, D., & Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In: *Proceedings of ICCV*.