Noname manuscript No.
(will be inserted by the editor)

Image Restoration via Simultaneous Sparse Coding and Gaussian Scale Mixture

Weisheng Dong · Guangming Shi · Yi Ma · Xin Li

the date of receipt and acceptance should be inserted later

Abstract Sparse coding has been shown 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 tune 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 an empirical Bayesian approach of addressing this issue - more specifically, a nonlocal extension of Gaussian scale mixture (GSM) model is developed based on 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. Unlike previous attacks to Bayesian image restoration, ours leads to closed-form solutions involving iterative shrinkage/filtering only and therefore admits computationally efficient implementation. Our solution can be viewed as an exemplar of demonstrating a new *variational* approach toward empirical Bayesian inference with parametric models, which has computational advantages over the existing variational Bayesian approaches. 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 · empirical Bayesian · alternative minimization · variational image restoration

W. Dong
School of Electronic Engineering, Xidian University, Xi'an, 710071, China
E-mail: wsdong@mail.xidian.edu.cn

G. Shi
School of Electronic Engineering, Xidian University, Xi'an, 710071, China
E-mail: gmshi@xidian.edu.cn

Yi Ma
Visual Computing Group, Microsoft Research of Asia, Beijing, China
E-mail: mayi@microsoft.com

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

1 Introduction

1.1 Background and Motivation

Sparse representation of signals/images has been widely studied by signal/image processing communities in the past decades. Historically, the idea of sparsity dated back to the idea of variable selection studied by statisticians in late 1970s [**?**] and coring operator invented by RCA researchers in early 1980s [1]. The birth of wavelet [2] or filter bank theory [3] or multi-resolution analysis [4] in late 1980s rapidly sparkled the interest in sparse representation, which has found successful applications into image coding [5], [6], [7] and denoising [8], [9], [10]. 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 [11]) and over-complete dictionary learning from training data (e.g., K-SVD [12, 13], multiscale dictionary learning [16], online dictionary learning [14] and

the non-parametric Bayesian dictionary learning [15]); the latter include the use of Gaussian mixture models [26,32], the variational Bayesian models [17,29], the universal models [24], and the centralized Laplacian model [34] for sparse coefficients.

More recently, a class of nonlocal image restoration techniques [19], [20], [21], [41], [22], [23,34] 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 nonlocal simultaneous sparse coding (SSC) [21]. Our own recent works [22], [23] 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 [24] 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 simultaneous sparse coding (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 [36] used in this work) over the positive scaling variables. We show that the empirical Bayesian estimates of both sparse coefficients and scaling variables can be efficiently calculated by alternating minimization. By characterizing the set of sparse coefficients of similar patches with the *same* prior distribution (i.e., the same nonzero means and positive scaling variables), we can effectively exploit the local and nonlocal dependencies among the sparse coefficients, which have been shown important for image restoration applications [20–22]. Our solution to SSC-GSM can be viewed as an exemplar of demonstrating a new *variational* approach toward empirical Bayesian inference with parametric models. By contrast to existing variational Bayesian approaches [17,29], the derived SSC-GSM based image restoration algorithms admits computationally efficient implementations. More importantly, SSC-GSM based image restoration is capable of both achieving local spatial adaptation and exploiting nonlocal dependencies within an image. Our experimental results have shown that SSC-GSM can restore images whose subjective/objective qualities are often higher than other state-of-the-art methods. Visual quality improvements are attributed to better preservation of edge sharpness and suppression of undesirable artifacts for some images.

1.2 Relationship to Other Competing Approaches

The connection between sparse coding and Bayesian inference has been previously studied in sparse Bayesian learning [27], [28] and more recently in Bayesian compressive sensing [17] and latent variable Bayesian models for promoting sparsity [29]. Despite offering a generic theoretical foundation as well as promising results, the Bayesian inference techniques along this line of works often involve potentially expensive sampling (e.g., approximated solutions for some choice of prior are achieved in [29]). By contrast, our empirical Bayesian 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) [30] and field-of-GSM [25]. In [30], the 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 [30] is very similar to the weights derived in the reweighted l_1-norm minimization [38]. 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 a more generic Bayesian inference of both scaling variables and sparse coefficients via efficient alternating optimization method. In [25], the field of Gaussian scale mixture model was constructed using the product of two independent homogeneous Gaussian Markov random fields (hGMRFs) to exploit the dependencies between adjacent blocks. Despite similar motivations to exploit the dependencies between the scaling variables, the techniques used in [25] is significantly different and requires a lot more computations than our SSC-GSM formulation.

The proposed work is also related to nonparametric Bayesian dictionary learning [31] and solving inverse problems with piecewise linear estimators [32] but ours is motivated by the connection between nonlocal SSC strategy and local parametric GSM model. When compared with previous works on image denoising (e.g., K-SVD denoising [13], spatially adaptive singular-value thresholding [23] and Expected Patch Log Likelihood (EPLL) [26]), SSC-GSM targets at a more general framework of combining parametric Bayesian 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) [33] and Nonlocal centralized sparse representation (NCSR) [34]) and the gain in terms of ISNR is as much as $1.5dB$ over IDD-BM3D [33] for some test images (e.g., $butterfly$ image - please refer to Fig. 4).

The reminder of this paper is organized as follows. In Sec. 2, we formulate the Bayesian sparse coding problem with Gaussian scale mixture model and generalize it into the new SSC-GSM model. In Sec. 3, we elaborate on the details of how to solve SSC-GSM by alternative minimization and emphasize the analytical solutions for both subproblems. In Sec. 4, we study the application of SSC-GSM into image restoration and discuss efficient implementation of SSC-GSM based image restoration algorithms. In Sec. 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 Sec. 6, we make some conclusions about the relationship of sparse coding to image restoration as well as perspectives about the future directions.

2 Bayesian Structured Sparse Coding: Connecting Gaussian Scale Mixture Models with Structured Sparsity

2.1 Bayesian Sparse Coding with Gaussian Scale Mixture Model

The basic idea behind sparse coding is to represent a signal $\boldsymbol{x} \in R^n$ (n is the size of an image patch) as the linear combination of basis vectors (dictionary elements) $\mathbf{D}\boldsymbol{\alpha}$ where $\mathbf{D} \in R^{n\times K}, n \leq K$ is the dictionary and the coefficients $\boldsymbol{\alpha} \in R^K$ 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:

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

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 $\boldsymbol{x}$ considering that real signal/image are non-stationary and may be degraded by noise and blur.

In this paper we propose to model the sparse coefficients $\boldsymbol{\alpha}$ with the Gaussian scale mixture (GSM) [35] model. The GSM model decomposes coefficient vector $\boldsymbol{\alpha}$ into the pointwise 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, the coefficients α_i is Gaussian with standard derivation 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

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

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 avoided by joint estimation of (α_i, θ_i). For a given observation $\boldsymbol{x} = \mathbf{D}\boldsymbol{\alpha}+\boldsymbol{n}$, where $\boldsymbol{n} \sim N(0, \sigma_n^2)$ denotes the additive Gaussian noise, we can formulate the following MAP estimator

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

where $P(\boldsymbol{x}|\boldsymbol{\alpha})$ 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

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

Instead of assuming the mean $\mu_i = 0$, inspired by our previous work NCSR [34] here we propose to use a biased-mean μ_i for α_i (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 [36] $P(\theta_i) \approx \frac{1}{\theta_i}$ - 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

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

where we have used $P(\boldsymbol{\theta}) = \sum_i P(\theta_i)$. Noting that Jeffrey's prior is unstable as $\theta_i \to 0$, we replace $\log\theta_i$ by $\log(\theta_i + \epsilon)$ where ϵ is a small positive number for numerical stability and rewrite $\sum_i \log(\theta_i + \epsilon)$ into $\log(\boldsymbol{\theta} + \epsilon)$ for notational simplicity. The above equation can then be further translated into the following Bayesian sparse coding (BSC) problem

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

Note that the matrix form of original GSM model is $\boldsymbol{\alpha} = \Lambda\boldsymbol{\beta}$ and $\boldsymbol{\mu} = \Lambda\boldsymbol{\gamma}$ where $\Lambda = diag(\theta_i) \in R^{K\times K}$ is a diagonal matrix characterizing the variance field for a chosen image patch. Accordingly, Bayesian sparse coding problem in Eq. (5) can be translated from $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ domain to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ domain as follows

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

In other words, BSC formulation of GSM model boils down to the joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. But unlike [18] 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. Therefore, our approach can be viewed as an empirical Bayesian approach toward joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ by connecting GSM with the SSC strategy as we will detail next.

2.2 Exploiting Simultaneous Sparse Coding 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 $\boldsymbol{\alpha}$'s should be characterized by the *same prior* - i.e., the density function with the same $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$. Therefore, if one consider the simultaneous Bayesian sparse coding of GSM models for the collection of m similar patches, a structured/group sparsity based extension of Eq. (7) can be written as

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

where $\mathbf{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_m]$ denotes the collection of m similar patches[1], $\mathbf{A} = \Lambda\mathbf{B}$ is the group representation of GSM model for sparse coefficients and their corresponding first-order and second-order statistics are characterized by $\Gamma = [\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_m] \in R^{K\times m}$ and $\mathbf{B} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m] \in R^{K\times m}$ respectively, wherein $\boldsymbol{\gamma}_j = \boldsymbol{\gamma}, j = 1, 2, \cdots, m$. From a collection of m similar patches, we have adopted the following strategy for estimating $\boldsymbol{\mu}$

$$\boldsymbol{\mu} = \sum_{j=1}^{m} w_j\boldsymbol{\alpha}_j, \tag{9}$$

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

$$\boldsymbol{\gamma} = \sum_{j=1}^{m} w_j\Lambda^{-1}\boldsymbol{\alpha}_j = \sum_{j=1}^{m} w_j\boldsymbol{\beta}_j, \tag{10}$$

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 $(\mathbf{B}, \boldsymbol{\theta})$ such that both dictionary learning and statistical modeling of sparse coefficients can be unified within the framework of Eq. (8).

3 Solving Bayesian Structured 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 $\mathbf{B}$ and $\boldsymbol{\theta}$. The key observation lies in that the two subproblems - minimization of $\mathbf{B}$ for a fixed $\boldsymbol{\theta}$ and minimization of $\boldsymbol{\theta}$ for a fixed $\mathbf{B}$ - both can be efficiently solved. Specifically, both subproblems admits closed-form solutions when the dictionary is orthogonal.

3.1 Solving $\boldsymbol{\theta}$ for a fixed $\mathbf{B}$

For a fixed $\mathbf{B}$, the first subproblem simply becomes

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{X} - \mathbf{D}\Lambda\mathbf{B}||_F^2 + 4\sigma_n^2\log(\boldsymbol{\theta}+\epsilon). \tag{11}$$

which can be rewritten as

$$\begin{aligned}\boldsymbol{\theta} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{X} - \sum_{i=1}^{K}\boldsymbol{d}_i\boldsymbol{\beta}^i\theta_i||_F^2 + 4\sigma_n^2\log(\boldsymbol{\theta}+\epsilon) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\tilde{\boldsymbol{x}} - \tilde{\mathbf{D}}\boldsymbol{\theta}||_2^2 + 4\sigma_n^2\log(\boldsymbol{\theta}+\epsilon),\end{aligned} \tag{12}$$

where the long vector $\tilde{\boldsymbol{x}} \in R^{nm}$ denotes the vectorization of the matrix $\mathbf{X}$, the matrix $\tilde{\mathbf{D}} = [\tilde{\boldsymbol{d}}_1, \tilde{\boldsymbol{d}}_2, \cdots, \tilde{\boldsymbol{d}}_K] \in R^{mn\times K}$ whose each column $\tilde{\boldsymbol{d}}_j$ denotes the vectorization of the rank-one matrix $\boldsymbol{d}_i\boldsymbol{\beta}^i$, and $\boldsymbol{\beta}^i \in R^m$ denotes the i-th row of matrix $\mathbf{B}$. Though $\log(\boldsymbol{\theta}+\epsilon)$ is non-convex, Eq. (12) can be solved by solving a sequence of reweighted l_1-minimization problems [38].

However, the optimization of Eq. (11) can be much simplified when the dictionary $\mathbf{D}$ is orthogonal (e.g., DCT or PCA basis). In the case of orthogonal dictionary, Eq. (11) can be rewritten as

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{A} - \Lambda\mathbf{B}||_F^2 + 4\sigma_n^2\log(\boldsymbol{\theta}+\epsilon). \tag{13}$$

where we have used $\mathbf{X} = \mathbf{D}\mathbf{A}$.

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

Although $\log(\boldsymbol{\theta}+\epsilon)$ is non-convex, we can efficiently solve it using a local minimization method for a local minimum. Let $f(\boldsymbol{\theta}) = \sum_{i=1}^{K}\log(\theta_i+\epsilon)$. We can approximate $f(\boldsymbol{\theta})$ by its first-order Taylor expansion, i.e.,

$$f(\boldsymbol{\theta}^{(k+1)}) = f(\boldsymbol{\theta}^{(k)}) + < \nabla f(\boldsymbol{\theta}^{(k)}), \boldsymbol{\theta}-\boldsymbol{\theta}^{(k)} > . \quad (14)$$

where $\boldsymbol{\theta}^{(k)}$ denotes the solution obtained in the k-th iteration. By using the fact that $\nabla f(\boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{K} 1/(\theta_i^{(k)}+\epsilon)$ and ignoring the constants in Eq. (14), Eq. (13) can be solved by iteratively minimizing

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{A}-\Lambda\mathbf{B}||_F^2 + 4\sigma_n^2||\mathbf{W}\boldsymbol{\theta}||_1. \quad (15)$$

where $\mathbf{W} = diag(\frac{1}{\theta_i^{(k)}+\epsilon})$ is the reweighting matrix that is often used in iterative reweighted l_1-minimization.

Since both Λ and $\mathbf{W}$ are diagonal, we can decompose the minimization problem in Eq. (15) into K parallel scalar optimization problems which admit highly efficient implementation. Let $\boldsymbol{\alpha}^i \in R^{1\times m}$ and $\boldsymbol{\beta}^i \in R^{1\times m}$ denote the i-th row of matrix $\mathbf{A} \in R^{n\times m}$ and $\mathbf{B} \in R^{n\times m}$, respectively. Eq. (15) can be rewritten as

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{K} ||(\boldsymbol{\alpha}^i)^T - (\boldsymbol{\beta}^i)^T\theta_i||_2^2 + 4\sigma_n^2 \sum_{i=1}^{K} \frac{\theta_i}{\theta_i^{(k)}+\epsilon}, \quad (16)$$

which can be conveniently decomposed into a sequence of independent scalar optimization problems

$$\theta_i^{(k+1)} = \underset{\theta_i}{\operatorname{argmin}} ||(\boldsymbol{\alpha}^i)^T - (\boldsymbol{\beta}^i)^T\theta_i||_2^2 + 4\sigma_n^2 \frac{\theta_i}{\theta_i^{(k)}+\epsilon}, \quad (17)$$

Now one can see this is standard l_2-l_1 optimization problem whose closed-form solution is given by

$$\theta_i^{(k+1)} = \frac{1}{\boldsymbol{\beta}^i(\boldsymbol{\beta}^i)^T}[\boldsymbol{\beta}^i(\boldsymbol{\alpha}^i)^T - \tau]_+, \quad (18)$$

where the threshold $\tau = \frac{4\sigma_n^2}{\theta_i^{(k)}+\epsilon}$ and $[\cdot]_+$ denotes the soft shrinkage operator.

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

The second subproblem is in fact easier to solve than the first one. It takes the following form

$$\mathbf{B} = \operatorname{argmin} ||\mathbf{X}-\mathbf{D}\Lambda\mathbf{B}||_F^2 + \sigma_n^2||\mathbf{B}-\Gamma||_F^2. \quad (19)$$

Since both terms are l_2, the closed-form solution to Eq. (19) is essentially a Wiener filtering

$$\mathbf{B} = (\hat{\mathbf{D}}^T\hat{\mathbf{D}} + \sigma_n^2\mathbf{I})^{-1}(\hat{\mathbf{D}}^T\mathbf{X} + \Gamma). \quad (20)$$

where $\hat{\mathbf{D}} = \mathbf{D}\Lambda$. Note that when $\mathbf{D}$ is orthogonal, Eq. (20) can be further simplified into

$$\mathbf{B} = (\Lambda^T\Lambda + \sigma_n^2\mathbf{I})^{-1}(\Lambda^T\mathbf{A} + \Gamma). \quad (21)$$

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

By alternatively solving both subproblems of Eqs. (11) and (19) for the estimates of Λ and $\mathbf{B}$, the image data matrix $\mathbf{X}$ can then be reconstructed as

$$\hat{\mathbf{X}} = \mathbf{D}\hat{\Lambda}\hat{\mathbf{B}}, \quad (22)$$

where $\hat{\Lambda}$ and $\hat{\mathbf{B}}$ denotes the final estimates of Λ 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: $\boldsymbol{y} = \mathbf{H}\boldsymbol{x}+\boldsymbol{w}$ where $\boldsymbol{x} \in R^N, \boldsymbol{y} \in R^M$ denotes the original and degraded images respectively, $\mathbf{H} \in R^{N\times M}$ is the degradation matrix and $\boldsymbol{w}$ is additive white Gaussian noise observing $N(0,\sigma_n^2)$. The whole-image reconstruction problem can be expressed as

$$\begin{aligned}(\boldsymbol{x}, \{\mathbf{B}_l\}, \{\boldsymbol{\theta}_l\}) = \underset{\boldsymbol{x},\{\mathbf{B}_l\},\{\boldsymbol{\theta}_l\}}{\operatorname{argmin}} & ||\boldsymbol{y}-\mathbf{H}\boldsymbol{x}||_2^2 \\ & + \sum_{l=1}^{L}\{\eta||\tilde{\mathbf{R}}\boldsymbol{x} - \mathbf{D}\Lambda_l\mathbf{B}_l||_F^2 \\ & + \sigma_n^2||\mathbf{B}-\Gamma||_F^2 + 4\sigma_n^2\log(\boldsymbol{\theta}_l+\epsilon)\}.\end{aligned} \quad (23)$$

where $\tilde{\mathbf{R}}_l\boldsymbol{x} \doteq [\mathbf{R}_{l_1}\boldsymbol{x}, \mathbf{R}_{l_2}\boldsymbol{x}, \cdots, \mathbf{R}_{l_m}\boldsymbol{x}] \in R^{n\times m}$ denotes the data matrix formed by the group of image patches similar to the l-th exemplar patch $\boldsymbol{x}_l$ (including $\boldsymbol{x}_l$ itself), $\mathbf{R}_l \in R^{n\times N}$ denotes the matrix extracting the $l-th$ patch $\boldsymbol{x}_l$ from $\boldsymbol{x}$, and L is the total number of exemplars extracted from the reconstructed image $\boldsymbol{x}$. Invoking the principle of alternative optimization again, we propose to solve the whole-image reconstruction problem in Eq. (23) by alternating the solutions to the following two subproblems:

4.1 Solving $\boldsymbol{x}$ for a fixed $\{\mathbf{B}_l\}, \{\boldsymbol{\theta}_l\}$

Let $\hat{\mathbf{X}}_l = \mathbf{D}\Lambda_l\mathbf{B}_l$. When $\{\mathbf{B}_l\}$ and $\{\boldsymbol{\theta}_l\}$ are fixed, so is $\{\hat{\mathbf{X}}_l\}$. Therefore, Eq. (23) reduces to the following l_2-optimization

problem

$$\boldsymbol{x} = \underset{\boldsymbol{x}}{\operatorname{argmin}} ||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||_2^2 + \sum_{l=1}^{L} \eta ||\mathbf{R}\boldsymbol{x}_l - \hat{\mathbf{X}}_l||_F^2. \quad (24)$$

which admits the following closed-form solution

$$\boldsymbol{x} = (\mathbf{H}^T\mathbf{H} + \eta \sum_{l=1}^{L} \tilde{\mathbf{R}}_l^\top \tilde{\mathbf{R}}_l)^{-1} (\mathbf{H}^T\boldsymbol{y} + \eta \sum_{l=1}^{L} \tilde{\mathbf{R}}_l^\top \hat{\mathbf{X}}_l). \quad (25)$$

where $\tilde{\mathbf{R}}_l^\top \tilde{\mathbf{R}}_l \doteq \sum_{j=1}^{m} \mathbf{R}_j^\top \mathbf{R}_j$, $\tilde{\mathbf{R}}_l^\top \hat{\mathbf{X}}_l \doteq \sum_{j=1}^{m} \mathbf{R}_j^\top \hat{\boldsymbol{x}}_{l_j}$ and $\hat{\boldsymbol{x}}_{l_j}$ 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 inversed in Eq. (25) is diagonal, and its inverse can be solved easily. Actually, similar to the K-SVD approach Eq. (25) can be computed by weighted averaging each reconstructed patches sets $\tilde{\mathbf{X}}_l$. For image deblurring and super-resolution applications, Eq. (25) can be computed by using a conjugate gradient (CG) algorithm.

4.2 Solving $\{\mathbf{B}_l\}, \{\boldsymbol{\theta}_l\}$ for a fixed $\boldsymbol{x}$

When $\boldsymbol{x}$ is fixed, the first term in Eq. (23) go away and the subproblem boils down to a sequence of patch-level SSC-GSM problems formed for each exemplar - i.e.,

$$(\mathbf{B}_l, \boldsymbol{\theta}_l) = \underset{\mathbf{B}_l, \boldsymbol{\theta}_l}{\operatorname{argmin}} ||\mathbf{X}_l - \mathbf{D}\Lambda_l\mathbf{B}_l||_F^2 + \frac{\sigma_n^2}{\eta}||\mathbf{B} - \Gamma||_F^2 + \frac{4\sigma_n^2}{\eta}\log(\boldsymbol{\theta}_l + \epsilon). \quad (26)$$

where we use $\mathbf{X}_l = \tilde{\mathbf{R}}_l\boldsymbol{x}$. 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 [21], here similar to NCSR [34] we learn the principle component analysis (PCA) based dictionary for each dataset. 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.

Algorithm 1. SSC-GSM-based Image Restoration

• Initialization:

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

(b) Set parameters η;

(c) Obtain data matrices $\{\mathbf{X}_l\}$'s from $\hat{\boldsymbol{x}}$ (though kNN search) for each exemplar and compute the PCA basis $\{\mathbf{D}_l\}$ for each $\mathbf{X}_l$.

• Outer loop (solve Eq. (23) by alternative optimization): Iterate on $k = 1, 2, \cdots, k_{max}$

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

(b) Estimate biased means $\boldsymbol{\mu}$ using Eq. (9) for each $\mathbf{X}_l$;

(c) Inner loop (solve Eq. (26) for each data $\mathbf{X}_l$): iterate on $J = 1, 2, \cdots, J$;

(I) update $\boldsymbol{\theta}_l$ for fixed $\mathbf{B}_l$ using Eq. (18);

(II) update $\mathbf{B}_l$ for fixed $\boldsymbol{\theta}_l$ using Eq. (21);

(d) Reconstruct $\mathbf{X}_l$'s from $\boldsymbol{\theta}_l$ and $\mathbf{B}_l$ using Eq. (22);

(e) If $mod(k, k_0) = 0$, update the PCA basis $\{\mathbf{D}_l\}$ for each $\mathbf{X}_l$;

(f) Patch-to-image transformation: obtain reconstructed $\hat{\boldsymbol{x}}^{(k+1)}$ from $\{\mathbf{X}_l\}$'s by solving Eq. (25);

• Output: $\hat{\boldsymbol{x}}^{(k+1)}$.

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 the identity matrix, **Algorithm 1** is an image denoising algorithm using iterative regularization technique [39]. When $\mathbf{H}$ is a blur matrix or reduced blur matrix, Eq. (23) 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.

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 [34]. 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. To evaluate the quality of restored images, both PSNR and SSIM [40] 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://www.csee.wvu.edu/~xinl/source.html`.

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) [41] (it is an enhanced version of BM3D denoising [20] in which local spatial adaptation is achieved by shape-adaptive PCA), learned simultaneous sparse coding (LSSC) [21] and nonlocally centralized sparse representation (NCSR) denoising [34]. As can be seen from Table I, 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 20 seconds to denoise a 256×256 image on a PC with an Intel i7-2600 processor at 3.4GHz.

Figs. 1 and 2 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 levels respectively. It can be observed from Fig. 1 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. 2, 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) [33] and nonlocally centralized sparse representation (NCSR) denoising [34]. 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 kernal - 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 II includes the PSNR/SSIM comparison results for a collection of 11 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 $0.13dB$). 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 nonlocal dependencies within those images.

Figs. 3-5 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 III includes the PSNR/SSIM comparison for a set of 10 test images among four competing approaches. It can be seen that SSC-GSM outperforms others in most situations (8 out of 10). Visual quality comparison as shown in Figs. 7 and 6 also justifies the superiority of SSC-GSM to other SR techniques.

6 Conclusions

In this paper, we proposed a new image model named SSC-GSM that connects simultaneous sparse coding with Gaussian scale mixture 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

Table 1 The PSNR (dB) results by different denoising methods. In each cell, the results of the four denoising methods are reported in the following order: top left-BM3D-SAPCA [41]; top right-LSSC [21]; bottom left-NCSR [34]; bottom right-proposed SSC-GSM

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**

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. Our solution to SSC-GSM can be viewed as an exemplar of demonstrating a new variational approach toward empirical Bayesian inference with parametric models. 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 [38]. It is also possible to incorporate dictionary $\mathbf{D}$ into the optimization problem formulated in Eq. (5); and from this perspective, we can view BSSD as a Bayesian 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 [45],[31] as well as the idea of integrating over hidden variacles like BLS-GSM [18] (i.e., from empirical Bayesian to fully Bayesian).

References

1. C. Carlson, E. Adelson, and C. Anderson, "System for coring an image-representing signal," Jun 1985, US Patent 4,523,230.
2. I. Daubechies, "Orthonormal bases of compactly supported bases," *Communications On Pure and Applied Mathematics*, vol. 41, pp. 909–996, 1988.
3. M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Processing*, pp. 219–244, 1986.
4. S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $l^2(\mathbf{r})$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69–87, 1989.
5. J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Acous. Speech Sig. Proc.*, vol. 41, no. 12, pp. 3445–3462, 1993.
6. A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, 1996.

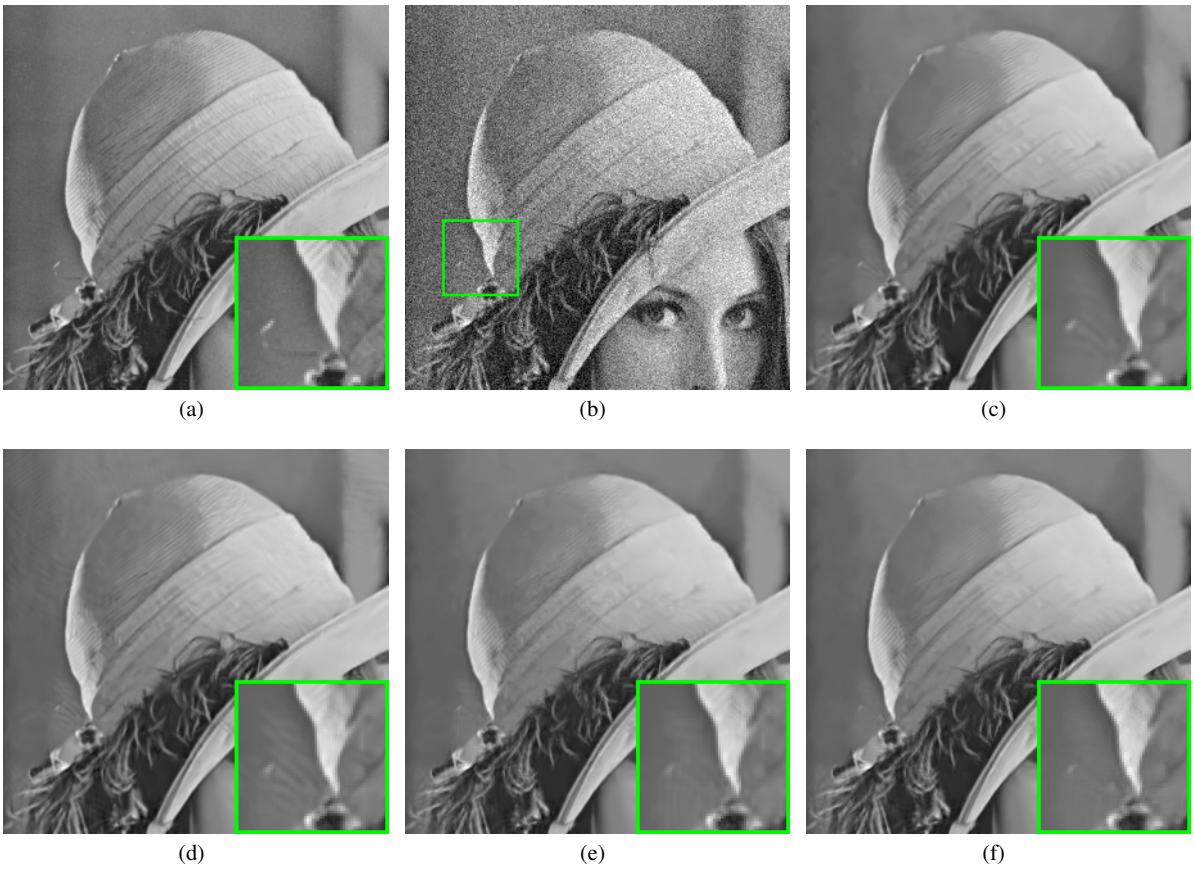


Fig. 1 Denoising performance comparison on the *Lena* image with moderate noise corruption. (a) Original image; (b) Noisy image ($\sigma_n = 20$); denoised images by (c) BM3D-SAPCA [41] (PSNR=**33.20** dB, SSIM=**0.8803**); (d) LSSC [21] (PSNR=32.88 dB, SSIM=0.8742); (e) NCSR [34] (PSNR=32.92 dB, SSIM=0.8760); (f) Proposed SSC-GSM (PSNR=33.08, SSIM=0.8787).

7. D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*. Kluwer, 2001.
8. D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
9. I. K. M.K. Mihcak and K. Ramchandran, "Local statistical modeling of wavelet image coefficients and its application to denoising," in *IEEE International Conference on Acoust. Speech Signal Processing*, 1999, pp. 3253–3256.
10. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.
11. M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
12. M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations," *IEEE Trans. on Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2012.
13. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Process.*, vol. 21, no. 9, pp. 3850–3864, Sep. 2012.
14. J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *2009 IEEE 26th International Conference on Machine Learning*, pp. 689–696, 2009.
15. M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, "Non-parametric bayesian dictionary learning for sparse image representation," in *Advances in neural information processing systems*, pp. 2295–2303, 2009.
16. J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representation for image and video restoration,", *SIAM Multiscale Model. Simul.*, vol. 7, no. 1, pp. 214–241, Apr. 2008.
17. S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. on Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
18. J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, pp. 1338–1351, Nov 2003.
19. A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *CVPR*, vol. 2, pp. 60–65, 2005.
20. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
21. J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 2272–2279.
22. W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
23. W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: a low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

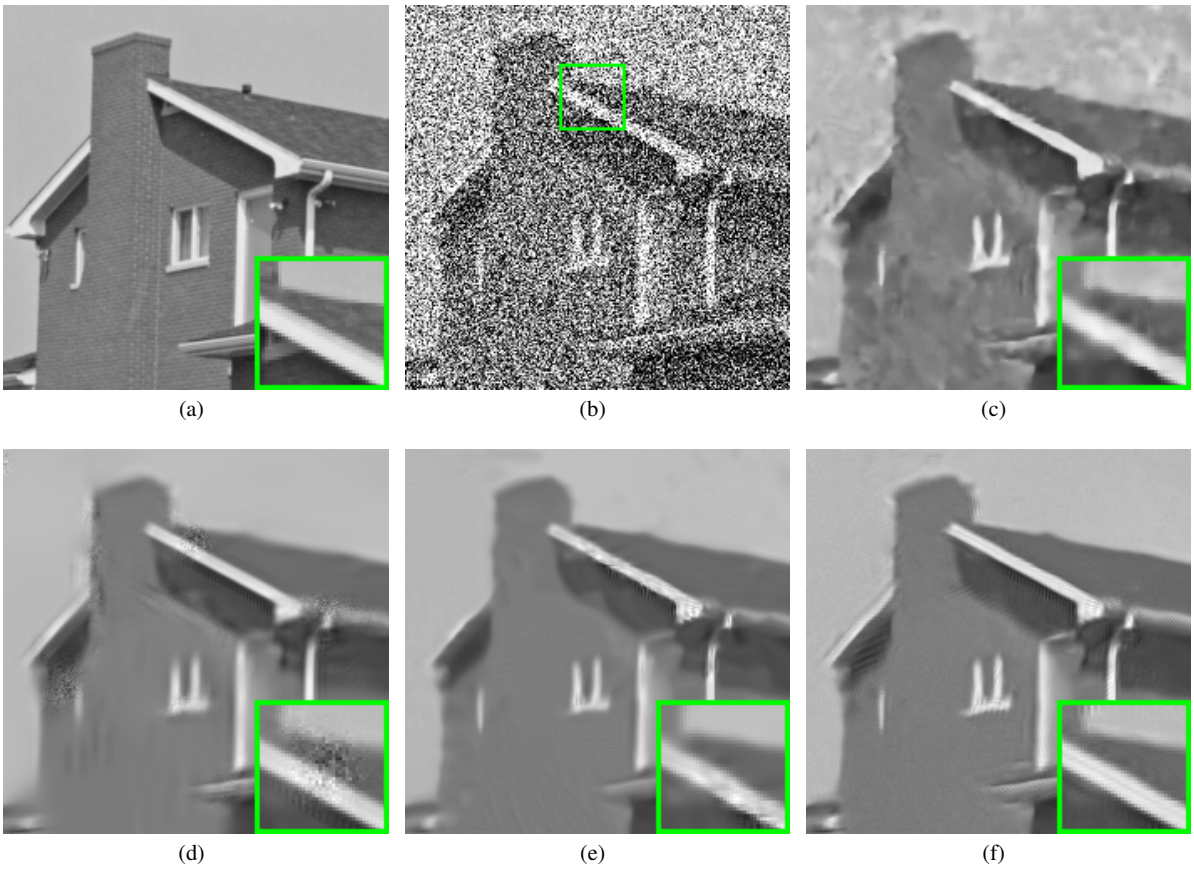


Fig. 2 Denoising performance comparison on the *House* image with strong noise corruption. (a) Original image; (b) Noisy image ($\sigma_n = 100$); denoised images by (c) BM3D-SAPCA [41] (PSNR=35.20 dB, SSIM=0.6767); (d) LSSC [21] (PSNR=25.63 dB, SSIM=0.7389); (e) NCSR [34] (PSNR=25.65 dB, SSIM=**0.7434**); (f) Proposed SSC-GSM (PSNR=**26.70**, SSIM=0.7430).

24. I. Ramirez and G. Sapiro, "Universal regularizers for robust sparse coding and modeling," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3850–3864, 2012.
25. S. Lyu and E. Simoncelli, "Modeling multiscale subbands of photographic images with fields of gaussian scale mixtures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 693–706, 2009.
26. D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. of ICCV*, 2011.
27. M. Tipping, "Sparse bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
28. D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
29. D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6236–6255, 2011.
30. P. Garrigues and B. A. Olshausen, "Group sparse coding with a laplacian scale mixture prior," in *Advances in neural information processing systems*, 2010, pp. 676–684.
31. M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 130–144, 2012.
32. G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.
33. A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1715–1728, 2012.
34. W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.
35. D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 1, pp. 99–102, 1974.
36. G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley and Sons, 2011, vol. 40.
37. X. Li and M. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *IEEE International Conference On Image Processing*, 2000, pp. 300–303.
38. E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
39. J. Xu and S. Osher, "Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 534–544, 2007.
40. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity." *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
41. V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, Aug. 2010.

Table 2 PSNR(dB) and SSIM results of the deblurred images.

	9×9 uniform blur, $\sigma_n = \sqrt{2}$										
Images	*Butterfly*	*Boats*	*C. Man*	*Starfish*	*Parrot*	*Lena*	*Barbara*	*Peppers*	*Leaves*	*House*	*Average*
FISTA [42]	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 [33]	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 [34]	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 [42]	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 [33]	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 [34]	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**

Table 3 PSNR(dB) and SSIM results(luminance components) of the reconstructed HR images.

	Noiseless									
Images	*Butterfly*	*Parrot*	*Plants*	*Hat*	*flower*	*Raccoon*	*Bike*	*Pathenon*	*Girl*	*Average*
TV [43]	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 [44]	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 [34]	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 [43]	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 [44]	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 [34]	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**

42. A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2419–2434, 2009.
43. A. Marquina and S. J. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.
44. J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
45. N. G. Polson and J. G. Scott, "Shrink globally, act locally: sparse bayesian regularization and prediction," *Bayesian Statistics*, vol. 9, pp. 501–538, 2010.

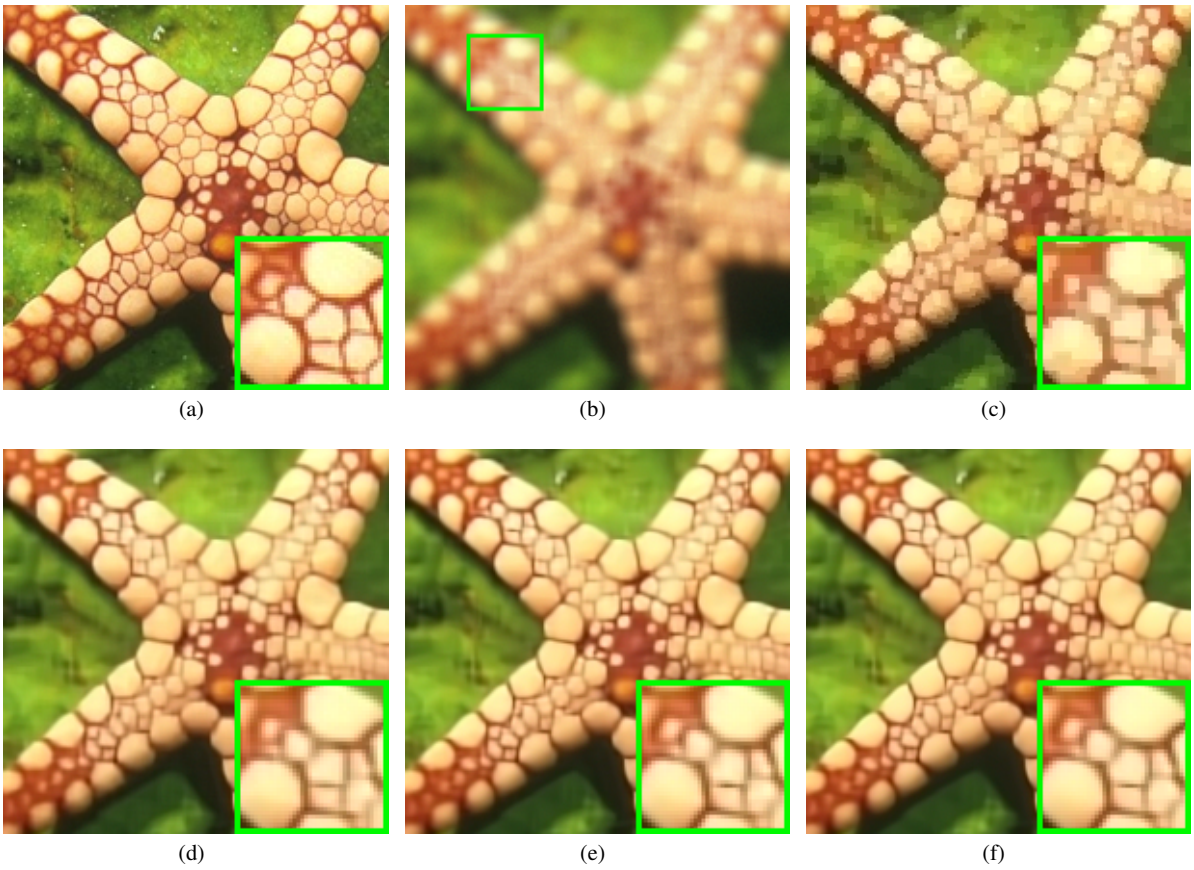
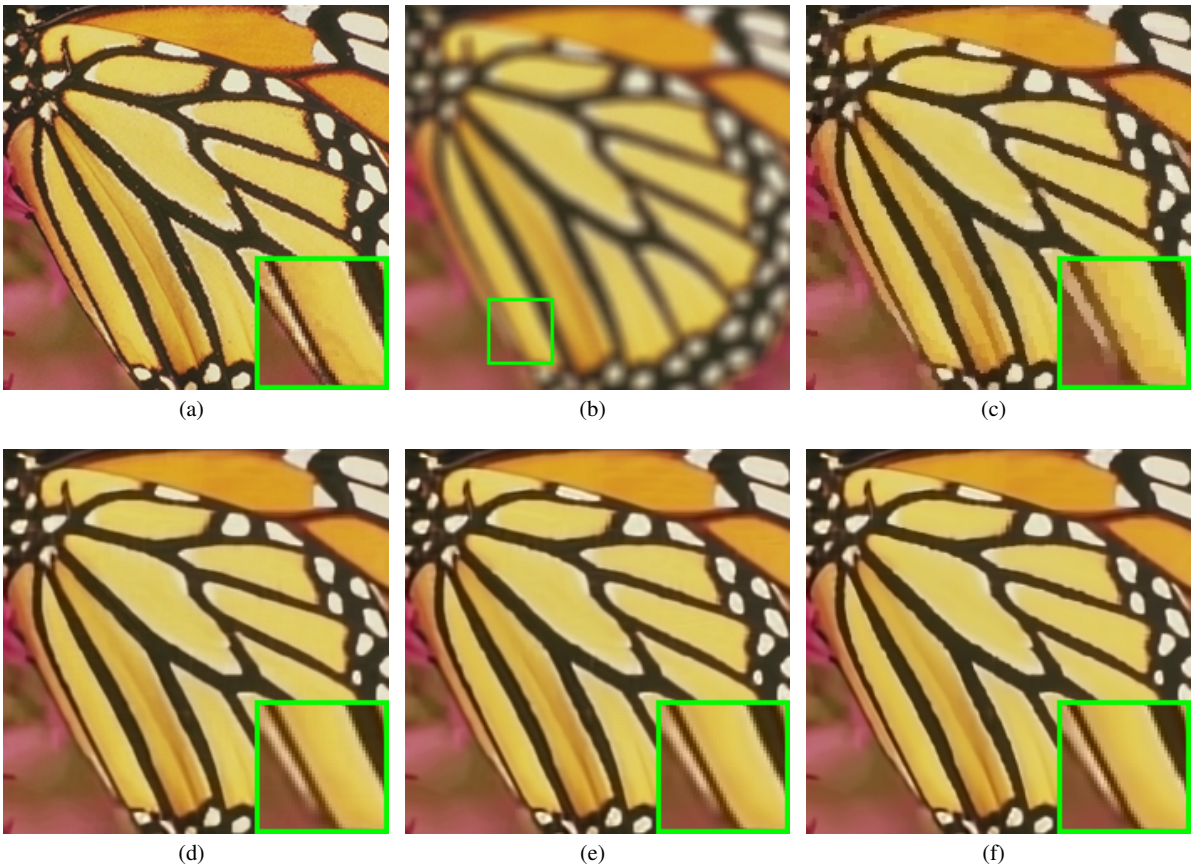


Fig. 3 Deblurring performance comparison on the *Starfish* image. (a) Original image; (b) Noisy and blurred image (9×9 uniform blur, $\sigma_n = \sqrt{2}$); deblurred images by (c) FISTA [42] (PSNR=27.75 dB, SSIM=0.8200); (d) IDD-BM3D [33] (PSNR=29.48 dB, SSIM=0.8640); (e) NCSR [34] (PSNR=30.28 dB, SSIM=0.8807); (f) Proposed SSC-GSM (PSNR=**30.58** dB, SSIM=**0.8862**).



(a) (b) (c)

(d) (e) (f)

Fig. 4 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 [42] (PSNR=28.37 dB, SSIM=0.9058); (d) IDD-BM3D [33] (PSNR=29.21 dB, SSIM=0.9216); (e) NCSR [34] (PSNR=29.68 dB, SSIM=0.9273); (f) Proposed SSC-GSM (PSNR=**30.45** dB, SSIM=**0.9377**).

Fig. 5 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 [42] (PSNR=25.03 dB, SSIM=0.7377); (d) IDD-BM3D [33] (PSNR=27.19 dB, SSIM=0.8231); (e) NCSR [34] (PSNR=27.91 dB, SSIM=0.8304); (f) Proposed SSC-GSM (PSNR=**28.42** dB, SSIM=**0.8462**).

Fig. 6 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 [43] (PSNR=31.34 dB, SSIM=0.8797); (d) Sparsity-based [44] (PSNR=31.55 dB, SSIM=0.8964); (e) NCSR [34] (PSNR=34.00 dB, SSIM=**0.9369**); (f) Proposed SSC-GSM (PSNR=**34.33** dB, SSIM=0.9236).

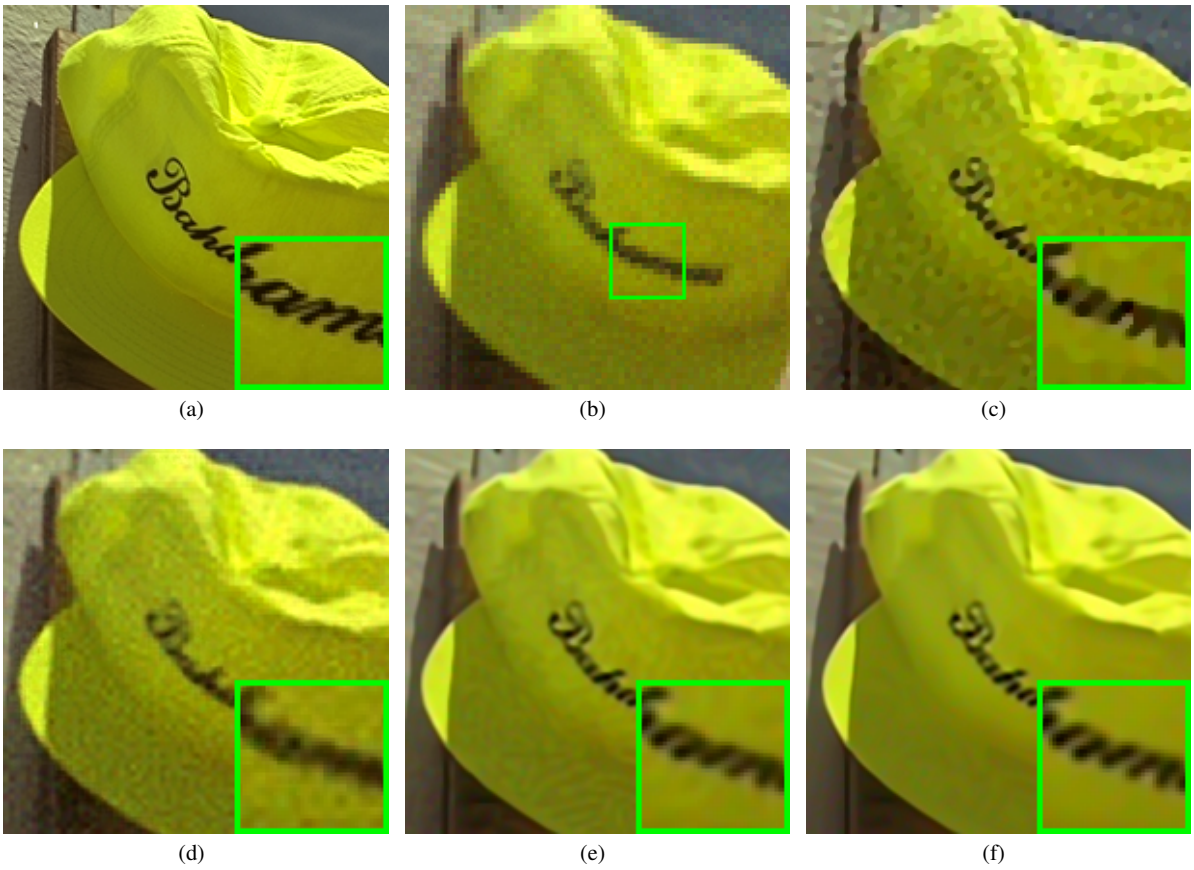


Fig. 7 Image super-resolution performance comparison on the *Hat* image (scaling factor 3, $\sigma_n = 5$). (a) Original image; (b) Low-resolution image; reconstructed images by (c) TV [43] (PSNR=28.13 dB, SSIM=0.7701); (d) Sparsity-based [44] (PSNR=28.31 dB, SSIM=0.7212); (e) NCSR [34] (PSNR=29.94 dB, SSIM=0.8238); (f) Proposed SSC-GSM (PSNR=**30.21** dB, SSIM=**0.8354**).