# IMAGE RESTORATION VIA BAYESIAN STRUCTURED SPARSE CODING

*Weisheng Dong[a], Xin Li[b], Yi Ma[c], Guangming Shi[a]*

[a]School of Electronic Engineering, Xidian University, Xi'an, China
[b]Lane Dept. of Computer Science and Electrical Engineering
West Virginia University, Morgantown WV26506-6109
[c] Visual Computing Group, Microsoft Research Asia

## ABSTRACT

In this work, we propose a Bayesian structured sparse coding (BSSC) framework containing a nonlocal extension of Gaussian scale mixture (GSM) model by exploiting structured sparsity. It is shown that the variances of sparse coefficients (the field of Gaussian scalars) - if treated as a latent variable - can be jointly estimated along with the unknown sparse coefficients via the the method of alternative optimization. When applied to image restoration, BSSC leads to closed-form solutions involving iterative shrinkage/filtering and therefore admits computationally efficient implementation. Our experimental results have shown that BSSC-based image restoration often delivers reconstructed images with higher subjective/objective qualities than other competing approaches including IDD-BM3D and NCSR.

***Index Terms***— Bayesian sparse coding, Gaussian scale mixture, structured sparsity, alternative minimization, variational image restoration.

## I. INTRODUCTION

Recently, a class of nonlocal image restoration techniques [2], [3], [4], [5], [6] 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 are characterized by the abundance of self-repeating patterns. Such observation has led to the formulation of so-called structured/group sparsity [7] (a.k.a. simultaneous sparse coding [4]). Our own recent works [5], [6] has more clearly verified the potential of exploiting structured sparsity 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 parameter in a variational setting or shape parameter 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.

In this work, we propose a new image representation named Bayesian structured sparse coding (BSSC) that connects existing Gaussian scale mixture (GSM) model [1] with structured sparsity. Our idea is to treat the variance field - i.e., the collection of Gaussian scalars - as a latent variable and resort to nonlocal similar patches for a robust estimation. Similar to Field-of-GSM [8] and Expected Patch Log Likelihood (EPLL) [9], our formulation admits a clean Bayesian treatment of hidden parameters; but unlike those previous works, we are able to obtain closed-form solutions to the proposed BSSC problem under a mild assumption with the dictionary being unitary. Our solution to BSSC 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, the derived BSCC-based image restoration algorithms admits computationally efficient implementations. More importantly, BSSC-based image restoration is capable of both achieving local spatial adaptation and exploiting nonlocal dependencies within an image. Our experimental results have shown that BSSC can deliver restored images whose subjective/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 for some images.

## II. BAYESIAN STRUCTURED SPARSE CODING: CONNECTING GAUSSIAN SCALE MIXTURE WITH STRUCTURED SPARSITY

GSM decomposes coefficient vector $\boldsymbol{\alpha}$ into the product of a Gaussian vector $\boldsymbol{\beta}$ and a hidden scalar multiplier $\boldsymbol{\theta}$ -i.e., $\alpha_i = \theta_i \beta_i$. As articulated in [1], the sparsity related to coefficients $\boldsymbol{\alpha}$ can now be characterized by the prior density $P(\boldsymbol{\theta})$ specified for the multiplier $\boldsymbol{\theta}$. More specifically, one can formulate the following maximum a posterior (MAP) estimation problem for the target signal $\boldsymbol{x} = \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{n}$ ($\boldsymbol{n} \sim N(0, \sigma_n^2)$ is a term characterizing the approximation errors of sparse coding)

$$\begin{aligned}(\boldsymbol{\alpha}, \boldsymbol{\theta}) &= \operatorname{argmax} logP(\boldsymbol{x}|\boldsymbol{\alpha}, \boldsymbol{\theta})P(\boldsymbol{\alpha}, \boldsymbol{\theta}) \\ &= \operatorname{argmax} logP(\boldsymbol{x}|\boldsymbol{\alpha}) + logP(\boldsymbol{\alpha}|\boldsymbol{\theta}) + logP(\boldsymbol{\theta}).\end{aligned} \tag{1}$$

where $P(\boldsymbol{x}|\boldsymbol{\alpha})$ is the likelihood term observing standard Gaussian with variance $\sigma_n^2$. Inspired by our previous work NCSR [14], we propose to introduce a biased-mean term to the GSM model - that is, $P(\boldsymbol{\alpha}|\boldsymbol{\theta})$ can be written as

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

where $\mu_i$ is the biased mean for $\alpha_i$ (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 prior density $P(\boldsymbol{\theta})$. It has been suggested in the literature that noninformative prior $P(\theta_i) \approx \frac{1}{\theta_i}$ - a.k.a. Jeffrey's prior - is often the favorable choice. Therefore, we have also adopted this option in this work, which translates Eq. (1) into

$$(\boldsymbol{\alpha},\boldsymbol{\theta}) = \underset{\boldsymbol{\alpha},\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{2\sigma_n^2}||\boldsymbol{x}-\mathbf{D}\boldsymbol{\alpha}||_2^2 + \sum_i log(\theta_i\sqrt{2\pi}) + \sum_i \frac{(\alpha_i-\mu_i)^2}{2\theta_i^2} + \sum_i log\theta_i. \tag{3}$$

where we have used $P(\boldsymbol{\theta}) = \sum_i P(\theta_i)$ under the assumption with Jeffrey's prior and $\mathbf{D} \in R^{n\times K}$ denotes the dictionary ($K$ is the size of the dictionary). The above equation can be further simplified into the following Bayesian sparse coding (BSC) problem

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

Such BSC 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.

Unlike [1] 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 variance. In the matrix form, we have $\boldsymbol{\alpha} = \Lambda\boldsymbol{\beta}$ 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. From a collection of $m$ similar patches, we have adopted the following strategy for estimating $\boldsymbol{\mu}$

$$\boldsymbol{\mu} = \sum_{j=1}^{m} w_j \boldsymbol{\alpha}_j, \tag{5}$$

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

$$\boldsymbol{\gamma} = \sum_{j=1}^{m} w_j \Lambda^{-1}\boldsymbol{\alpha}_j = \sum_{j=1}^{m} w_j \boldsymbol{\beta}, \tag{6}$$

Accordingly, Bayesian sparse coding problem in Eq. (3) can be rewritten as

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

which can be solved by alternatively minimizing the objective functional with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

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

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

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. When compared against previous formulation on structured sparsity (or simultaneous sparse coding [4]), one can see both dictionary learning and statistical modeling of sparse coefficients are unified within the framework of Eq. (8). We call such new formulation *Bayesian structured sparse coding* (BSSC) and propose to solve it via alternating minimization as follows.

**A. Solving $\boldsymbol{\theta}$ for a fixed $\mathbf{B}$**

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

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{X}-\mathbf{D}\Lambda\mathbf{B}||_F^2 + 4\sigma_n^2 log\boldsymbol{\theta}. \tag{9}$$

When the dictionary $\mathbf{D}$ is unitary (e.g., PCA or DCT basis)[2], the above optimization can be further simplified by the energy conservation property of unitary matrices. Note that

$$||\mathbf{X}-\mathbf{D}\Lambda\mathbf{B}||_F^2 = ||\mathbf{D}(\mathbf{A}-\Lambda\mathbf{B})||_F^2 = ||\mathbf{A}-\Lambda\mathbf{B}||_F^2. \tag{10}$$

where we have used $\mathbf{X} = \mathbf{D}\mathbf{A}$. If one substitutes Eq. (10) into Eq. (9), it becomes

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

Although $log\boldsymbol{\theta}$ is non-convex, we note that $log\boldsymbol{\theta} = \sum_{i=1}^{K} log\theta_i$ (due to $P(\boldsymbol{\theta}) = \prod_i P(\theta_i)$). Therefore one can approximate $f(\boldsymbol{\theta}) = log\boldsymbol{\theta}$ by its first-order Taylor expansion - that is

$$f(\boldsymbol{\theta}^{(k+1)}) = f(\boldsymbol{\theta}^{(k)}) + < \nabla f(\boldsymbol{\theta}^{(k)}), \boldsymbol{\theta}-\boldsymbol{\theta}^{(k)} > . \tag{12}$$

where $\boldsymbol{\theta}(k)$ denotes the solution obtained for the $k$-th iteration. Since the first-order derivative of $log\theta$ is $\frac{1}{\theta}$, one can approximate Eq. (11) by

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} ||\mathbf{A}-\Lambda\mathbf{B}||_F^2 + 4\sigma_n^2||\mathbf{W}\boldsymbol{\theta}||_1. \tag{13}$$

where $\mathbf{W} = diag(\frac{1}{\theta_i+\epsilon})$ ($\epsilon$ is a small positive) is the reweighting matrix that is often used in iterative reweighted $l_1$-minimization [10]. Therefore, our derivation here can also be viewed as a Bayesian interpretation of reweighting strategy - it is connected with the Jeffrey's prior we have adopted. Indeed, the optimization problem in the form of Eq. (13) has been widely studied in the literature and can be efficiently solved by iterative shrinkage algorithm as we will show next.

Since both $\Lambda$ and $\mathbf{W}$ are diagonal, we can decompose the minimization problem in Eq. (13) 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 $\mathbf{B} \in R^{n\times m}$. Eq. (13) can be rewritten as

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{K} ||(\boldsymbol{\alpha}^i)^T - (\boldsymbol{\beta}^i)^T\theta_i||_2^2 + 4\sigma_n^2 \sum_{i=1}^{K} \frac{\theta_i}{\theta_i+\epsilon}, \tag{14}$$

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

$$\theta_i^{(k+1)} = \underset{\theta_i}{\operatorname{argmin}} ||(\boldsymbol{\alpha}^i)^T - (\boldsymbol{\beta}^i)^T\theta_i||_2^2 + 4\sigma_n^2 \frac{\theta_i}{\theta_i^{(k)}+\epsilon}, \tag{15}$$

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

$$\theta_i^{(k+1)} = \frac{1}{\boldsymbol{\beta}^i(\boldsymbol{\beta}^i)^T}[\boldsymbol{\beta}^i(\boldsymbol{\alpha}^i)^T - \tau]_+, \tag{16}$$

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

[2]In this case, we have $n = K$ which is the same as the size of an image patch.

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

**B. 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

$$\mathbf{B} = \operatorname{argmin} ||\mathbf{Y} - \mathbf{D}\Lambda\mathbf{B}||_F^2 + \sigma_n^2||\mathbf{B} - \Gamma||_F^2. \tag{17}$$

Since both terms are $l_2$, the closed-form solution to Eq. (17) is essentially a Wiener filtering

$$\mathbf{B} = (\hat{\mathbf{D}}^T\hat{\mathbf{D}} + \sigma_n^2\mathbf{I})^{-1}(\hat{\mathbf{D}}^T\mathbf{Y} + \Gamma). \tag{18}$$

where $\hat{\mathbf{D}} = \mathbf{D}\Lambda$. Note that when $\mathbf{D}$ is orthogonal, Eq. (18) can be further simplified into

$$\mathbf{B} = (\Lambda^T\Lambda + \sigma_n^2\mathbf{I})^{-1}(\Lambda^T\mathbf{A} + \Gamma). \tag{19}$$

where $\Lambda^T\Lambda + \sigma_n^2\mathbf{I}$ is a diagonal matrix and therefore its inverse can be easily computed. The reconstruction of image data matrix $\mathbf{X}$ is then obtained by

$$\mathbf{X} = \mathbf{D}\Lambda\mathbf{B}. \tag{20}$$

Putting things together, a complete image restoration based on BSSC can be summarized as follows.

**Algorithm 1. BSSC-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 initial regularization parameters $\lambda$ and $\delta$;
- Outer loop: Iterate on $k = 1, 2, \cdots, k_{max}$
  - (a) Landweber iteration: $\hat{\boldsymbol{x}}^{(k+1/2)} = \hat{\boldsymbol{x}}^{(k)} + \delta\mathbf{H}^T(\boldsymbol{y} - \mathbf{H}\hat{\boldsymbol{x}}^{(k)})$;
  - (b) Image-to-patch transformation: obtain data matrices $\{X_l\}$'s for each exemplar (though kNN search);
  - (c) Inner loop: iterate on $l = 1, 2, ...L$;
    - (I) estimated biased means $\boldsymbol{\mu}$ using Eq. (5);
    - (II) update $\boldsymbol{\theta}_l i$ for fixed $\mathbf{B}_l$ using Eq. (16);
    - (III) update $\mathbf{B}_l$ for fixed $\boldsymbol{\theta}_l$ using Eq. (19);
    - (IV) Reconstruct $X_l$'s from $\boldsymbol{\theta}_l$ and $\mathbf{B}_l$ using Eq. (20);
  - (d) Patch-to-image transformation: obtain reconstructed $\hat{\boldsymbol{x}}^{(k+1)}$ from $\{X_l\}$'s;
- Output: $\hat{\boldsymbol{x}}^{(k+1)}$.

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, the Landweber iteration in Algorithm 1 degenerates into an iterative regularization technique useful for image denoising [**?**]. When $\mathbf{H}$ is a reduced blur matrix, Eq. (**??**) becomes the standard formulation of image super-resolution problem. When compared against our previous work NCSR [14], the key novelty of the new BSSC algorithm lies in the implementation of inner loop - the core problem as formulated in Eq. (**??** attempts to achieve simultaneous spatially local adaptation and nonlocal robustness. If the morale behind the story of NCSR (as well as many other nonlocal image models developed in the literature including BM3D and IDD-BM3D) is the advocate the importance of nonlocal self-similarity around edges and textures; the new message we attempt to convey through this work is that *locality still matters*. 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.

## III. EXPERIMENTAL RESULTS

We have compared BSSC-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) [13] and nonlocally centralized sparse representation (NCSR) denoising [14]. Table I includes the PSNR/SSIM comparison results for a collection of 11 images among four competing methods. It can be observed that BSSC clearly outperforms all other three for 10 out of 11 images (the only exception is the $house$ image for which IDD-BM3D slightly outperforms BSSC 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 BSSC is capable of striking a better tradeoff between exploiting local and nonlocal dependencies within those images. Fig. 1 show the visual comparison of deblurring results for a test image $barbara$ respectively. Since this image contains abundant textures, visual quality improvements achieved by BSSC are readily observable. Such experimental findings clearly suggest that the BSSC model is a stronger prior for the class of photographic images containing strong edges/textures.

## IV. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a new framework named Bayesian structured sparse coding that connects structured sparsity with Gaussian scale mixture. image restoration. BSSC can be viewed as the unification of two previous models: GSM and NCSR - it 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 original BSSC 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 applied to image restoration, BSSC leads to computationally efficient algorithms involving iterative shrinkage/filtering only. Our solution to BSSC 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 BSSC can both preserve the sharpness of edges and suppress undesirable artifacts more effectively than other competing approaches.

In addition to image restoration, BSSC 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 BSSC problem by reducing it to iterative reweighted $l_1$-minimization problem [10]. It is also possible to incorporate dictionary $\mathbf{D}$ into the optimization problem formulated in Eq. (3); 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 BSSC to recent advances in Bayesian nonparametric inference [12],[11].

**Acknowledgements**

This work was partially supported by grant NSF-ECCS-0968730, Natural Science Foundation of China under Grant 61100154, Grant 61227004, and the Fundamental Research Funds of the Central Universities of China under Grant BDY081424.

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

| | $9\times9$ uniform blur, $\sigma_n=\sqrt{2}$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Images | *Butterfly* | *Boats* | *C. Man* | *Starfish* | *Parrot* | *Lena* | *Barbara* | *Peppers* | *Leaves* | *House* | *Average* |
| FISTA [15] | 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 [13] | 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 [14] | 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 BSSC** | **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 [15] | 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 [13] | 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 [14] | 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 BSSC** | **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** |


**Fig. 1**. 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 [15] (PSNR=25.03, SSIM=0.7377); (d) IDD-BM3D [13] (PSNR=27.19 dB, SSIM=0.8231); (e) NCSR [14] (PSNR=27.91 dB, SSIM=0.8304); (f) Proposed BSSC (PSNR=**28.42**, SSIM=**0.8462**).

## V. REFERENCES


[1] 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.

[2] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” *CVPR*, vol. 2, pp. 60–65, 2005.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] J. Huang and T. Zhang, “The benefit of group sparsity,” *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[8] 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.

[9] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in *Proc. of ICCV*, 2011.

[10] E.J. Candes and M.B. Wakin and S.P. Boyd, “Enhancing sparsity by reweighted l1 minimization,” in *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.