Patch Group based Bayesian Learning for Blind Image Denoising

Jun Xu[1], Dongwei Ren[1,2], Lei Zhang[1]**, David Zhang[1]

[1]Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Abstract. Most existing image denoising methods assume to know the noise distributions, e.g., Gaussian noise, impulse noise, etc. However, in practice the noise distribution is usually unknown and is more complex, making image denoising still a challenging problem. In this paper, we propose a novel blind image denoising method under the Bayesian learning framework, which automatically performs noise inference and reconstructs the latent clean image. By utilizing the patch group (PG) based image nonlocal self-similarity prior, we model the PG variations as Mixture of Gaussians, whose parameters, including the number of components, are automatically inferred by variational Bayesian method. We then employ nonparametric Bayesian dictionary learning to extract the latent clean structures from the PG variations. The dictionaries and coefficients are automatically inferred by Gibbs sampling. The proposed method is evaluated on images with Gaussian noise, images with mixed Gaussian and impulse noise, and real noisy photographed images, in comparison with state-of-the-art denoising methods. Experimental results show that our proposed method performs consistently well on all types of noisy images in terms of both quantitative measure and visual quality, while those competing methods can only work well on the specific type of noisy images they are designed for and perform poorly on other types of noisy images. The proposed method provides a good solution to blind image denoising.

1   Introduction

Image denoising is an important problem in image processing and computer vision. Most existing methods are designed to deal with specific types of noise, e.g., Gaussian noise, mixed Gaussian and impulse noise, etc. Gaussian noise removal is a fundamental problem and has received intensive research interests with many representative work [1–8]. Gaussian noise removal is not only an independent task but also can be incorporated into other tasks, e.g., the removal of mixed Gaussian and impulse noise. The 'first-impulse-then-Gaussian' strategy is commonly adopted by methods of [9–11], which are designed specifically for mixed Gaussian and impulse noise. However, noise in real images is more complex than simple Gaussian or mixed Gaussian and impulse distribution. Besides, noise is usually unknown for existing methods. This makes image denoising still a challenging problem.

To the best of our knowledge, the study of blind image denoising can be traced back to the BLS-GSM model [12]. In [12], Portilla et al. proposed to use scale mixture

of Gaussian in overcomplete oriented pyramids to estimate the latent clean images. In [13], Portilla proposed to use a correlated Gaussian model for noise estimation of each wavelet subband. Based on the robust statistics theory [14], Rabie modeled the noisy pixels as outliers, which could be removed via Lorentzian robust estimator [15]. Liu et al. proposed to use 'noise level function' (NLF) to estimate the noise and then use Gaussian conditional random field to obtain the latent clean image [16]. Recently, Gong et al. proposed an optimization based method [17], which models the data fitting term by weighted sum of ℓ_1 and ℓ_2 norms and the regularization term by sparsity prior in the wavelet transform domain. Later, Lebrun el al. proposed a multiscale denoising algorithm called 'Noise Clinic' [18] for blind image denoising task. This method generalizes the NL-Bayes [19] to deal with signal and frequency dependent noise.

Despite the success of these methods, they have many limitations. On one hand, as suggested in [16, 18], Gaussian noise, assumed by [13, 15, 16], may be inflexible for more complex noise in real images. Hence, better approximation to the noise could bring better image denoising performance [16, 18]. On the other hand, the method [17] needs tune specific parameters for different types of noise. This makes the proposed method not a strictly "blind" image denoising. Based on these observations, it is still needed to design an robust and effective model for blind image denoising. Few assumption and no parameter tuning would bring extra points.

In this paper, we propose a new method for blind image denoising task. The key factor of success is to employ the Mixture of Gaussian (MoG) model to fit the patches extracted from the image. Since the noise in real image is unknown, we utilize variational Bayesian inference to determine all the parameters, including the number of components. That is, the noise is modeled by a MoG adapted to the testing image. This data driven property makes our model able to deal with blind noise. Then we employ the nonparametric Bayesian model [20] to reconstruct the latent clean structures in each component. Specificly, the beta-Bernoulli process [21, 22] is suitable for this task. In our proposed method, the noise in each component is assumed to be Gaussian. The parameters of the beta-Bernoulli process are automatically determined by nonparametric strategies such as the Gibbs sampling. The proposed model is tested on Gaussian noise, mixed Gaussian and impulse noise,and real noise in photographed images.

To summarize, our paper has the following contributions:
- We proposed a noval framework for blind image denoising problem;
- The proposed model is more robust on image denoising tasks than the competing methods;
- We demonstrated that, the performance of the Beta Process Factor Analysis (BPFA) model can be largely improved by structural clustering strategy and Non-local self similarity property;
- We achieve comparableor even better performance on blind image denoising tasks than the competing methods.

The remainder of this paper is organized as follows. Section II introduces the related workd. Section III introduces the patch group based Mixture of Gaussian model infered by variational Bayesian method. In section IV, we will formulate the proposed PG based nonparametric Bayesian dictionary learning model. Section V summarizes the overall algorithm. In section VI, we will present the experimental results, as well as discussions, on blind image denoising tasks. Section VII concludes this paper.

2   Related Work

Structural clustering are employed by many image denoising methods. For example, both EPLL [7] and PLE [23] utilized the Mixture of Gaussian (MoG) model [24] for clustering similar patches. NCSR [25] utilized the k-means algorithm. The recently proposed Patch Group Prior based Denoising (PGPD) [26] employ the MoG model to learn the non-local self-similarity (NSS) prior of images. However, all these methods need preset the number of clusters. In [27], the Beta-Bernoulli Factor Analysis (BPFA) model is further extended by nonparametric clustering processes such as Dirichlet process (DP) [28] or probit stick-breaking processes (PSBP) [29]. However, the authors in [27] pointed out that, the resulting 'clustering-aided' models achieved similar performance on image denoising to the original BPFA model [27]. In this paper, we demonstrate that, we could indeed improve the performance of Bayesian methods on image denoising tasks, if we utilize the structural clustering strategy properly.

    Dictionary learning is very useful for image denoising tasks. The dictionary can be chosen off-the-shelf (wavelets and curvelets) or can be learned from natural image patches. The seminal work of K-SVD [4] has demonstrated that dictionary learning not only can help achieves promising denoising performance, but also can be used in other image processing applications. In image denoising, dictionary learning is commonly combined with the non-local self-similarity (NSS) prior [2, 26], sparse prior [5, 6, 25], and low rank prior [30], etc. Though work well on Gaussian noise removal, the above methods perform poorly on other types of noises, especially noise in real noisy images.
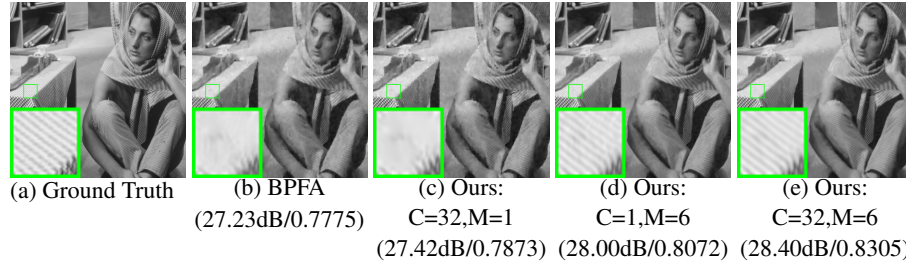
3   Patch Group based Bayesian Learning for Structural Clustering

3.1   Patch Group based Non-local Self Similarity Property

Non-local self similarity (NSS) is a common image prior employed by many image restoration methods [2, 5, 6, 25, 30, 26]. In PGPD [26], the patch group (PG) based NSS prior is learned on clean natural images for efficient image denoising. In this paper, however, we apply the PG based NSS prior directly on testing noisy images. Given a noisy image, we firstly extract image patches of size $p \times p$. Then, for each image patch, we find the M most similar patches to it in a large enough local window of size $W \times W$. The similarity measurement is based on Euclidean distance, i.e., ℓ_2 norm, which is commonly used in other methods [5, 6, 25, 30, 26]. In this work, we set $p = 8$, $M = 6$, $W = 31$. The PG is a set of similar patches $\{\mathbf{x}_m\}_{m=1}^{M}$, in which the $\mathbf{x}_m \in \mathbb{R}^{p^2 \times 1}$ is the mth patch vector. The group mean of this PG is $\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m$. The mth group mean subtracted patch vector is $\overline{\mathbf{x}}_m = \mathbf{x}_m - \boldsymbol{\mu}$. The $\overline{\mathbf{X}} \triangleq \{\overline{\mathbf{x}}_m\}$, $m = 1, ..., M$ is called the PG variations. In Figure 1, we show one example of PG, PG mean, and the PG variations after group mean subtraction. The PG mean is the main structure of this PG. Here, we proposed to embed this idea into the BPFA model [27] and compare the performance on the Gaussian noise removal. The results are listed in Figure 2. When comparing the images (b) and (d), we can see that the introduce of PG based NSS prior can really boost the performance on image denoising of the BPFA model. The improvements on PSNR is nearly 0.8dB and on SSIM is nearly 0.03. In particular, the image quality in (d) is much better than that in (b).

Fig. 1. An example of PG, PG mean (duplicated by five), and PG variations after mean subtraction. The PG is from the image 'House' corrupted by Gaussian noise.

(a) Ground Truth   (b) BPFA   (c) Ours: (d) Ours: (e) Ours:
                    (27.23dB/0.7775)   C=32,M=1   C=1,M=6   C=32,M=6
                                      (27.42dB/0.7873)   (28.00dB/0.8072)   (28.40dB/0.8305)

Fig. 2. Denoised images of *Barbara* and PSNR/SSIM results by BPFA and our models (the standard deviation of Gaussian noise is $\sigma = 40$).

3.2   PG based Bayesian Learning for Structural Clustering

Given a noisy image, we can extracted N PG variations, the m-th variation in the n-th PG is defined as $\overline{\mathbf{x}}_{n,m}$, where $n = 1, ..., N$ and $m = 1, ..., M$. We propose to utilize Mixture of Gaussian (MoG) with C components, to divide the sundry PG variations into different clusters. The relationship between PG variation $\overline{\mathbf{x}}_{n,m}$ and the MoG model relies on a latent variable defined as $\mathbf{z}_{n,m}$, which is consisted of C elements $z_{n,m,c}$ for $c = 1, ..., C$. If PG variation $\overline{\mathbf{x}}_{n,m}$ belongs to the cth component, $z_{n,m,c} = 1$; and otherwise $z_{n,m,c} = 0$. A natural problem is how to determine C, the number of clusters, as well as other parameters in the MoG model. A reliable solution is to initialize C as a large number and then employ the Dirichlet prior to estimate it [24].

In order to fully estimate the parameters of the MoG model, we can resort to the posterior distribution $p(\{\mathbf{z}_{n,m}\}|\{\overline{\mathbf{x}}_{n,m}\})$ given $\{\mathbf{z}_{n,m}\}$ ($n = 1, ..., N$ and $m = 1, ..., M$). Though its analytical form is not computationally tractable, we could use variational Bayesian inference as an approximational technique. Assuming that the mixing coefficients vector is $\boldsymbol{\pi}$, the conditional distribution of the latent variables $\{\mathbf{z}_{n,m}\}$ is

$$p(\{\mathbf{z}_{n,m}\}|\boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \prod_{c=1}^{C} \boldsymbol{\pi}_c^{z_{n,m,c}}. \tag{1}$$

The conditional distribution of the PG variations is

$$p(\{\overline{\mathbf{x}}_{n,m}\}|\{\mathbf{z}_{n,m}\}, \{\boldsymbol{\mu}_c\}, \{\boldsymbol{\Sigma}_c\}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \prod_{c=1}^{C} \mathcal{N}(\overline{\mathbf{x}}_{n,m}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)^{z_{n,m,c}} \tag{2}$$

where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are the corresponding mean vector and covariance matrix of the c-th component of the MoG model. After introducing corresponding priors over the parameters $\{\mathbf{z}_{n,m}\}$, $\{\boldsymbol{\mu}_c\}$ and $\{\boldsymbol{\Sigma}_c\}$, we can write the joint distribution of all the variables by

$$p(\{\overline{\mathbf{x}}_{n,m}\}, \{\mathbf{z}_{n,m}\}, \boldsymbol{\pi}, \{\boldsymbol{\mu}_c\}, \{\boldsymbol{\Sigma}_c\}) = p(\{\overline{\mathbf{x}}_{n,m}\}|\{\mathbf{z}_{n,m}\}, \{\boldsymbol{\mu}_c\},$$
$$\{\boldsymbol{\Sigma}_c\}) \times p(\{\mathbf{z}_{n,m}\}|\boldsymbol{\pi}) \times p(\boldsymbol{\pi}) \times p(\boldsymbol{\mu}_c|\boldsymbol{\Sigma}_c) \times p(\{\boldsymbol{\Sigma}_c\}), \tag{3}$$

where the mixing coefficients vector $\boldsymbol{\pi}$ is assumed to follow Dirichlet distribution.

Finally, we can perform variational maximization (M step) and expectation estimation of the responsibilities (E step) alternatively. By taking suitable parameters for the equation 3, we can re-write the lower bound as a function of the parameters. This lower bound can be used to determine the posterior distribution over the number of components C in the MoG model. Please refer to [24] for more details. We use the Matlab code implemented in Pattern Recognition and Machine Learning Toolbox `http://www.mathworks.com/matlabcentral/fileexchange/55826-pattern-recognition-and-machine-learning-toolbox`.

3.3 Discussion

Based on above description, we discuss as follows the advantages of the proposed model for PG based modeling directly on noisy images: Firstly, all the parameters of the proposed model are automatically estimated from the noisy image via variational Bayesian inference. This is a major advantage over PGPD [26]. Secondly, we demonstrated that structural clustering can indeed boost the performance of Bayesian methods such as BPFA [27]. This is demonstrated in the images (b) and (c) of Figure 2. Thirdly, the Non-local Self Similarity prior can be used to improve the performance of BPFA [27]. This can be demonstrated by comparing the images ((b), (c), (e)) or ((b), (d), (e)).

4 Patch Group based Bayesian Dictionary Learning

4.1 Truncated beta-Bernoulli Process for Dictionary Learning

Once we have clustered similar PG variations into different components, we can extract the latent clean PG variations using sparse or low rank image priors. These priors are also frequently employed by many state-of-the-art methods [5–7, 25, 30, 26] for image restoration tasks. We do not fixed the number of atoms in the dictionary learning. This is different from the previous methods, including PGPD. Instead, we set a large number K, which makes our Bayesian model a truncated beta-Bernoulli process. We employ the beta-Bernoulli process [20–22] to seek the sparse priors on infinite feature space.

In this paper, we express noisy PG variations $\overline{\mathbf{X}}$ as

$$\overline{\mathbf{X}} = \mathbf{DW} + \mathbf{V}, \tag{4}$$

where $\overline{\mathbf{X}} \in \mathbb{R}^{p^2 \times M}$, the coefficients vector $\mathbf{W} \in \mathbb{R}^{K \times M}$, and the noise term $\mathbf{V} \in \mathbb{R}^{p^2 \times M}$. The matrix $\mathbf{D} \in \mathbb{R}^{p^2 \times K}$ contains K dictionary atoms. The coefficients matrix $\mathbf{W}$ is represented by $\mathbf{W} = \mathbf{B} \odot \mathbf{S}$. The matrices $\mathbf{B}, \mathbf{S} \in \mathbb{R}^{K \times M}$ are the binary matrix and the coefficients matrix, respectively and $\odot$ is the element-wise product. We denote $\overline{\mathbf{x}}$ as any column of the PG variations $\overline{\mathbf{X}}$ and $\mathbf{w}, \mathbf{b} \in \{0, 1\}^K$, and $\mathbf{s}$ as corresponding columns of the coefficients matrix $\mathbf{W}$, the binary matrix $\mathbf{B}$, and the coefficients matrix $\mathbf{S}$. We denote $\{\mathbf{d}_k\}_{j=1}^K$ as columns of the dictionary $\mathbf{D}$, $\{\mathbf{s}_j\}_{j=1}^n$ as columns of $\mathbf{S}$, and $\{\mathbf{v}_j\}_{j=1}^n$ as columns of the noise matrix $\mathbf{V}$. Each column $\mathbf{w}$ is represented by a binary vector $\mathbf{b} \in \{0, 1\}^K$ and a coefficient vector $\mathbf{s} \in \mathbb{R}^{K \times 1}$, i.e., $\mathbf{w} = \mathbf{b} \odot \mathbf{s}$. Then we can impose suitable priors on these parameters, i.e., $\mathbf{b}$, $\{\mathbf{d}_k\}_{k=1}^K$, $\{\mathbf{s}_k\}_{k=1}^K$, and $\{\mathbf{v}_j\}_{j=1}^n$. The binary vector $\mathbf{b} \in \{0, 1\}^K$ denotes which of the columns (or atoms) in $\mathbf{D}$ are used for the representation of $\overline{\mathbf{x}}$. At beginning, we do not know the suitable K. We can set $K \to \infty$ and impose sparse prior on $\mathbf{b} \in \{0, 1\}^K$ to limit the number of

atoms in $\mathbf{D}$ used for representing each PG variation $\overline{\mathbf{x}}$ extracted from noisy images. The beta-Bernoulli process [20–22] provides a convenient way for this purpose. For inference convenience, we impose independent Gaussian priors on $\{\mathbf{d}_k\}_{k=1}^K$, $\{\mathbf{s}_k\}_{k=1}^K$, and $\{\mathbf{v}_j\}_{j=1}^n$.

The dictionary learning model for PG variations is

$$\overline{\mathbf{X}} = \mathbf{DW} + \mathbf{V}, \mathbf{W} = \mathbf{B} \odot \mathbf{S}, \tag{5}$$

where $\odot$ is element-wise product. The beta-Bernoulli process [22] for binary vector is:

$$\mathbf{b} \sim \prod_{k=1}^K Bernoulli(\pi_k), \boldsymbol{\pi} \sim \prod_{k=1}^K Beta(\frac{a}{K}, \frac{b(K-1)}{K}), \tag{6}$$

$$\mathbf{d}_k \sim \mathcal{N}(0, P^{-1}\mathbf{I}_P), \mathbf{s}_j \sim \mathcal{N}(0, \gamma_s^{-1}\mathbf{I}_K), \mathbf{v}_j \sim \mathcal{N}(0, \gamma_v^{-1}\mathbf{I}_P), \tag{7}$$

$$\gamma_s \sim Gamma(c, d), \gamma_v \sim Gamma(e, f), \tag{8}$$

where $\{a, b\}, \{c, d\}, \{e, f\}$ are the corresponding hyper parameters of the parameters $\boldsymbol{\pi}$, γ_s, and γ_v in the conjugate hyper priors in equations (6), (8). The inference procedures we take for the Bayesian model is Gibbs sampling [31] and we ignore the procedures here. As we can see, the noise estimation is integrated into the overall model in Gibbs sampling process. That is th reason why the proposed algorithm can be used to deal with blind noise. This is also different from the other non-blind image denoising methods [2, 5–7, 25, 30, 26], since they do not have the ability to estimate the noise within the noisy images.

We compare the proposed model with the BPFA model on the dictionary elements. With the help of PG based NSS prior and structural clustering, the dictionary learned by the proposed model is more representative and discriminative than that learned by the BPFA model [27]. Take the image "Barbara" for an example, it is corrupted by Gaussian noise with $\sigma = 40$. The initial number of dictionary elements is 512 for both methods. In Figure 2 (a), we demonstrate the dictionary elements learned by the BPFA model [27]. In the (b), (c), and (d) of Figure 2, we demonstrate the dictionary elements of three components learned by the proposed method. Noted that the BPFA learned from the original patches while the proposed method learned from patch group variations. It can be seen that these dictionary elements express the latent clean structures of the noisy PG variations in this component. For different Gaussian components, we can see that the number of elements are automatically determined by the nonparametric Bayesian inference. The dictionary elements in the subfigures (a) to (d) in Figure 3 are supplemented by black patches to make sure that these subfigures are square.

(a)   (b)   (c)   (d)
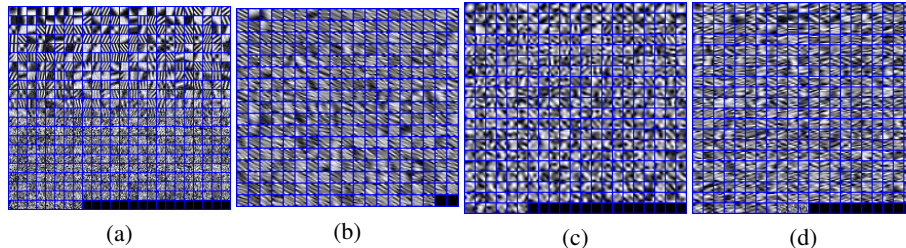
Fig. 3. The dictionary elements learned by the BPFA model ((a)) and the proposed method ((b), (c), and (d)).

Algorithm 1: The Overall Algorithm

Input: Noisy image $\mathbf{y}, a = b = 1, c = d = e = f = 10^{-6}$

1. Extract non-local similar patch groups $\{\mathbf{X}\}$;
2. For each PG, calculate its mean $\boldsymbol{\mu}_x$ and form PG variations $\overline{\mathbf{X}}$;
3. Estimate the parameters in MoG model via variational Bayesian inference;
4.   **for** each cluster
5.   **for** $t = 1 : IteNum$ **do** Gibbs Sampling:
6.   Sample $\mathbf{b}$ from equation (6)
7.   Sample $\mathbf{d}_k \sim \mathcal{N}(0, P^{-1}\mathbf{I}_P)$
8.   Sample $\mathbf{s}_j \sim \mathcal{N}(0, \gamma_s^{-1}\mathbf{I}_K), \gamma_s \sim Gamma(c, d)$
9.   Sample $\mathbf{v}_j \sim \mathcal{N}(0, \gamma_v^{-1}\mathbf{I}_P), \gamma_v \sim Gamma(e, f)$
10.   **end for**
11.   Recover each patch group via $\hat{\mathbf{X}} = \mathbf{DW} + \boldsymbol{\mu}_x$;
12.   **end for**
13. Average the recovered patch groups to form the recovered image $\hat{\mathbf{y}}$;
14. **Output:** The recovered image $\hat{\mathbf{y}}$.

4.2   Discussion

Our method are different from the method of BPFA [27]. Basicly, our model learns flexible number of dictionary atoms on patch groups while BPFA fixs the number of dictioary atoms learned over patches. This modification makes the dictionary more representative and discriminative than those in BPFA [27]. What's more, BPFA is only applied in (non-blind) denoising Gaussian noise while our model is applied to (blind) denoising multiple noises such as Gaussian noise, mixed Gaussian and impulse noise, and real noise. This will be demonstrated in the experimental section.

5   Summary of The Overall Algorithm

The overall algorithm is consisted of three parts. The first is the PG based Bayesian learning for automatically estimating the parameters of the MoG model, including the number of components. After the clustering, we employ a divide-and-conquer strategy. The second part is, for each component, we employ the PG based nonparametric Bayesian dictionary learning to sample the dictionary $\mathbf{D}$ as well as coefficients matrix $\mathbf{W}$. Then, the final denoised PG variations is simply calculated as $\hat{\mathbf{X}} = \mathbf{DW}$. That is the key factor why the proposed algorithm is able to deal with blind image denoising. After this is done for each component, the third part is to recover the denoised image by aggregating the denoised PGs. The overall algorithm is summarized in Algorithm 1. The Matlab source code of our algorithm can be downloaded at `http://www4.comp.polyu.edu.hk/~csjunxu/code/PGBL_BID.zip`.

6   Experiments

In this section, we perform image denoising experiments on various types of noise, including synthetic noise and real noise in real world images. The synthetic noise includes

Fig. 4. The 20 widely used test images.

additive white Gaussian noise (AWGN), mixed Gaussian and random value impulse noise (RVIN). The synthetic image denoising experiments are performed on 20 widely used natural images listed in Figure 4. In the synthetic noise removal experiments, we compare the proposed method with other state-of-the-art methods such as [27, 5, 26, 30, 9, 11, 18]. The BPFA [27] BM3D [5], PGPD [26], and WNNM [30] are designed especially for Gaussian noise removal. The Two Phase [9] and WESNR [11] are designed especially for mixed Gaussian and impulse noise removal. The "Noise Clinic" method [18] is designed especially for real noise removal. It is a state-of-the-art blind denoising method. We evaluate all these methods on PSNR, SSIM [32], and visual quality. On real noise removal, besides the above methods, we also compare with the commercial software Neat Image [33]. This software is embedded in Photoshop CS, a famous commercial software for image processing. The codes or executive package of these methods are provided on the corresponding websites. Since there is no ground truth in real noise removal, we only compare the visual quality of the recovered images by these methods.

6.1   Implementation Details

For the BM3D [5], WNNM [30], and PGPD [26], the input standard derivation of the noise is a key parameter. We employ a robust noise estimation method [34] to estimate the noise standard derivation for these methods. The methods Two Phase [9] and WESNR [11] do not need the noise level as input when performing denoising tasks. But they need some preprocessing for the noise removal tasks. However, our proposed method doesn't need noise estimation nor image preprocessing.

In the proposed method, we set the initial number of components C as 32. The final number of components will be automatically determined by variational Bayesian inference introduced in section 2. The patch size is fixed as 8×8, so the dimension of each patch vector is 64. The number of dictionary atoms K is set to be 256 when number of PG variations in this component is less than 10^4, otherwise the number is set to 512. The hyperparameters are fixed as $a = b = 1$, $c = d = e = f = 10^{-6}$ in all the experiments. We do not tune these paramters in our experiments. The number of patches in a group is set to 6. We set the maximal iteration number as 10. The $IteNum$ is set as 50. The proposed algorithm will be terminated when the noise variance of each component is less than or equal to 1.

6.2   Additive White Gaussian Noise Removal

Here, we compare the proposed method on Gaussian noise removal with other competing methods: BPFA [27], BM3D [5], and PGPD [26], WNNM [30], Two Phase [9],

WNSER [11], Noise Clinic [18]. As a common experimental setting, we add additive white Gaussian noise with zero mean and standard deviation σ to the test images. The denoising experiments are performed on multiple noise levels of $\sigma = 30, 40, 50, 75$.

The averaged results on PSNR and SSIM are listed in Tables 1 and 2. We can see that the PSNR and SSIM results of our proposed method are much better than the BPFA, Two Phase, and WESNR, Noise Clinic methods. The results of the proposed method are comparable to BM3D when the noise levels are higher than 30. Though getting inferior results on PSNR and SSIM when compared to WNNM and PGPD, the proposed method can achieve similar or even better performance when compared with other methods. For instance, from Figure 4 and Figure 5, we can see that the proposed method generates better image quality and less artifacts on the image "House" and "Hill" than the other methods. Considering that the proposed method is fully blind, it is more convincing to see that our proposed method achieves better image quality results than the Noise Clinic [18], which is a state-of-the-art blind image denoising method. We want to mention again that BM3D, PGPD, and WNNM can only work well on the Gaussian noise, which they are designed for, but perform poorly on other types of noise, which will be demonstrated later.

Table 1. Average PSNR(dB) results of different algorithms on 20 natural images corrupted by Gaussian noise.
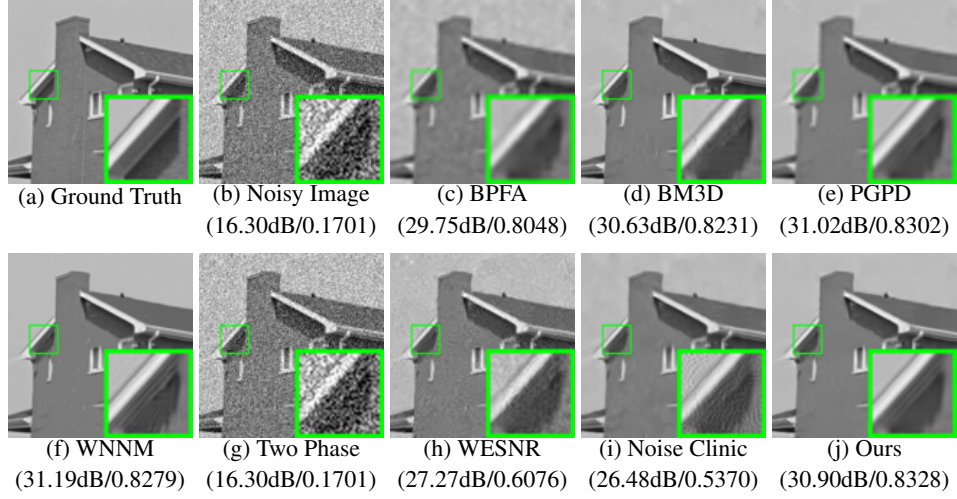
σ	BPFA	BM3D	WNNM	PGPD	Two Phase	WESNR	Noise Clinic	Ours
30	28.81	29.11	29.35	29.13	18.84	27.44	26.81	28.65
40	27.45	27.68	28.03	27.88	16.52	25.07	24.90	27.59
50	26.38	26.80	27.06	26.89	14.80	22.02	23.47	26.68
75	24.40	25.04	25.25	25.11	11.95	9.02	21.31	24.91

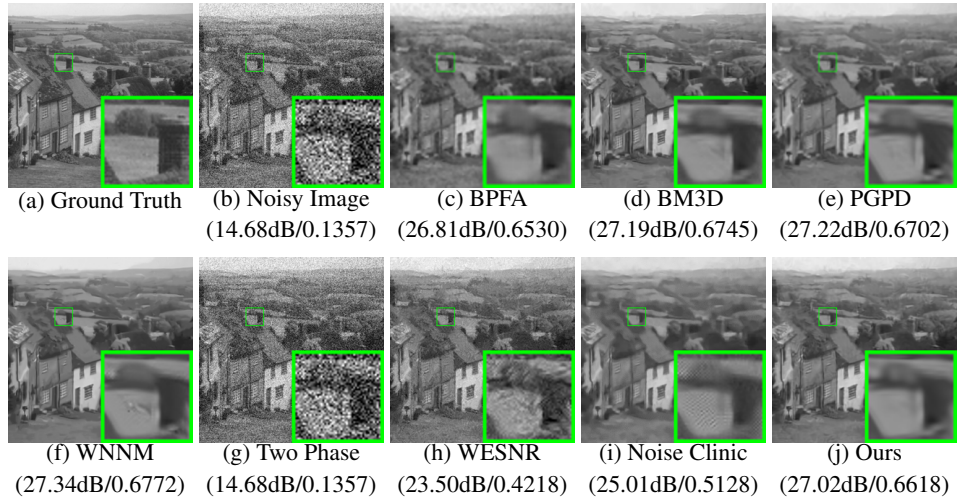**Table 2.** Average SSIM results of different algorithms on 20 natural images corrupted by Gaussian noise.

σ	BPFA	BM3D	WNNM	PGPD	Two Phase	WESNR	Noise Clinic	Ours
30	0.7988	0.8108	0.8156	0.8089	0.3137	0.7430	0.6520	0.7971
40	0.7576	0.7706	0.7784	0.7747	0.2361	0.6063	0.5631	0.7668
50	0.7216	0.7430	0.7510	0.7435	0.1855	0.4483	0.4990	0.7382
75	0.6470	0.6803	0.6897	0.6825	0.1147	0.1749	0.4289	0.6766

6.3   Mixed Gaussian and Impulse Noise Removal

Here, we compare the proposed method on mixed Gaussian and impulse noise with the compared methods [27, 5, 26, 30, 9, 11, 18]. We consider Random Value Impulse Noise (RVIN) here. The pixels in the testing image corrupted by RVIN is distributed between 0 and 255. This is much harder than salt and pepper noise, which is only 0 or 255 values. In the synthetic noise, the standard derivations of the Gaussian noise are $\sigma = 10, 20$ and the ratios of the impulse noise are $0.15, 0.30$, respectively. The RVIN noise is generated by the "impulsenoise" function used by WESNR [11].

(a) Ground Truth	(b) Noisy Image	(c) BPFA	(d) BM3D	(e) PGPD
	(16.30dB/0.1701)	(29.75dB/0.8048)	(30.63dB/0.8231)	(31.02dB/0.8302)
(f) WNNM	(g) Two Phase	(h) WESNR	(i) Noise Clinic	(j) Ours
(31.19dB/0.8279)	(16.30dB/0.1701)	(27.27dB/0.6076)	(26.48dB/0.5370)	(30.90dB/0.8328)

Fig. 5. Denoised images of *House* and PSNR/SSIM results by different methods (the standard deviation of Gaussian noise is $\sigma = 40$).

(a) Ground Truth	(b) Noisy Image	(c) BPFA	(d) BM3D	(e) PGPD
	(14.68dB/0.1357)	(26.81dB/0.6530)	(27.19dB/0.6745)	(27.22dB/0.6702)
(f) WNNM	(g) Two Phase	(h) WESNR	(i) Noise Clinic	(j) Ours
(27.34dB/0.6772)	(14.68dB/0.1357)	(23.50dB/0.4218)	(25.01dB/0.5128)	(27.02dB/0.6618)

Fig. 6. Denoised images of *Hill* and PSNR/SSIM results by different methods (the standard deviation of Gaussian noise is $\sigma = 50$).

For BPFA [27], BM3D [5], and PGPD [26], WNNM [30], they are designed to deal with Gaussian noise. Hence, we still employ the noise estimation method [34] to estimate the noise levels σ. We also compare the proposed method with the methods Two Phase [9] and WESNR [11], which are designed especially for the mixed Gaussian and RVIN noise. Noted that both Two Phase [9] and WESNR [11] employ Adaptive Median Filter (AMF) [35] to preprocess the image before performing image denoising task. We also compare with the Noise Clinic [18].

The results on PSNR and SSIM are listed in Table 3 and Table 4. As we can see, the performance of the proposed method is comparable or better than other methods. For the visual quality comparison, the proposed method can generate better results than other methods. Take the image "Couple" for an example, from the Figure 7, the proposed method removes the noise clearly while all other methods remain some noise or generate some artifacts. From the results on "Barbara" listed in Figure 8, the proposed method achieves higher SSIM and generates better image quality than all other methods. We have to mention again that the Two Phase and WESNR are two state-of-the-art methods designed especially for the mixed Gaussian and impulse noise. But they perform poorly on other types of noises such as Gaussian.

Table 3. Average PSNR(dB) results of different algorithms on 20 natural images corrupted by mixed of Gaussian and RVIN noise.
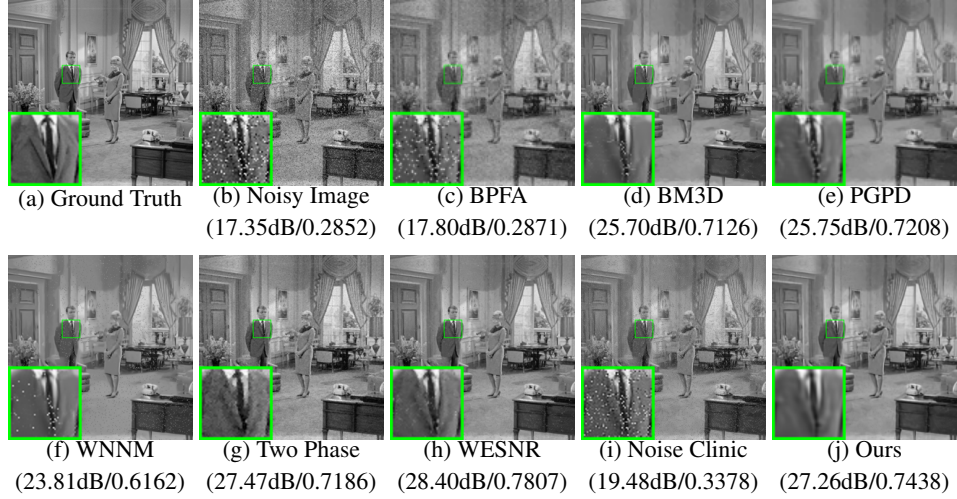
σ, Ratio	BPFA	BM3D	WNNM	PGPD	Two Phase	WESNR	Noise Clinic	Ours
10, 0.15	17.12	25.18	22.98	25.41	27.28	27.37	18.66	27.17
10, 0.30	14.19	21.80	21.40	21.74	26.12	21.50	16.44	22.17
20, 0.15	17.62	25.13	23.57	25.33	24.43	27.24	19.66	26.12
20, 0.30	17.61	21.73	21.40	21.64	23.61	22.69	14.46	21.89

Table 4. Average SSIM results of different algorithms on 20 natural images corrupted by mixed of Gaussian and RVIN noise.
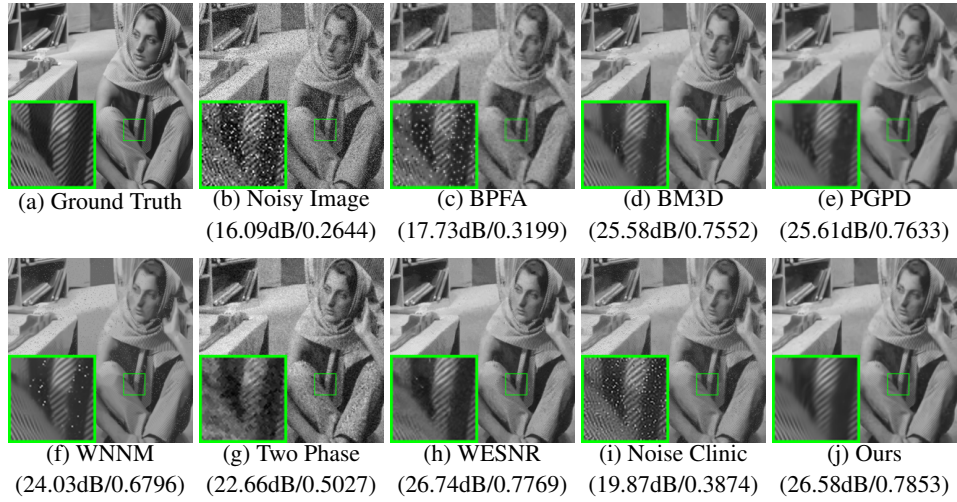
σ, Ratio	BPFA	BM3D	WNNM	PGPD	Two Phase	WESNR	Noise Clinic	Ours
10, 0.15	0.2749	0.7037	0.5806	0.7242	0.7091	0.7499	0.3198	0.7459
10, 0.30	0.1576	0.6444	0.6038	0.6405	0.6783	0.4970	0.2042	0.6559
20, 0.15	0.2821	0.7132	0.6190	0.7220	0.5468	0.7568	0.3423	0.7312
20, 0.30	0.2821	0.6414	0.6310	0.6371	0.5183	0.5871	0.1346	0.6470

6.4 Real Noisy Image Denoising

In this section, we will test the proposed method on real noise removal. Since the color images are in RGB channels, we firstly transform the color images into YCbCr channels, then perfrom denoising on the Y channel, and finally transform the denoised YCbCr channel image back into RGB channels. The denoised images are cropped to size of 800×600 for better visualization. We do not compare with WNNM [30] and Two Phase [9] here since they achieve worse denoising quality than the other methods such as BM3D [5], WESNR [11], and Noise Clinic [18]. For BM3D and PGPD, the input noise level σ is still estimated by [34]. We also compare with the Neat Image [33], a commercial software embedded in Photoshop CS. In this paper, we take three real noisy images for examples, which are "SolvayConf1927", "Girls", and "Windmill". The image "SolvayConf1927" is an old image provided by the Noise Clinic website on
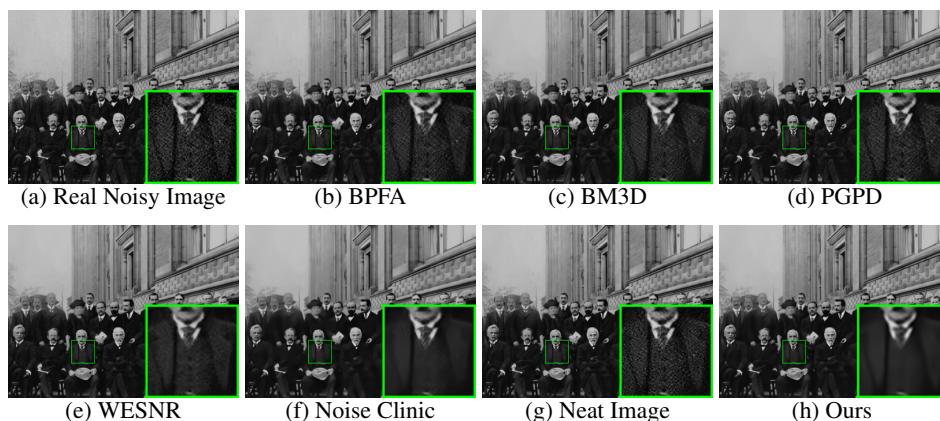
(a) Ground Truth	(b) Noisy Image	(c) BPFA	(d) BM3D	(e) PGPD
	(17.35dB/0.2852)	(17.80dB/0.2871)	(25.70dB/0.7126)	(25.75dB/0.7208)
(f) WNNM	(g) Two Phase	(h) WESNR	(i) Noise Clinic	(j) Ours
(23.81dB/0.6162)	(27.47dB/0.7186)	(28.40dB/0.7807)	(19.48dB/0.3378)	(27.26dB/0.7438)

Fig. 7. Denoised images of *Couple* by different methods (the mixed Gaussian and Random Value Impulse Noise is with $\sigma = 10$ and ratio 0.15). The images are better viewed by zooming in on screen.

(a) Ground Truth	(b) Noisy Image	(c) BPFA	(d) BM3D	(e) PGPD
	(16.09dB/0.2644)	(17.73dB/0.3199)	(25.58dB/0.7552)	(25.61dB/0.7633)
(f) WNNM	(g) Two Phase	(h) WESNR	(i) Noise Clinic	(j) Ours
(24.03dB/0.6796)	(22.66dB/0.5027)	(26.74dB/0.7769)	(19.87dB/0.3874)	(26.58dB/0.7853)

Fig. 8. Denoised images of *Barbara* by different methods (the mixed Gaussian and RVIN noise is with $\sigma = 20$ and ratio 0.15). The images are better viewed by zooming in on screen.

IPOL [36] while the other two images are provided by the Neat Image website [33]. The denoised images are evaluated in Figures 9 to 11.
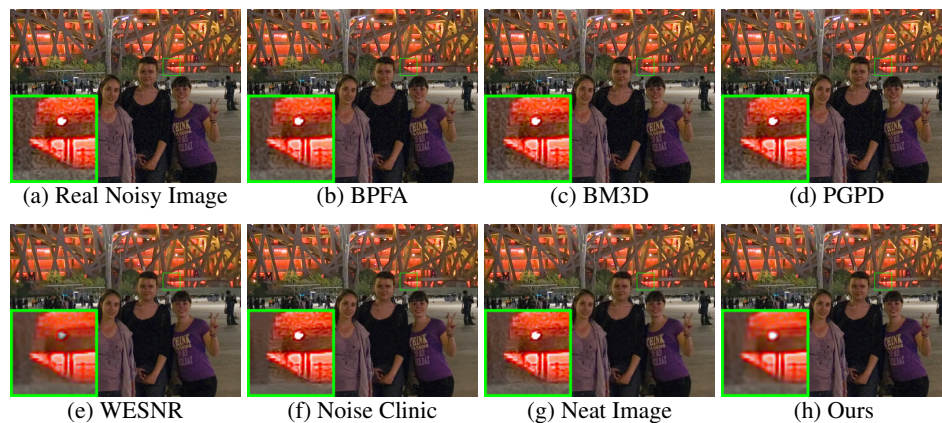
From the results listed in Figures 9, 10, and 11, we can see that the proposed method can remove real noise while preserving details better than other methods. For example, in the Figure 10, the image "Girls" are damaged by night shot noise and hardly recoverable. However, the proposed method can denoise the heavy noise, restore the details under the noise, and make the image looks better. The details of the building in the background is zoomed in to demonstrate the advantages of our proposed method than other compared methods. Another example is the image "Windmill", the noise in night sky is traditionally difficult to remove, since it is very hard to distinct the image noise from the faint stars in the sky. Our method not only reduces the noise while generating much less artifacts, but also preserves details, i.e., the faint stars, better than other methods.

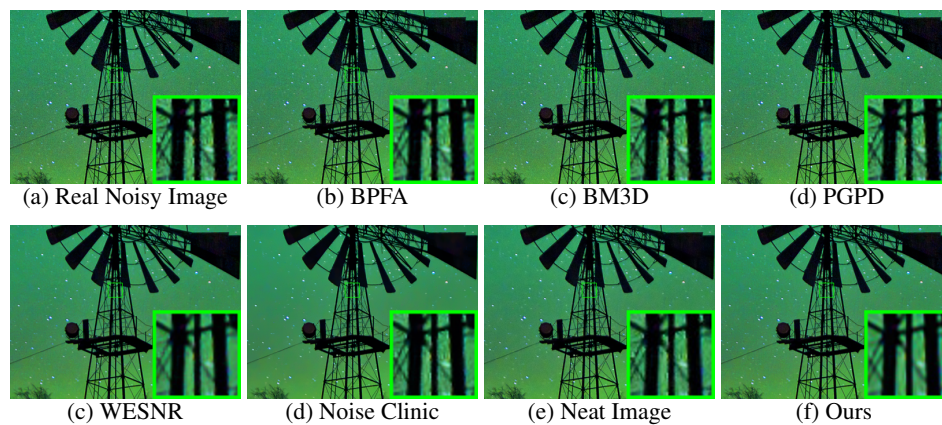| (a) Real Noisy Image | (b) BPFA | (c) BM3D | (d) PGPD |
| (e) WESNR | (f) Noise Clinic | (g) Neat Image | (h) Ours |

Fig. 9. Denoised images of the old image "SolvayConf1927" by different methods. The images are better viewed by zooming in on screen.

7   Conclusion

Image denoising is a commonly encountered problem in real life. However, most denoising methods [27, 5–7, 30, 26, 9, 11] need to know the noise distributions, such as Gaussian noise or mixed Gaussian and impulse noise, as well as the noise intensity. In this paper, we developed a novel blind image denoising method by patch group (PG) based nonlocal self-similarity prior modeling. We modeled the PG variations by Mixture of Gaussians [24] whose parameters, including its number of components, are inferred by variational Bayesian method. For each component, we employed nonparametric Bayesian dictionary learning [27, 21, 22] to reconstruct the latent clean images. The proposed method can deal with unknown or arbitrary noise without knowing the noise distribution. From the experimental results on removing Gaussian noise, mixed Gaussian and random value impulse noise , and the noise in real images, we demonstrated that the proposed method achieves comparable PSNR/SSIM measurements and even better visual quality than those methods [27, 5, 30, 26, 9, 11, 18, 33], which are specially designed for specific types of noises.

| (a) Real Noisy Image | (b) BPFA | (c) BM3D | (d) PGPD |

| (e) WESNR | (f) Noise Clinic | (g) Neat Image | (h) Ours |

Fig. 10. Denoised images of the image "Girls" by different methods. The images are better viewed by zooming in on screen.

| (a) Real Noisy Image | (b) BPFA | (c) BM3D | (d) PGPD |

| (c) WESNR | (d) Noise Clinic | (e) Neat Image | (f) Ours |

Fig. 11. Photo courtesy of Alexander Semenov. Denoised images of the image "Windmill" by different methods. The images are better viewed by zooming in on screen.

References

1. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60** (1992) 259–268
2. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. CVPR (2005) 60–65
3. Roth, S., Black, M.J.: Fields of Experts. International Journal of Computer Vision **82** (2009) 205–229
4. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. Image Processing, IEEE Transactions on **15** (2006) 3736–3745
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. Image Processing, IEEE Transactions on **16** (2007) 2080–2095
6. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. ICCV (2009) 2272–2279
7. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. ICCV (2011) 479–486
8. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2392–2399
9. Cai, J.F., Chan, R.H., Nikolova, M.: Fast two-phase image deblurring under impulse noise. Journal of Mathematical Imaging and Vision **36** (2010) 46–53
10. Xiao, Y., Zeng, T., Yu, J., Ng, M.K.: Restoration of images corrupted by mixed gaussian-impulse noise via l1l0 minimization. Pattern Recognition **44** (2011) 1708 – 1720
11. Jiang, J., Zhang, L., Yang, J.: Mixed noise removal by weighted encoding with sparse non-local regularization. Image Processing, IEEE Transactions on **23** (2014) 2651–2662
12. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussians in the wavelet domain. Image Processing, IEEE Transactions on **12** (2003) 1338–1351
13. Portilla, J.: Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. Image Processing, 2004. ICIP '04. 2004 International Conference on **2** (2004) 1217–1220
14. Huber, P.J.: Robust statistics. Springer (2011)
15. Rabie, T.: Robust estimation approach for blind denoising. Image Processing, IEEE Transactions on **14** (2005) 1755–1765
16. Liu, C., Szeliski, R., Kang, S.B., Zitnick, C.L., Freeman, W.T.: Automatic estimation and removal of noise from a single image. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 299–314
17. Gong, Z., Shen, Z., Toh, K.C.: Image restoration with mixed or unknown noises. Multiscale Modeling & Simulation **12** (2014) 458–487
18. Lebrun, M., Colom, M., Morel, J.M.: Multiscale image blind denoising. Image Processing, IEEE Transactions on **24** (2015) 3149–3161
19. Lebrun, M., Buades, A., Morel, J.M.: A nonlocal bayesian image denoising algorithm. SIAM Journal on Imaging Sciences **6** (2013) 1665–1688
20. Hjort, N.L.: Nonparametric bayes estimators based on beta processes in models for life history data. The Annals of Statistics (1990) 1259–1294
21. Thibaux, R., Jordan, M.I.: Hierarchical beta processes and the indian buffet process. In: International conference on artificial intelligence and statistics. (2007) 564–571
22. Paisley, J., Carin, L.: Nonparametric factor analysis with beta process priors. In: Proceedings of the 26th Annual International Conference on Machine Learning, ACM (2009) 777–784

23. Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity. Image Processing, IEEE Transactions on **21** (2012) 2481–2499
24. Bishop, C.M.: Pattern recognition and machine learning. (2006)
25. Dong, W., Zhang, L., Shi, G., Li, X.: Nonlocally centralized sparse representation for image restoration. Image Processing, IEEE Transactions on **22** (2013) 1620–1630
26. Xu, J., Zhang, L., Zuo, W., Zhang, D., Feng, X.: Patch group based nonlocal self-similarity prior learning for image denoising. The IEEE International Conference on Computer Vision (ICCV) (2015) 244–252
27. Zhou, M., Chen, H., Ren, L., Sapiro, G., Carin, L., Paisley, J.W.: Non-parametric bayesian dictionary learning for sparse image representations. In: NIPS. (2009) 2295–2303
28. Ferguson, T.S.: A bayesian analysis of some nonparametric problems. The annals of statistics (1973) 209–230
29. Ren, L., Du, L., Carin, L., Dunson, D.: Logistic stick-breaking process. The Journal of Machine Learning Research **12** (2011) 203–239
30. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. CVPR (2014) 2862–2869
31. Zhou, M., Chen, H., Paisley, J., Ren, L., Li, L., Xing, Z., Dunson, D., Sapiro, G., Carin, L.: Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. Image Processing, IEEE Transactions on **21** (2012) 130–144
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on **13** (2004) 600–612
33. ABSoft, N.: Neat image. (`https://ni.neatvideo.com/home`)
34. Liu, X., Tanaka, M., Okutomi, M.: Single-image noise level estimation for blind denoising. Image Processing, IEEE Transactions on **22** (2013) 5226–5237
35. Hwang, H., Haddad, R.: Adaptive median filters: new algorithms and results. Image Processing, IEEE Transactions on **4** (1995) 499–502
36. Lebrun, M., Colom, M., Morel, J.M.: The noise clinic: a blind image denoising algorithm. (`http://www.ipol.im/pub/art/2015/125/`) Accessed 01 28, 2015.