Robust Mixed Noise Removal with Non-parametric Bayesian Sparse Outlier Model

Peixian Zhuang [1], Wei Wang [1], Delu Zeng [1], Xinghao Ding *[1]

[1] *College of information science and engineering, Xiamen University, China*

* dxh@xmu.edu.cn

***Abstract*—This paper proposes a novel non-parametric Bayesian framework for solving mixed noise removal problem. In order to removing unstable effects of outlier noise such as salt-and-pepper in the training data, we decompose the observed data model into three components terms of ideal data, Gaussian noise and sparse outlier. And the proposed model employs spike-slab sparse prior to find the sparser coefficients of desired data term and outlier noise. Note that the proposed non-parametric Bayesian model can infer the noise statistics from the training data and have been robust to the mixed noise without tuning of model parameters. Experimental results demonstrate our proposed algorithm performs well with mixed noise and achieves better performance over other state-of-the-art methods.**

I. Introduction

Image denoising problems become well-known and difficult based on the fact that images can be easily contaminated by different types of noises produced from environmental conditions, sensor disturbance or uneven illuminations in practical applications [1], [2]. And previous works of denoising mainly aims at removing Gaussian or salt-and-pepper noise respectively, but few literatures focus on the removal of mixed noises. Variational methods are well-known for Gaussian noise removal but oversmooth image textures details, whereafter they combine with nonlocal models for better preserve textures details [3], [4], [5]. KSVD [6] employs the iterative method to adaptively learn a dictionary to better fit the training data. BM3D [7] uses block matching and 3D collaborative filter to enhance the sparsity and achieve better performance for denoising. BPFA [1]automatically estimates the Gaussian noise variance of corrupted images by using nonparametric Bayesian technique with beta process [8], [9]. Note that the above sparse representation-based image denoising algorithms (e.g.[1], [6], [7], [10], [11], [12]) are not robust to outlier noise in images since most of them are developed based on the assumption that the images are corrupted with the white noise. However, salt-and-pepper noise is another type that drawn from the impulse distribution [13].And this commonly-encountered outlier noise (i.e., salt-and-pepper noise) in the images will produce the unreliable results in the sparse representation. As far as we known, there are few sparse representation models that are developed for the mixed noise removal. [14] presents a sparse outlier model, which is not suitable for the mixed denoising. [2] builds the weighted KSVD algorithm(WKSVD) to remove the mixed noise, but it produces several recovery errors of texture structures during the noise-removal processing. [15] proposes a robust Bayesian RVM algorithm and achieves better performance than the original RVM for handling the outlier noise. However, this method needs the fixed dictionary, which well not adapts to various kinds of data. Henceforth, it is desirable to develop an efficient sparse representation-based model for mixed noise removal.

2014 IEEE 16th International Workshop on Multimedia Signal Processing (MMSP), Sep. 22-24, 2014, Jakarta, Indonesia.
978-1-4799-5896-2/14/$31.00

A novel nonparametric Bayesian model is proposed in this paper for mixed noise removal, in which the spike-slab sparse prior [16], [17], [18]can be used to find the much sparser and more accurate solution. Unlike other existing sparse representation methods that assume the noise to be Gaussian noise, the proposed algorithm considers the noise as the mixture of Gaussian and outlier noise. In our approach, the observed data is decomposed into three components: a desired data term, a Gaussian noise term and a sparse outlier term. A dictionary, parameters for sparse coefficients of desired data and outlier noise can be learned and inferred from the training data. Note that another merit of the proposed model is that it can automatically determine other hyper-parameters. Experimental results demonstrate that the proposed nonparametric Bayesian model can be effective to remove the mixed noise and markedly better than other state-of-the-art methods on various kinds of images. The remainder of the paper is organized as follows. Firstly, we present the novel nonparametric Bayesian model in section 2. Then the model inference process is proposed in section 3 in detail. And the experimental results are shown and discussed in section 4. Finally, the conclusion and discussion are given in the last section.

II. NONPARAMETRIC BAYESIAN MODEL CONSTRUCTION

In this work, we decompose the noisy data Y into three components: ideal image term X, Gaussian noise term N and outlier noise term S. Given observation $Y = X + N + S$, Our objective is to recover as high quality as possible. And we use image patches to construct the model due to the limitations of computing time and memory. Let $X = [x_1, x_2, ..., x_N]$, where x_i denotes an image patch, i.e.$x_i = D\alpha_i$, D means the dictionary learned from the training data, and the atom d_i is drawn from a Gaussian distribution. α_i is the sparse representation coefficients, and π_k obeys the Bernoulli distribution as $Beta(a_0, b_0)$. n_i is the noise term and drawn from a Gaussian distribution. The sparse outlier is s_i, we impose s_i a sparse spike-slab prior: $s_{ip} \sim (1-q_{ip})\delta + q_{ip}N(0,\ \gamma_s^{-1}),\ \ p = 1, 2, ..., P$. P is each observed data dimension, q_{ip} is the probability that the pixel is outlier, which is drawn from a beta distribution with m_0 and n_0 two parameters: $q_{ip} \sim Beta(m_0, n_0)$. As the expectation of beta distribution is $E(q_{ip}) = \frac{m_0}{m_0+n_0}$, we set $\frac{m_0}{m_0+n_0} << 1$ to guarantee a small q_i so that the corresponding s_{ip} has a big probability to be spike term. The nonparametric Bayesian model can be summarized as follows:

$$\begin{aligned}
&y_i = \mathbf{D}\alpha_i + n_i + s_i \\
&d_k \sim N(0, P^{-1}\mathbf{I}_P), n_i \sim N(0, \gamma_n^{-1}) \\
&\alpha_i \sim \prod_{k=1}^{K} ((1-\pi_k)\delta + \pi_k N(0,\ \gamma_\alpha^{-1})) \\
&s_i \sim \prod_{p=1}^{P} ((1-q_p)\delta + q_p N(0,\ \gamma_s^{-1})) \\
&\pi_k \sim Beta(a_0, b_0), q_p \sim Beta(m_0, n_0) \\
&\gamma_\alpha \sim \Gamma(c_0, d_0), \gamma_s \sim \Gamma(e_0, f_0), \gamma_n \sim \Gamma(g_0, h_0)
\end{aligned} \tag{1}$$

Although outlier noise and sparse coefficients have the same form, they represent different meanings. In the nonparametric Bayesian model, all train data $\mathbf{y} = [y_1, y_2, ..., y_N]$, use a common dictionary D. π_k is the probability that the atom d_k is used, so all data share the same π_k. While each outlier noise is independent, therefore, each q_{ip} corresponds to a noise pixel. Non-informative gamma hyper-priors are typically imposed on γ_α,γ_s and γ_n, which are used to infer the precision of their corresponding variances.

III. MODEL INFERENCE

In our nonparametric Bayesian model, every parameter has been imposed on a prior. The posterior distribution can be inferred from the training data and Bayesian inference. In this section, Markov Chain Monte Carlo (MCMC) algorithm has been employed to compute the approximate inference. And samples are drawn from the following conditional distribution:

A. Sample data term

1. sample D:

$$p(d_k|-) \sim \prod_{i=1}^{N} N(y_i|D\alpha_i, {\gamma_n}^{-1}\mathbf{I}_P)N(d_k|0, P^{-1}\mathbf{I}_P) \tag{2}$$

The posterior distribution of d_k also obeys the Gaussian distribution:

$$p(d_k|-) \sim N(\mu_{d_k}, \Sigma_{d_k}) \tag{3}$$

And its corresponding mean and variance can be expressed as:

$$\Sigma_{d_k} = (PI + \gamma_n \sum_{i=1}^{N} \alpha_{ik}^2)^{-1} \tag{4}$$

$$\mu_{d_k} = \gamma_n \Sigma_{d_k} \sum_{i=1}^{N} \alpha_{ik} z_{ik} \tilde{y}_i^{-k} \tag{5}$$

where $\tilde{y}_i^{-k} = y_i - D\alpha_i - s_i + d_k\alpha_{ik}$.

2. sample α_{ik}:

$$p(\alpha_{ik}|-) \sim N(\tilde{y}_i^{-k}|d_k\alpha_{ik}, \gamma_n^{-1}\mathbf{I}_P)((1-\pi_k)\delta + \pi_k N(0,\ \gamma_\alpha^{-1})) \tag{6}$$

The posterior distribution of α_{ik} can be inferred as:

$$p(\alpha_{ik}|-) \sim (1-\pi_{ik}^*)\delta + \pi_{ik}^* N(\mu_{\alpha_{ik}}^*, \Sigma_{\alpha_{ik}}^*) \tag{7}$$

where

$$\Sigma_{\alpha_{ik}}^* = (\gamma_\alpha + \gamma_n d_k^T d_k)^{-1} \tag{8}$$

$$\mu_{\alpha_{ik}}^* = \gamma_n \Sigma_{\alpha_{ik}}^* (d_k^T \tilde{y}_i^{-k}) \tag{9}$$

$$\ln \frac{\pi_{ik}^*}{1-\pi_{ik}^*} = \frac{\pi_{ik}}{1-\pi_{ik}} \frac{N(0|0, \gamma_\alpha^{-1})}{N(0|\mu_{\alpha_{ik}}^*, \Sigma_{\alpha_{ik}}^*)} \tag{10}$$

3. sample π_k:

$$p(\pi_k|-) \sim Beta(a_0 + \sum_{i=1}^{N} 1_{\alpha_{ik}\neq 0}, b_0 + N - \sum_{i=1}^{N} 1_{\alpha_{ik}\neq 0}) \tag{11}$$

4. sample γ_α:

$$\begin{aligned}
p(\gamma_\alpha|-) &\sim \Gamma(\gamma_\alpha|c_0, d_0) \prod_{i=1}^{N} N(\alpha_i|0, \gamma_\alpha^{-1} I_K) \\
&\sim \Gamma(c_0 + \frac{1}{2}KN, d_0 + \frac{1}{2}\sum_{i=1}^{N} \alpha_i^T \alpha_i)
\end{aligned} \tag{12}$$

B. outlier noise term

Exclude the ideal data before the calculation of outlier noise parameters and thus let

$$\hat{y}_i = y_i - D\alpha_i \tag{13}$$

1. sample s_i: The sampling process of s_i is same as that of α_{ik}. The distributions of s_i is described as follows:

$$p(s_i|-) \sim q_p N(s_i|\mu_{s_i}^*, \Sigma_{s_i}^*) + (1-q_p)\delta \tag{14}$$

where

$$\Sigma_{s_i}^* = (\gamma_s + \gamma_n)^{-1} \tag{15}$$

$$\mu_{s_i}^* = \gamma_n (\Sigma_{s_i}^* \circ \hat{y}_i) \tag{16}$$

$$\ln \frac{q_{ip}^*}{1-q_{ip}^*} = \frac{q_{ip}}{1-q_{ip}} \frac{N(0|0, \gamma_s^{-1})}{N(0|\mu_{s_{ip}}^*, \Sigma_{s_{ip}}^*)} \tag{17}$$

2. sample q_{ip}:

$$p(q_{ip}|-) \sim Beta(m_0 + 1_{s\neq 0}, n_0 + 1_{s=0}) \tag{18}$$

3. sample γ_s:

$$
\begin{aligned}
p(\gamma_s|-) &\sim \Gamma(\gamma_s|e_0, f_0)\prod_{i=1}^{N} N(s_i|0, \gamma_s^{-1}I_P) \\
&\sim \Gamma(e_0 + \frac{1}{2}PN, f_0 + \frac{1}{2}\sum_{i=1}^{N} s_i^T s_i)
\end{aligned}
\tag{19}
$$

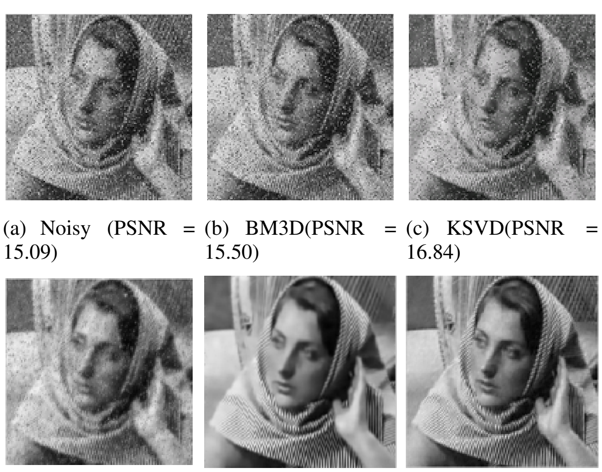
C. sample Gaussian noise

$$
\begin{aligned}
p(\gamma_n|-) &\sim \Gamma(\gamma_n|g_0, h_0)\prod_{i=1}^{N} N(y_i|D\alpha_i, \gamma_n^{-1}I_P) \\
&\sim \Gamma(g_0 + \frac{1}{2}PN, h_0 + \frac{1}{2}\sum_{i=1}^{N} \|y_i - D\alpha_i - s_i\|^2)
\end{aligned}
\tag{20}
$$

IV. EXPERIMENTS RESULTS

A. Mixed Noise Removal

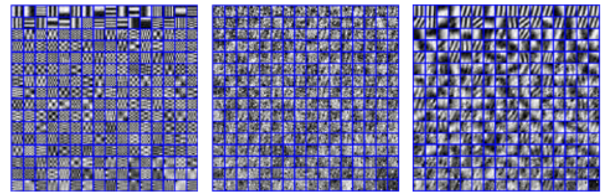
In order to evaluate the performance of our proposed model, we add both Gaussian and salt-and-pepper noise with different levels into three test images: Lena, House, Barbara. The resolution is 256×256, and patch size of 8×8. The dictionary dimension is 64 and the number of dictionary atoms is set to 256. Gamma parameters of c_0, d_0, e_0, f_0, g_0 and h_0 are all set to 10^{-6}, and these beta parameters are fixed as $a_0 = 1$, $b_0 = N/8$, $m_0 = 0.1$ and $n_0 = 0.9$. In our experiments, we compare the proposed model with the commonly-used median filter and three popular sparse representation algorithms of BPFA [1], KSVD [13], BM3D [6]and WKSVD [2]under the standard measure of Peak Signal to Noise Ratio (PSNR).



(a) Noisy (PSNR = 15.09) (b) BM3D(PSNR = 15.50) (c) KSVD(PSNR = 16.84)

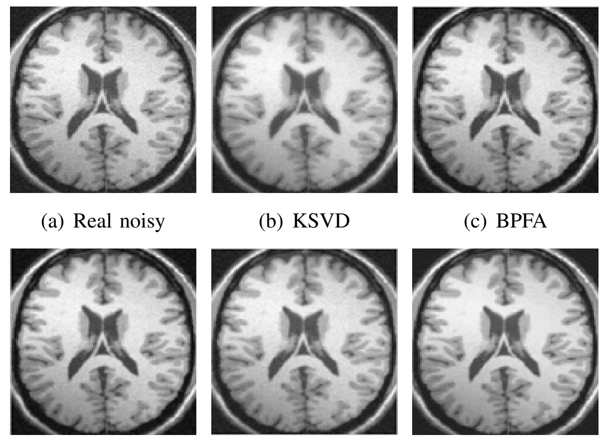
(d) BPFA(PSNR = 23.22) (e) WKSVD (PSNR = 29.94) (f) Ours(PSNR = 32.39)

Fig. 1. Comparison of different denoising results under the mixed noise (Gaussian variance 10 and 10% salt-and-pepper).



(a) KSVD (b) BPFA (c) Ours

Fig. 2. Dictionary learned from Barbara under the mixed noise (Gaussian variance 20 and 20% salt-and-pepper).



(a) Real noisy (b) KSVD (c) BPFA

(d) BM3D (e) WKSVD (f) Ours

Fig. 3. Comparison of different denoising results for real brain MR image.

Figure 1 shows denoising results for Barbara image corrupted by mixture noise with Gaussian (variance 10) and 10 percentage salt-and-pepper noise. As shown in Figure 1, BM3D and KSVD have the worse performance. Because KSVD needs to set the noise variance in advance, but the real variance is hard to be accurately estimated. BM3D and BPFA effectively work on the removal of Gaussian noise but has little effects on the salt-and-pepper noise. WKSVD is effectively to remove the mixed noise but produces mistake textures in the scarf. However, Our proposed method can outstandingly remove the mixed noise and achieve better visual quality than other methods. Figure 2 shows the dictionaries of different sparse representation-based methods with KSVD, BPFA, BM3D and the proposed method. It can be further seen that our model has better detail structures information than that of other two methods. Table 1 shows the denoising performance of different algorithms on the different levels of mixed noise, where σ_s is Gaussian noise variance and s is the density of the salt-and-pepper noise. Note that our proposed method can mostly achieve the highest PSNR on various noised images. It means that the proposed algorithm outperforms other methods. To deeply verify the efficiency of the proposed method, the denoising results of the real brain MR noisy image have shown in Figure 3. It can be drawn that our proposed method can achieve the better denoising effects than other four algorithms. And KSVD has the worst denoising effect due to the blur and contrast distortions arisen

TABLE I
PSNR FOR DIFFERENT DENOISING METHODS

Images	$\sigma_n + s$	Noisy	BM3D	KSVD	BPFA	WKSVD	Ours
House	10+10%	15.28	15.65	16.98	24.63	33.04	**34.07**
	10+20%	12.28	12.58	14.02	21.75	32.24	**32.47**
	10+30%	10.56	10.83	11.78	19.59	**30.70**	30.49
	20+10%	14.75	16.64	16.36	24.15	30.05	**31.69**
	20+20%	12.13	12.94	13.15	21.33	28.17	**30.43**
	20+30%	10.44	11.10	11.21	19.54	26.07	**28.30**
Barbara	10+10%	15.09	15.50	16.84	23.22	30.04	**32.39**
	10+20%	12.33	12.49	13.83	20.73	28.19	**30.86**
	10+30%	10.56	10.71	11.73	19.06	25.50	**27.80**
	20+10%	14.64	16.24	16.20	23.05	26.23	**29.25**
	20+20%	11.99	12.82	12.97	20.34	24.12	**27.74**
	20+30%	10.38	10.96	11.05	19.06	21.40	**25.62**
Lena	10+10%	15.18	15.65	17.04	23.50	**30.54**	29.41
	10+20%	12.25	12.53	13.99	21.04	**29.21**	29.14
	10+30%	10.59	10.75	11.79	19.22	27.21	**27.35**
	20+10%	14.60	16.49	16.42	23.40	27.97	**28.77**
	20+20%	12.13	12.87	13.11	20.92	25.94	**27.41**
	20+30%	10.48	11.00	11.16	19.07	23.97	**26.02**

in the denoised image, then BPFA and BM3D have the worse results due to some noise residual, and bad blurs have occurred in WKSVD. Therefore, our proposed method can perform better visual quality than other methods in the removal of the complicated and mixed noise.

B. Separation Removal for Gaussian and Salt-and-Pepper Noise

To further validation of the performance for our proposed model, we add Gaussian noise and salt-and-pepper noise into test images of Lena, House, Barbara respectively. And the denoising results for different algorithms on different noisy levels have shown in Table 2 and 3. Table 2 shows BM3D under the matched variance has the highest PSNR but is worse in the dismatched in lined with the actual practice, and our proposed method can perform much better PSNR than KSVD and BM3D with dismatched variance except 30. However, From the salt-and-pepper noise removal results of Table 3, it can be seen that the proposed method can always achieve the highest PSNR and outstandingly overwhelms other methods on the noised images. It means that our proposed method can outperform well with Gaussian and salt-and-pepper noise respectively, BM3D only fits well for Gaussian noise removal but fails in salt-and-pepper, and median filter(MF)is only suitable for removing salt-and-pepper noise with helpless for Gaussian noise.WKSVD and the proposed method both work well for the Gaussian noise, but our method can achieve better denoising results than WKSVD for the mixed noise from Table 1.

TABLE III
PSNR FOR SALT-AND-PEPPER NOISE REMOVAL

Images	s	BM3D	KSVD	BPFA	MF	Ours
House	10%	15.44	15.89	21.36	30.93	**36.27**
	20%	12.46	12.78	18.58	25.88	**27.17**
	30%	10.73	10.89	16.96	21.05	**22.48**
Barbara	10%	15.34	15.78	20.72	25.85	**36.32**
	20%	12.38	12.67	18.48	24.11	**25.74**
	30%	10.62	10.91	16.71	21.24	**21.31**
Lena	10%	15.47	15.95	21.13	28.34	**30.11**
	20%	12.42	12.75	18.52	25.36	**25.98**
	30%	10.66	10.84	16.74	21.40	**22.24**

V. DISCUSSION AND CONCLUSION

In this paper, we have presented a novel non-parametric Bayesian framework to solve sparse outlier problem, and also have considered the observed data as three components of an ideal data term, a Gaussian noise term and an outlier noise term. The spike-slab sparse prior has been employed to fit the sparse coefficients and outlier noise. And the proposed nonparametric Bayesian model can automatically estimate Gaussian noise variance with no need of changing our parameters under different levels for the mixed noise. Simulation

TABLE II
PSNR FOR GAUSSIAN NOISE REMOVAL

Images	σ_n	BM3D(matched/dismatched)	KSVD(matched/dismathed)	BPFA	WKSVD	Ours
House	10	**36.72**/32.58	34.96/32.56	35.30	35.34	35.29
	20	**33.79**/32.30	32.12/32.02	32.86	32.93	32.80
	30	**32.13**/31.61	30.12/30.06	30.92	30.75	30.87
Barbara	10	**34.78**/31.10	33.28/30.48	33.49	33.28	33.43
	20	**31.25**/29.89	30.01/29.28	29.02	24.80	30.16
	30	**29.06**/28.78	28.13/28.01	28.08	26.87	28.02
Lena	10	**33.87**/31.46	30.51/31.30	31.70	34.06	31.62
	20	**30.44**/30.35	29.47/29.24	30.33	26.37	29.50
	30	**29.29**/29.04	27.68/27.17	27.90	28.14	27.79

results demonstrate our model can outperform well with mixed noise than other existing methods. Due to the space limitations, we briefly note that our proposed algorithm can also be used for target detection, background and foreground separation in videos like the outlier in training dataset. And the work for these applications will be left for further study.

Acknowledgment

The project is supported by the National Natural Science Foundation of China (No. 30900328, 61172179, 61103121, 81301278), the Natural Science Foundation of Fujian Province of China (No. 2012J05160), The National Key Technology R&D Program (2012BAI07B06), the Fundamental Research Funds for the Central Universities (No. 2011121051, 2013121023), the NCETFJ.

References

[1] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, and L. Carin, "Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, January 2011.

[2] J. Liu, X. C. Tai, H. Huang, and Z. Huan, "A weighted dictionary learning model for denoising images corrupted by mixed noise," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1108–1120, March 2013.

[3] X. Liu and L. Huang, "A new nonlocal total variation regularization algorithms for image denoising," *Mathematics and Computers in Simulation.*, vol. 97, pp. 224–233, March 2014.

[4] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *International Journal of Computer Vision.*, vol. 86, no. 1, pp. 1–32, January 2010.

[5] G. Katkovnik and S. Osher, "Nonlocal operator with applications to image processing," *SIAM Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, November 2008.

[6] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing.*, vol. 54, no. 11, pp. 4311–4322, November 2006.

[7] K. Dabov, A. Foi, V. Katkovnik, and K Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, August 2007.

[8] B. Chen, J. Paisley, and L Carin, "Sparse linear regression with beta process priors.," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing*, March 2010, pp. 1234–1237.

[9] John Paisley and Lawrence Carin, "Nonparametric factor analysis with beta process priors.," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 777–784.

[10] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach.," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, February 2013.

[11] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering.," *2011 IEEE Conference on Computer Vision and Pattern Recognition.*, pp. 457–464, June 2011.

[12] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[13] R. H. Chan, C. W. Ho, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and edge-preserving regularization," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479–1485, October 2005.

[14] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, December 2011.

[15] K. Mitra, A. Veeraraghavan, and R. Chellappa, "Robust rvm regression using sparse outlier model," *2010 IEEE Conference on Computer Vision and Pattern Recognition.pages*, pp. 1887–1894, June 2010.

[16] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: frequentist and bayesian strategies," *Annals of Statistics*, vol. 33, no. 2, pp. 730–773, April 2005.

[17] G. Malsiner-Walli and H. Wagner, "Comparing spike and slab priors for bayesian variable selection," *Austrian Journal of Statistics*, vol. 40, no. 4, pp. 241–264, 2011.

[18] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Transactions on Signal Processing.*, vol. 57, no. 9, pp. 3488–3497, September 2009.