Image Denoising with Wavelet Markov Fields of Experts

Xiang Wu, Zhi-Guo Shi, and Lei Liu

Abstract—**Image denoising methods based on Markov random field (MRF) are often shown over-smooth phenomenon for strong noise image. Wavelet analysis has good time-frequency local ability and preserves the image edge information well for image denoising problem. Based on wavelet analysis and MRF theory, we propose a wavelet markov field of experts (WMFoE) framework to deal with image denoising problems. The noise image is divided into low-frequency and high-frequency component, and MRF are used to deal with low-frequency component. For high-frequency component, a clustering based soft-threshold method is used to remove the noise signal. Then, the restored image can be gotten by reconstruction from different components. Experiment results show that our method not only gets good PSNR and SSIM values but also preserves image edge information especially for strong noise image, compared with BM3D etc. state-of-art methods.**

Index Terms—**Markov random field, image denoising, wavelet transformation, k-means clustering.**

I. INTRODUCTION

Image restoration has been studied for decades in image processing. Traditionally, removing noise is achieved by linear filter processing such as Wiener filter, or Bayesian inference using image prior knowledge.

Since Donoho proposed the soft-thresholding function [1], lots of research on image denoising has been focused on methods on wavelet domains, because high-frequency components in wavelet domain contain image noise signal and edge for natural image. Therefore, in thresholding-based method, it is important to set a thresholding function for distinguishing noise signal from wavelet high-frequency coefficients. S. Grace Chang, Bin Yu and Martin Vetterli developed an adaptive data driven threshold for image denoising via wavelet soft-thresholding function [2]. They present the thresholding function derived from Bayesian framework and the wavelet coefficients can be modeled as a distribution such as Generalized Gaussian Distribution. This model are demonstrated to be efficient for a wide range level of Gaussian noise.

On the other hand, it is known for decades that image denoising can be expressed as energy minimization process of pixel-labelling task in Markov Random Fields (MRF). Richard Szeliski *et al.* do image denoising with smoothness-based priors in MRF [3], with fast MRF

inference solution method such as ICM, graph cut and belief propagation etc. They provide a set of energy minimization benchmarks and compare different energy minimization methods for image denoising problem. S. Roth developed a framework Fields of Experts (FoE) [4] to learn a generic expressive image prior knowledge that capture the statistic feature of natural scene images.

In this paper, we propose the Wavelet Markov Fields of Experts (WMFoE) algorithm to remove image noise. The WMFoE algorithm deals with coefficients separately in wavelet domain, by introducing virtues of wavelet analysis and MRF models on image denoising problem. The outline of this paper follows: In Section II, we describe the proposed WMFoE algorithm. The experimental results are given in Section III. The conclusions and directions for future work are listed in Section IV.

II. WAVELET MARKOV FIELDS OF EXPERTS

In this section, we will give the framework of the proposed Wavelet Markov Fields of Experts algorithm. The Flowchart of our method is shown in Fig. 1.

A. Framework of WMFoE Algorithm

The discrete wavelet transform kernel function can be represented as one separable 2-D scaling function and three separable 2-D wavelets

$$\phi(x, y) = \phi(x)\phi(y) \tag{1}$$

$$\psi^H(x, y) = \psi(x)\phi(y) \tag{2}$$

$$\psi^V(x, y) = \phi(x)\psi(y) \tag{3}$$

$$\psi^D(x, y) = \psi(x)\psi(y) \tag{4}$$

where $\phi(x)$ is the scaling function, $\psi(y)$ is the wavelets function, and $\psi^H(x, y)$, $\psi^V(x, y)$, and $\psi^D(x, y)$ are called horizontal, vertical and diagonal wavelets. The 2-D signal $f(x, y) \in L^2(R)$, wavelet transform formulation can be written as:

$$w_k = f_k + g_k^H + g_k^V + g_k^D \tag{5}$$

where f_k is low-frequency coefficients, g_k^H, g_k^V, and g_k^D are the horizontal, vertical and diagonal ones at scale k yields.

According to the wavelet transform, we can define the neighborhood factor S_i^2 as:

$$S_i^2 = \frac{1}{M} \sum_{w_j \in N_i} w_j^2 \tag{6}$$

where N_i is the set of sites neighboring i and M is the number of sites in the neighborhood.

Suppose the noisy image **y** and original image **x** can be formulated as

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \tag{7}$$
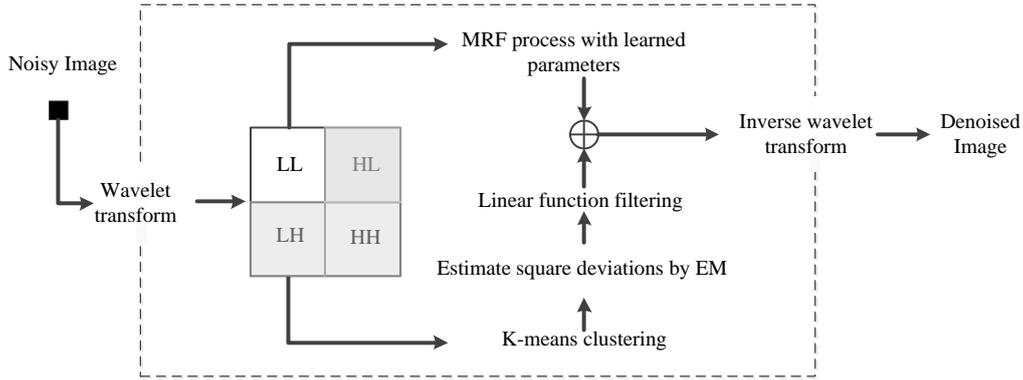
Fig. 1. Flowchart of the proposed image denoising algorithm.

where additive Gaussian white noise $v \sim N(0, \sigma^2)$. Our algorithm firstly starts with wavelet transform

$$w_y = w_x + w_v \tag{8}$$

where w_y is coefficients of noisy image, w_x, and w_v are the transformation of signal **x** and **v**. And then, we adopt differentmethod for low-frequency and high-frequency wavelet transform coefficients.

For purpose of removing the w_v from noisy image coefficients w_y **,** the threshold function method has been proposed and developed, however, the threshold function for high-frequency coefficients g_y have some disadvantages for recovery coefficient g_x. For example, hard-thresholding can produce ringing artifact and the soft-thresholding would lead to edges over-smoothed. Therefore, in this paper, we apply a simple but efficient linear method instead of thresholding function which is expressed as

$$\hat{g}_x(i,j) = \beta_{ij} g_y(i,j) \tag{9}$$

where the shrinkage factor β_{ij} can be defined as

$$\beta_{ij} = \max(1 - \sqrt{2\sigma^2 \log n} \big/ S_{ij}^2, 0) \tag{10}$$

σ is the square deviation of sample signal, the parameter n is the sample length of signal similar to Donoho thresholding function and S_{ij}^2 can be computed by equation (6). Besides, the factor formula means it takes non negative value.

On the other hand, we construct Markov Random Field for low frequent coefficients to remove the noise signal in low-frequency component. For our analysis of MRF in low-level vision, the probability density function is written as

$$p(\mathbf{x}; \Theta) = \frac{1}{Z(\Theta)} e^{-E_{\mathrm{WMFoE}}(\mathbf{x}; \Theta)} \tag{11}$$

with the energy function

$$E_{\mathrm{WMFoE}}(\mathbf{x}; \Theta) = \sum_{x \in N_i} \sum_{j=1}^{M} \log \psi_j(W_j^T \mathbf{x}; \alpha_j) \tag{12}$$

where W_j and α_j are the parameters of WMFoE model, M is the number of filters W_j. The function $\psi(\mathbf{x})$ has the form over FoEs model with Student-t experts [4]

$$\psi_j(W_j^T \mathbf{x}; \alpha_j) = (1 + \frac{1}{2}(W_j^T \mathbf{x})^2)^{-\alpha_j} \tag{13}$$

For Markov labeling problems, the maximum a posterior (MAP) solution, which is based on Bayes statistics theory in estimation and decision-making, is equivalently found by minimizing the posterior energy function [5]. Thus, we perform the posterior energy function of low-frequency coefficients $f_{\mathbf{x}}$ as

$$E(f_x \mid f_y) = E(f_y \mid f_x) + E_{\mathrm{WMFoE}}(f_x) \tag{14}$$

where the prior energy function as unary term

$$E(f_y \mid f_x) = \frac{1}{2\sigma^2}(f_y - f_x)^2 \tag{15}$$

and WMFoE energy function as the pairwise term. We use the gradient descent method to minimize the energy by

$$f_x^{(t+1)} = f_x^{(t)} - \alpha \nabla E(f_x \mid f_y) \tag{16}$$

where the directional derivative ∇E is expressed as

$$\nabla_{f_x} E = \frac{1}{\sigma^2} |f_y - f_x| + \sum_{j=1}^{M} \frac{\partial}{\partial f_x} \psi_j(W_j^T \mathbf{x}; \alpha_j) \tag{17}$$

After denoising on low-frequency and high-frequency components, we use inverse wavelet transform for wavelet reconstruction to get denoised image $\hat{\mathbf{x}}$.

B. Wavelet Coefficients Estimation

For the high-frequency coefficients, we use equation (9) to update the wavelet coefficients, so we need to compute the square deviations. Due to statistical features of image, we perform distribution fitting which is realized by Expectation Maximization (EM) algorithm to compute square deviations σ^2 of horizontal, vertical and diagonal coefficients respectively [6]. In fact, it is not appropriate to model all the wavelet coefficients in one subband with only one random variable for noisy image. Gathering pixel coefficients with some similarities is a good way for classification. Therefore, we define the features U_i of the site i of wavelet coefficient's four or eight nearest neighbours for clustering method, such as k-means, to estimate the square deviations locally and exactly.

C. MRF Parameters Learning

To better apply our method for different images, for low-frequency coefficients, it is essential to get suitable parameters for our WMFoE model. Generally, Maximum Likelihood Estimation (MLE) is used to get the suitable parameters on training sets. But according to Hinton's work, MLE has disadvantages for computing the parameters W and α [7]. So, for the low-frequency components of wavelet coefficients, we rely on the wavelet Markov fields of experts and train the filters W_j and weights α_j by contrastive divergence [7]. Contrastive divergence is an efficient method to get maximum likelihood through minimizing the Kullback-Liebler divergence between the data distribution and equilibrium distribution over the visible variables written as

$$\Delta\Theta = \eta \left[\left\langle \frac{\partial E_{\text{WMFoE}}}{\partial\Theta} \right\rangle_P - \left\langle \frac{\partial E_{\text{WMFoE}}}{\partial\Theta} \right\rangle_X \right] \quad (18)$$

where $\langle\ \rangle_P$ denotes the model distribution that we proposed as equation (11), $\langle\ \rangle_x$ denotes the average distribution of training data X, and η is the learning rate.

Due to the large number of natural image datasets for training, it is important to use subsampling strategy to estimate the parameters of our WMFoE model. Instead of inefficient direct sampling, Markov Chain Monte Carlo (MCMC) method is used. Running MCMC and Gibbs sampler for iterations starting from the training datasets will get the samples close to the target distribution. Then we perform Stochastic Gradient Descent (SGD) optimization method to update the parameters of WMFoE model.

III. EXPERIMENTS

In this section, we will evaluate and compare our algorithm with other methods on image denoising.

A. Wavelet Selection and Parameter Learning

The image denoising performance of our model is also based on the selection of wavelet function. A good wavelet transform function should not only make the high-frequency coefficients have histogram close to the distribution that we select, but also can concentrate noise signal on high-frequency. We select Discrete Meyer (D-Meyer) wavelets as the kernel function because of its high vanishing moments which can eliminate the correlation of pixels leading to get smaller high-frequency coefficients and concentrating on the energy of image.

Besides, learning suitable parameters determine the efficiency and universality of the proposed model. We train the datasets from Berkeley Segmentation Database [8] to get the suitable parameters of our WMFoE model. We employ contrastive divergence and MCMC method with SDG for learning filters W_j and weights α_j. We train 5×5 cliques' 24 filters of WMFoE model (shown in Fig. 2) on 200 natural images taken from Berkeley Segmentation Database. The properties of our low-frequency coefficients' MRF model are obtained by MCMC sampling images of 50×50 image patches after wavelet transform, and then the sampling data is

used by SGD with learning rate $\eta=0.1$. Training the WMFoE model is time-consuming, but occurs offline.

B. Image Denoising

We present experiments conducted with our proposed algorithm WMFoE in this section. The evaluation of our algorithm performance depends on two measurements:

1) The Peak Signal-to-Noise Ratio (PSNR) defined as

$$\text{PSNR}=10\lg\frac{255^2}{\sum(\text{x}-\hat{\text{x}})^2} \quad (19)$$

The PSNR is a very widely used evaluation criterion for denoising but has some limitation that it does not fully reflect the quality of restored image.

2) Structural similarity index (SSIM) [9] provides a more efficient method to evaluate the results of image denoising. The SSIM values range between 0 and 1, where the values which is closed to 1 means a perfect restoration method.

According to the Section III-A, we decide to employ Discrete Meyer wavelets as the wavelet function and 5x5 cliques' 24 filters of WMFoE model that we train as the low-frequency component model parameters. For the high-frequency component, we initialize the number of clusters k =35, the 3×3 region to compute the neighborhood factor in equation (6) and we use k-means to cluster pixels' feature which is composed by four nearest neighbors plus itself wavelet coefficients. To estimate the distribution of high-frequency coefficients, we employ the Laplacian distribution model.
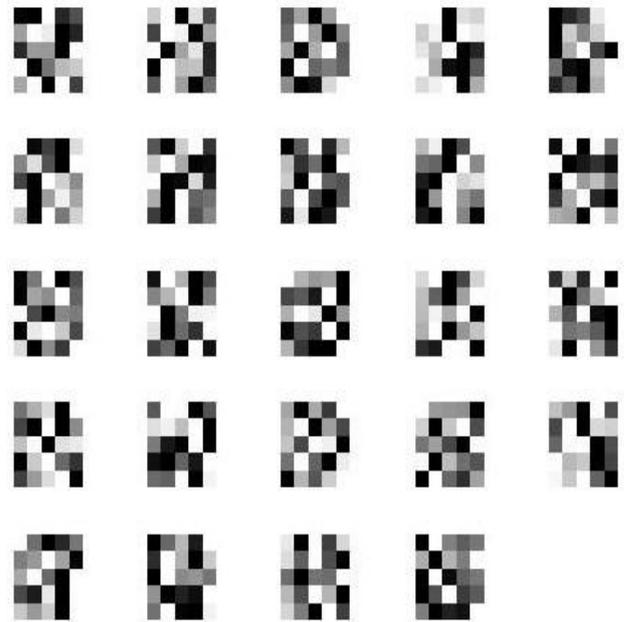
Fig. 2. 5×5 cliques' 24 filters of WMFoE model learned by contrastive divergence on Berkeley segmentation database.

Table I and Table II provides PSNR and SSIM values of the test images with different Additive White Gaussian Noise (AWGN) and we find our WMFoE algorithm works well for test images. Portilla *et al.* [10] and Roth S *et al.* [4] present denoising result on this dataset. We find that our denoising results are close to theirs (mostly within 0.3dB) and in some cases even outperform theirs (by about 0.5dB). Besides,

according to experimental results on SSIM value, the denoising performance is good especially when the image has lots of textures.

TABLE I: OUR WMFoE ALGORITHM PERFORMANCE WITH PSNR

σ^2/PSNR	Barbara	Boat	House	Lena	Peppers
10/28.13	33.58	32.35	35.58	35.30	34.07
15/24.61	32.04	31.17	34.31	33.65	32.72
20/22.11	30.96	30.13	33.41	32.37	31.66
25/20.17	30.01	29.21	32.60	31.33	30.68
50/14.15	26.97	26.40	29.53	27.97	27.50
75/10.63	24.42	24.62	26.29	25.32	24.67
100/8.13	22.18	22.82	23.53	22.85	22.22

TABLE II: OUR WMFoE ALGORITHM PERFORMANCE WITH SSIM

σ^2/PSNR	Barbara	Boat	House	Lena	Peppers
10/28.13	0.926	0.909	0.893	0.943	0.921
15/24.61	0.897	0.885	0.871	0.922	0.902
20/22.11	0.873	0.858	0.856	0.901	0.884
25/20.17	0.848	0.826	0.841	0.880	0.866
50/14.15	0.743	0.710	0.769	0.787	0.778
75/10.63	0.629	0.643	0.677	0.689	0.647
100/8.13	0.529	0.565	0.576	0.570	0.539

To prove the performance of our WMFoE method further, we use BM3D [11], LMMSE [12], GHP [13] and FoE [4] for comparison. The codes of all the method above for comparison is provided by the authors on the Internet and we use the authors' parameters setting. We employ all the denoising method on Lena, Barbara, House, and Pepper images with square deviation σ^2=10, 15, 20, 25, 50 AWGN and the results are shown in Fig. 3 and Fig. 4.

Table III and Table IV show the PSNR and SSIM values for different denoising method. We can see that our WMFoE algorithm is better than LMMSE and FoE method.

Besides, our method also gets similar performance compared with BM3D and GHP on PSNR and SSIM measure. The experiments shows the different performance between WMFoE and other methods are more notable to hold the detail information while others are over-smoothed.

In terms of PSNR value, the WMFoE performed slightly worse than the state-of-the-art BM3D for one of some noise levels, but the performance difference was not statistically significant in these cases. Overall, in majority of the cases the WMFoE denoising method performed good qualities and held more detail information than other methods.

TABLE III: IMAGE DENOISING PERFORMANCE COMPARISON WITH PSNR

σ^2	10	15	20	25	50	10	15	20	25	50
	Lena					Barbara				
BM3D	35.83	34.21	33.03	32.08	29.08	34.87	33.08	31.77	30.75	27.51
LMMSE	34.64	32.19	30.63	29.51	26.54	32.83	30.37	29.09	28.33	25.85
GHP	35.23	33.99	32.33	31.33	27.06	34.46	33.40	31.81	31.04	27.11
FoE	35.04	33.27	31.92	30.82	26.49	32.83	30.22	28.32	27.04	23.15
WMFoE	**35.30**	**33.65**	**32.37**	**31.33**	**27.97**	**33.58**	**32.04**	**30.96**	**30.01**	**26.97**
	House					Pepper				
BM3D	36.37	34.75	33.54	32.67	29.65	34.38	32.31	30.87	29.80	26.46
LMMSE	35.28	33.32	31.97	31.19	28.75	34.38	31.89	30.36	29.23	26.26
GHP	36.67	35.46	33.40	32.60	27.63	32.40	31.76	29.97	29.48	26.57
FoE	35.06	33.48	32.17	31.11	26.74	34.28	32.03	30.58	29.20	24.52
WMFoE	**35.58**	**34.31**	**33.41**	**32.60**	**29.53**	**34.07**	**32.72**	**31.66**	**30.68**	**27.50**

TABLE IV: IMAGE DENOISING PERFORMANCE COMPARISON WITH SSIM

| σ^2	10	15	20	25	50	10	15	20	25	50
	Lena					Barbara				
BM3D	0.953	0.931	0.907	0.887	0.808	0.951	0.922	0.892	0.867	0.769
LMMSE	0.931	0.899	0.871	0.846	0.752	0.904	0.857	0.826	0.802	0.703
GHP	0.952	0.934	0.895	0.879	0.708	0.946	0.932	0.910	0.893	0.778
FoE	0.898	0.876	0.854	0.834	0.741	0.918	0.884	0.841	0.805	0.622
WMFoE	**0.943**	**0.922**	**0.901**	**0.880**	**0.787**	**0.926**	**0.897**	**0.873**	**0.848**	**0.743**
	House					Pepper				
BM3D	0.923	0.890	0.871	0.861	0.826	0.933	0.900	0.889	0.853	0.759
LMMSE	0.880	0.861	0.846	0.835	0.781	0.921	0.895	0.872	0.850	0.772
GHP	0.934	0.903	0.854	0.825	0.604	0.931	0.911	0.883	0.856	0.712
FoE	0.880	0.866	0.850	0.836	0.763	0.932	0.901	0.879	0.853	0.735
WMFoE	**0.893**	**0.871**	**0.856**	**0.841**	**0.769**	**0.921**	**0.902**	**0.884**	**0.866**	**0.778**
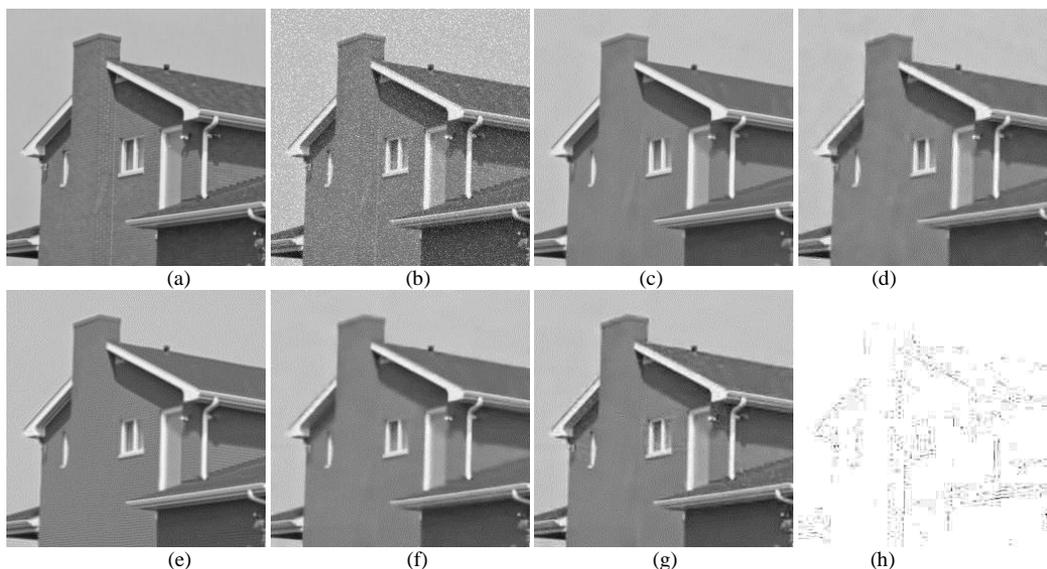
Fig. 3. House denoising results. (a) Original image. (b) Noisy image with AWGN (σ^2 =25). (c) BM3D; PSNR=32.67dB. (d) D LMMSE; PSNR =31.19dB. (e) GHP; PSNR=32.60dB. (f) FoE; PSNR =31.11dB. (h) WMFoE; PSNR=32.60dB. (h) Difference between original image and our denoising method output.
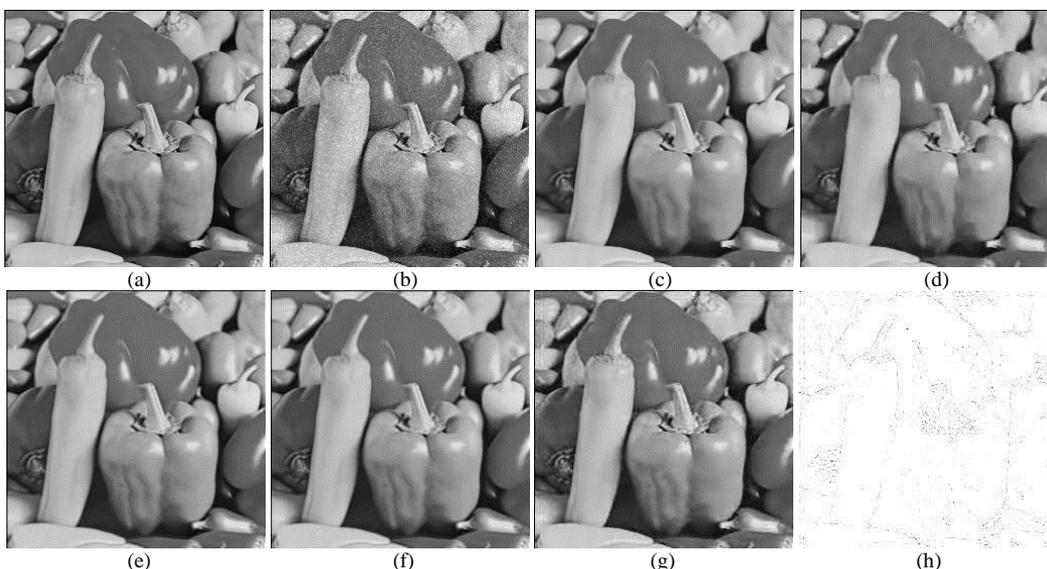

Fig. 4. Peppers denoising results. (a) Original image. (b) Noisy image with AWGN (σ^2 =25). (c) BM3D; PSNR=29.80dB. (d) LMMSE; PSNR =29.23dB. (e) GHP; PSNR=29.48dB. (f) FoE; PSNR =29.20dB. (h) WMFoE; PSNR=30.68dB. (h) Difference between original image and our denoising method output.

IV. CONCLUSION

In this paper, we proposed a Wavelet Markov Fields of Experts (WMFoE) model to solve the image denoising problems. The experimental results demonstrate that our method has similar PSNR or SSIM values compared with BM3D, LMMSE, GHP, FoE and so on. Due to dealing with noisy image by dividing wavelet coefficients into low and high-frequency components, we could not only get good quality of image denoising results, but also make the algorithm faster by parallel computation. While our WMFoE hold the detail edge information while other methods show over-smoothed phenomenon across different image region.

Although WMFoE algorithm has a good quality of image restoration, the scheme has some parameters to initialize and those would affect the results of denoising. Yet, a few images with weak noise do not get a good PSNR or SSIM values. This situation might be caused by high-frequency coefficients processing, because those components contain both noise signal and edge information of image. In the future, we will extend WMFoE algorithm for increasing the quality of weak noisy image and other image denoising tasks.

REFERENCES

[1] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.
[2] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol.9, no.9, pp. 1532-1546, 2000.
[3] R. Szeliski, R. Zabih, D. Scharstein *et al.*, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068-1080, 2008.
[4] S. Roth and M. J. Black. "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205-229, 2009.
[5] S. Z. Li, *Markov random field modeling in image analysis*, Springer, 2009.
[6] A. Hyvàrinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics*, Springer, 2009.
[7] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol.14, no. 8, pp. 1771-1800, 2002.

[8] D. Martin *et al.*, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE. ICCV.* vol. 2, pp. 416-423, 2001.

[9] Z. Wang *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[10] J. Portilla *et al.*, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338-1351, 2003.

[11] K. Dabov *et al.*, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080-2095, 2007.

[12] L. Zhang, P. Bao, and X. L. Wu, "Multiscale LMMSE-based image denoising with optimal wavelet selection," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 4, pp. 469-481, 2005.

[13] W. Zuo, L. Zhang, C. Song *et al.*, "Texture enhanced image denoising via gradient histogram preservation," in *Proc. IEEE. CVPR*, pp. 1203-1210, 2013.

Xiang Wu received his B.S degree from the School of Computer and Communication Engineering, University of Science and Technology Beijing, China, in 2013. He is a master student in the School of Computer and Communication Engineering, University of Science and Technology Beijing. His current research interests include machine learning and image processing.

Zhi-Guo Shi received his B.S. and M.S. degrees from the School of Information Engineering, University of Science and Technology, Beijing, China, in 1996 and 2000, respectively, and his Ph.D. degree from Institute of software Chinese Academy of Sciences, China, in 2008. He engaged in a visiting scholar at Ryerson University in Canada from April 2013 to April 2014. He is an associate professor at the faculty of School of Computer and Communication Engineering, University of Science and Technology Beijing. His current research interests include torque measurement and estimation of the robotic joint, multi-robot collaboration system and biomedical micro-robot.
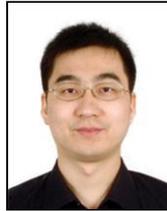
Lei Liu was born in 1980. He received his B.S. degree from the School of Mechanical Engineering, University of Science and Technology in 2003, and his Ph.D. degree from the School of Automation and Electrical Engineering, University of Science and Technology in 2011. Currently, he is a lecturer at the faculty of School of Computer and Communication Engineering, University of Science and Technology Beijing. He was a visiting scholar at School of Electrical and Computer Engineering at Georgia Tech. from February 2009 to February 2011. His current research interests include image processing, computer vision, pattern recognition, machine learning.