Learning Texture Image Prior for Super Resolution Using Restricted Boltzmann Machine

Chulmoo Kang(✉), Minui Hong, and Suk I. Yoo

Department of Computer Science and Engineering,
Seoul National University, Seoul, Republic of Korea
{nkarma,alsdml123,sukinyoo}@snu.ac.kr

Abstract. Field of Expert (FoE) [1], which is one of the most popular probabilistic models of natural image prior, has been successfully applied to super resolution. Piecewise smoothness imposed on natural images is, however, a relatively limited model for texture image. In the field of deep learning, various approaches for texture modeling using the Restricted Boltzmann Machine (RBM) achieves or surpasses the state-of-the-art on many tasks such as texture synthesis and inpainting. In this paper, we apply the convolutional RBM (cRBM) to learning a texture prior. The maximum a posteriori (MAP) framework is proposed to utilize the probabilistic texture model well. The experiment is done on the Brodatz Dataset, and our experimental results are shown to be comparable to those using FoE and other super resolution approaches.

Keywords: Texture image prior · Restricted boltzmann machine · Super resolution · Deep learning

1 Introduction

Super Resolution is a very active research topic in the image processing community. Formally speaking, super resolution(SR) estimates a high resolution(HR) image from one or multiple low resolution(LR) images. The problem is inherently under-determined because there are many possible high resolution images given a low resolution image.

There are several means of addressing the SR problem, such as interpolation-based super resolution, reconstruction-based super resolution and learning-based super resolution. Interpolation-based super resolution attempts to interpolate the HR image from the LR input [11]. These approaches usually blur high frequency details, however. Reconstruction-based approaches estimate an SR image using multiple low resolution images or by means of patch redundancy [12,13]. Learning-based techniques estimate high frequency details from a large training set of HR images that encode the relationship between HR and LR images [14]. These approaches have shown great promise and have been applied to SR in various ways. Recently, methods which exploit natural image priors for SR have been proposed in the literature [2]. The FoE framework, which imposes piecewise

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 215–224, 2015.
DOI: 10.1007/978-3-319-23231-7_20

smoothness on images, outperforms other natural image priors such as sparse coding or patch redundancy. However, the piecewise smoothness in this case is not equal to that associated with texture images; hence, the FoE framework is not suitable for modeling texture images [3]. In the field of deep learning, texture image modeling has been an active topic [4]. Following the recent exploration of the Restricted Boltzmann Machine(RBM) for texture image modeling, we choose the convolutional Gaussian Restricted Boltzmann Machine (cGRBM) [5,6]. The performance of this method is comparable to those of other modeling technique; moreover, its energy function in this case is tractable [7]. This tractable energy function is used to calculate the maximum a posteriori estimation. We explain this in detail in the following sections.

Although the modeling performance of cGRBM is acceptable, it can model only one texture image. A means of natural expansion in this case is to develop the ability to model the variety of textures seen in natural scenes. One means of addressing this problem is to train multiple texture image models individually and then to select the proper model for the given image. To do this, the implicit mixture RBM (imRBM) is used.

In this paper, exploring the implicit mixture of cGRBM, we suggest the maximum a posteriori (MAP) framework for super resolution into which a texture image prior is embedded. We investigate the mechanism how the implicit mixture of cGRBM automatically selects the correct texture among multiple cGRBMs.

The remainder of our paper is organized as follows: First, Section 2 describes the process of learning the texture image prior using the RBM; Section 3 presents details of our suggested framework. The results are presented in Section 4, and Section 5 concludes our paper with a discussion.

2 Learning Texture Image Prior

Modeling texture image prior knowledge can be understood by considering a combination of several repetitive features. Instead of hand-tuned feature selection, deep learning automatically selects the feature that captures the texture image structures. Following recent research [5,6], the convolutional Gaussian RBM matches or exceeds achieves the results of state-of-the-art method.

2.1 Modeling Individual Texture with the RBM

Boltzmann machines (Fig. 1 (a)) consider two sets of variables, the hidden unit $\mathbf{h}$, and the visible unit $\mathbf{x}$. They model a joint distribution of random variables with Boltzmann distribution (also called Gibbs distribution)

$$\frac{p(\mathbf{x}, \mathbf{h}) = \exp(-E(\mathbf{x}, \mathbf{h}))}{Z} \tag{1}$$

where is the energy function according to the model, and Z is a normalization constant which is also known as a partition function.

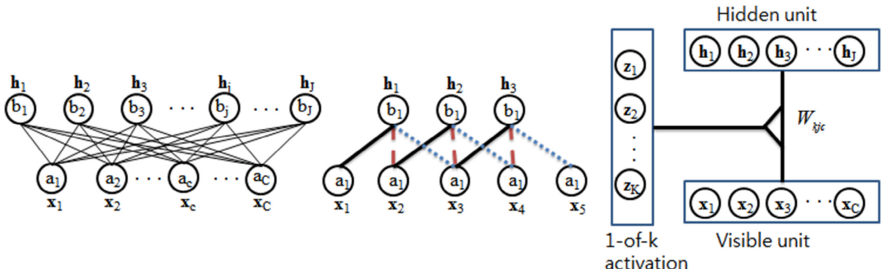


Fig. 1. (a) The undirected graph of an RBM with J hidden and C visible variables. (b) The convolutional RBM (identical color lines indeicate the same weights). (c) Schematic diagram of the implicit mixture of the RBM.

To resolve the poor scalability of the RBM, the convolutional Gaussian RBM (Fig. 1 (b)) is suggested. This model is spatially invariant and scalable to a realistic image size [8].

The energy function of cGRBM is written as

$$E_{cGRBM}(\mathbf{x}, \mathbf{h}) = \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} - \sum_{j \in filter} \sum_{c \in visible} h_{jc}(\mathbf{w}_j^T \mathbf{x} + b_j) \tag{2}$$

where the σ is standard deviation of Gaussian noise. Here, $\mathbf{w}_j$ determines the interaction between pairs of visible units $\mathbf{x}_c$ and hidden units h_{jc}. Thus, $\mathbf{w}_j$ values are the filters, b_j are the biases of hidden units, and c and j are indices for all overlapping image cliques and filters.

Although cGRBM facilitates texture model learning, it also restricts the number of texture models to one. Not being able to model a variety of textures, cGRBM is naturally extended to multiple Boltzmann machines.

2.2 Implicit Mixture Modeling for Multiple Textures

One way to address the problem of cGRBM modeling one texture per cGRBM is to use an implicit mixture of RBMs (imRBM) [9]. The literal meaning of implicit mixture is that it mixes one with another; however, in our work, the underlying role of imRBM is the selection of the proper cGRBM among several learned machines *automatically*. We explain this in detail in Section 3.

The energy function of the binary visible unit is

$$E(\mathbf{x}, \mathbf{h}, \mathbf{z}) = - \sum_{k \in \{1,2,\ldots,K\}} \sum_{j \in filter} \sum_{c \in visible} z_k h_{jc} \mathbf{w}_{jk}^T \mathbf{x}_c \tag{3}$$

where K is the number of RBMs (Fig. 1 (c)). Eq. (3) shows that imRBM is extended by including a discrete variable z with K possible states. In addition, $\mathbf{w}_{jk}$ is the filter which determines the interaction between h_{jc} and $\mathbf{x}_c$ of the kth

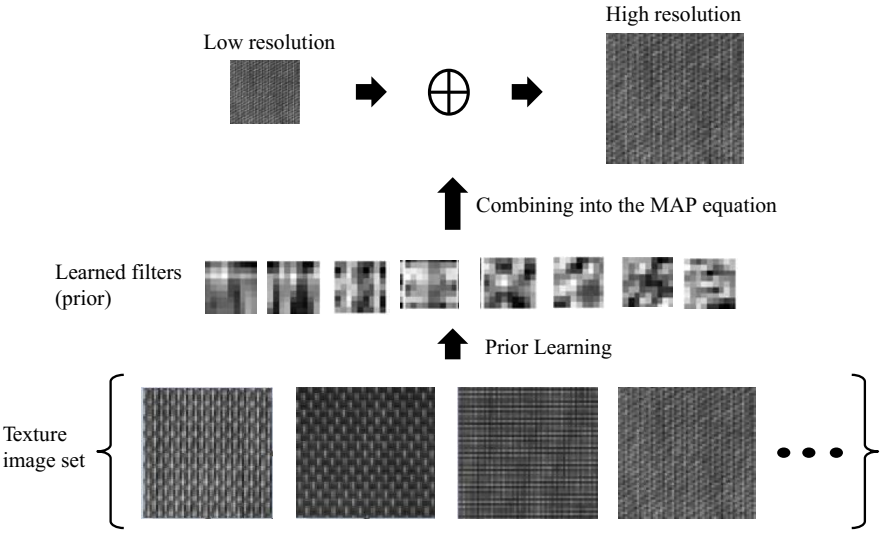


Fig. 2. A MAP framework for super resolution. In this proposed framework, we learn the filter as a prior. This prior is used to obtain the MAP solution analytically. The MAP solution is calculated using a gradient descent algorithm.

cGRBM. Eq. (3) can be adapted to a convolutional Gaussian visible unit [10]. Its energy function is given below.

$$\begin{aligned} E(\mathbf{x},\mathbf{h},\mathbf{z}) &= \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} - \sum_{k\in\{1,2,\ldots,K\}}\sum_{j\in filter}\sum_{c\in visible} z_k\left(h_{jc}\left(\mathbf{w}_{jk}^T\mathbf{x}_c + b_j\right)\right) \\ &= \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} - \sum_{k\in\{1,2,\ldots,K\}}\sum_{j\in filter}\sum_{c\in visible} h_{jc}\left(z_k\left(\mathbf{w}_{jk}^T\mathbf{x}_c + b_j\right)\right) \end{aligned} \tag{4}$$

3 A MAP Framework for Super Resolution

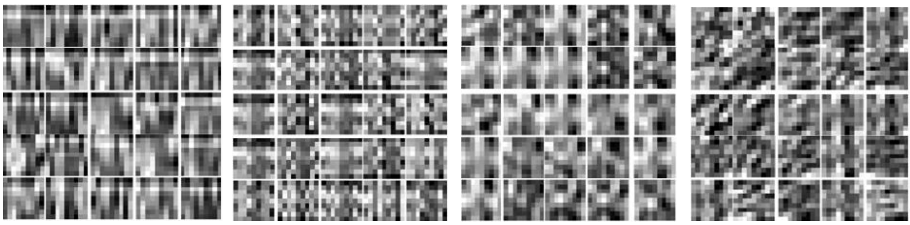
To utilize the deep learning MAP framework well, we formulate probabilistic models for a priori.

3.1 Image Formation Modeling

The low resolution image, $\mathbf{y}$, the image formation process is usually modeled as the convolution of the high resolution image, $\mathbf{x}$, followed by down sampling.

$$\mathbf{y} = \mathbf{DHx} + \mathbf{e} \tag{5}$$

Here, $\mathbf{D}$ is the downsampling operator, $\mathbf{H}$ is the blurring operator and $\mathbf{e}$ denotes the noise added to the low resolution image. The noise is usually modeled



(a) D6 (b) D21 (c) D53 (d) D77

Fig. 3. The learned filters

as independent and identically distributed Gaussian noise. The super resolution problem, recovering $\mathbf{x}$ from $\mathbf{y}$, is a well-known ill-posed problem. To solve the under-determined problem, it is possible to restrict the solution space by a prior knowledge.

The super resolution problem is then formulated via regularized least square regression

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{DHx}\|^2 + \lambda' R(\mathbf{x}) \tag{6}$$

where $R(\mathbf{x})$ is the mathematical formula for a priori and λ' is a regularization parameter. The first term here serves as the log likelihood function which models the Gaussian noise. Eq. (6) can be also recognized as MAP estimation problem. To clarify this statement, we start with a reformulation of the MAP estimation equation,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) \tag{7}$$

where $p(\mathbf{y}|\mathbf{x})$ denotes the likelihood distribution and $p(\mathbf{x})$ is the prior distribution. The likelihood distribution can be written as follows:

$$p(\mathbf{y}|\mathbf{x}) = \exp\left(-\frac{\|\mathbf{y} - \mathbf{DHx}\|^2}{\sigma^2}\right) \tag{8}$$

Here, σ denotes the standard deviation of the noise. The prior distribution of $\mathbf{x}$ can be represented by energy in the context of graphical models

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_E} \exp\left(-E(\mathbf{x})\right) \tag{9}$$

In this equation, Z_E is the partition function. Putting Eq. (8) and Eq. (9) into Eq. (7), and taking the negative logarithm, we obtain the regularized least square formula shown in Eq. (6). At this point, we are ready to derive the prior distribution for the texture model.

3.2 Embedding Multiple Texture Prior in the MAP Framework

To obtain the probabilistic prior model, $p(\mathbf{x})$, with which to capture the structure of the texture images, we integrate the energy function shown in Eq. (4)

out of the latent variables over their domains [7]. The free energy is proportional to marginal distribution [16]. The free energy is given below.

$$\begin{aligned}
F(\mathbf{x}) &= -\log\left(\sum_{\mathbf{z}}\sum_{\mathbf{h}}\exp\left(-E\left(\mathbf{x},\mathbf{h},\mathbf{z}\right)\right)\right) \\
&= -\log\left(\sum_{\mathbf{z}}\sum_{\mathbf{h}}\exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}+\sum_{k}\sum_{j,c}z_k\left(h_{jc}\left(\mathbf{w}_{jk}^T\mathbf{x}_c+b_j\right)\right)\right)\right) \\
&= -\log\left(\sum_{\mathbf{z}}\sum_{\mathbf{h}}\exp\left(-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}+\sum_{k}\sum_{j,c}h_{jc}\left(z_k\left(\mathbf{W}_{kjc}\mathbf{x}+b_j\right)\right)\right)\right) \\
&= \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}-\log\left(\sum_{\mathbf{z}}\sum_{\mathbf{h}}\exp\left(\sum_{k}\sum_{j,c}h_{jc}\left(z_k\left(\mathbf{W}_{kjc}\mathbf{x}+b_j\right)\right)\right)\right) \\
&= \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}-\log\left(\sum_{\mathbf{z}}\prod_{j,c}\left(1+\exp\left(\sum_{k}z_k\left(\mathbf{W}_{kjc}\mathbf{x}+b_j\right)\right)\right)\right) \\
&= \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}-\log\left(\sum_{k=1}^{K}\prod_{j,c}\left(1+\exp\left(\mathbf{W}_{kjc}\mathbf{x}+b_j\right)\right)\right)
\end{aligned} \tag{10}$$

Here, $\mathbf{W}_{kjc}\mathbf{x} = \mathbf{w}_{jk}^T\mathbf{x}_c$, $\mathbf{W}_{kjc}$ is the tensor of the filters. In addition, $\mathbf{x}$ is the vectorized image.

The super resolution MAP framework with probabilistic texture prior is derived from combining Eq. (6) and Eq. (10).

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}}\left(\left\|\mathbf{y}-\mathbf{DHx}\right\|^2+\lambda\left(\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}-\log\sum_{k=1}^{K}\prod_{j,c}\left(1+\exp(\mathbf{W}_{kjc}\mathbf{x}+b_j)\right)\right)\right) \tag{11}$$

where λ is a regularization parameter.

To get the MAP solution, the optimization is performed by gradient descent. The gradient of $F(\mathbf{x})$ w.r.t. $\mathbf{x}$ is shown in Eq. (12).

$$\nabla F_{\mathbf{x}} = \frac{\mathbf{x}}{\sigma^2}-\log\left(\frac{\sum_{k=1}^{K}\left(AR_k\times\left(\frac{\mathbf{W}_{kjc}\times\exp(\mathbf{W}_{kjc}\mathbf{x}+b_j)}{1+\exp(\mathbf{W}_{kjc}\mathbf{x}+b_j)}\right)\right)}{\sum_{k=1}^{K}AR_k}\right) \tag{12}$$

where $AR_k = \prod_{j,c}\left(1+\exp\left(\mathbf{W}_{kjc}\mathbf{x}+b_j\right)\right)$.

AR_k represents correlations between the filter of the kth RBM and the given image. Due to the product in AR_k, the correlation of each clique and filter is amplified. As a result, AR_k of the correct RBM is dominant such that the others become meaningless. This indicates that only the correct RBM which best

captures the structure of the given low resolution image affects the gradient of $F(\mathbf{x})$. This gives us a good insight into the 1-of-K activation mechanism of implicit mixture model. Note that one cannot calculate AR_k directly(Under our experimental conditions, it is over exp (10000000)). Because the maximum index of the AR_ks is that of the logarithm of the AR_ks, it is sufficient to determine the correct machine to sum the correlations of each clique and filter. After determining the correct machine by summing the correlations, Eq. (12) can be calcuated approximately with the filters and biases of the correct machine.

4 Experiment

The dataset used in the experiments and for the learning of the model is described in section 4.1. Section 4.2 shows the performance of the proposed algorithm as compared to that of a state-of-the-art method.

4.1 Dataset and Learning cRBM

We use a range of four texture images from the Brodatz dataset : D6, D21, D53, and D77(http://www.ux.uis.no/∼tranden/brodatz.html). We apply rescaling similar to that used in earlier work in [3], in which the 640x640 textures were rescaled to 480x480 (Downscale ratio is 0.75.). We divide each image into a top half used for training and a bottom half for testing. The number of training images is 40 with the size of 76x76, which is randomly cropped in the top half image. The filter size is set to 9x9. The number of filters per cRBM is 25. The number of RBMs in the imcGRBM is 4. We learn the parameters of the models by approximate maximum likelihood, using stochastic gradient ascent based on the Persistent chains Contrastive Divergence method(PCD) [18]. The number of Markov chain transition was set to 1 (PCD-1), and the transition was done by Hybrid Monte Carlo (HMC) sampling with 30 sample steps and 20 Leapfrog steps [19]. The learning rate of the weights is 0.000005. Following the recent research [6], σ^2 is set to 0.03. Figure 3 shows the learned filters for several texture images.

To create a low resolution, we convolute the ground truth with a blur kernel for which support is 7 and sigma is 1. Down sampling operator, with a scale of 2, follows the blurred image.

4.2 Multiple Texture Super Resolution

We conduct several experiments for super resolution with a size of 60x60 for the low resolution images and a zoom factor of 2. Figure 4 shows the super resolution result compared to those of other approaches. Our result shows more visually pleasing images as compared to those of the other methods. The fifth row of Fig. 4 shows the comparison with super resolution using the natural image prior, FoE. The result with the proposed method shows a sharper image compared to the use of FoE. This may result from the piecewise smoothness of FoE, in contrast

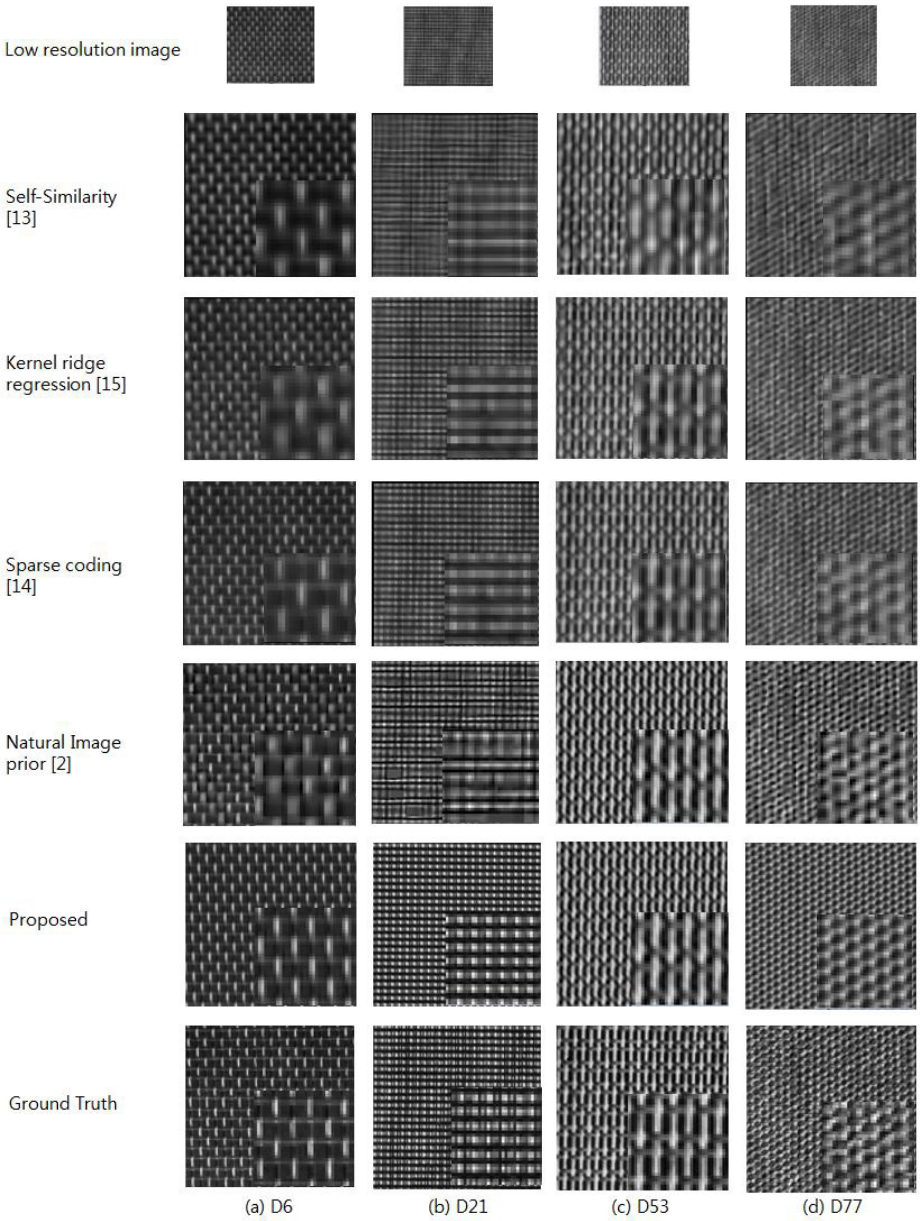


Fig. 4. Super resolution(X2) results of various approaches: Low resolution, Self-similarity, Kernel Ridge Regression, Sparse Coding, Natural Image Prior (FoE), the proposed algorithm and Ground truth (From the top row to the bottom row).

to our algorithm, which learns the image structure directly by finding the filter which is best correlated with the given low resolution image.

In addition to a qualitative comparison, we compare the result quantitatively. The well-known measure for super resolution is PSNR, which is defined as

$$PSNR = 10 \log_{10} \frac{(MaximumIntensityValue)^2}{MeanSquareError} \tag{13}$$

Another measure is the Structural Similarity Index (SSIM) [17]. As shown in Table 1, the proposed method can achieve better results than any other method such as the Self-similarity [13], Kernel Ridge Regression [15], Sparse Coding [14] and FoE [2] methods.

Table 1. Super Resolution (X2) Quality

Method		Self Similarity [13]	KRR [15]	Sparse Coding [14]	FoE [2]	Proposed
D6	PSNR	15.764	19.897	20.225	22.104	**25.201**
	SSIM	0.232	0.645	0.692	0.800	**0.913**
D21	PSNR	13.992	15.365	15.845	16.745	**22.074**
	SSIM	0.277	0.486	0.555	0.707	**0.937**
D53	PSNR	13.733	16.560	16.825	22.484	**23.136**
	SSIM	0.468	0.706	0.723	0.940	**0.955**
D77	PSNR	14.930	17.218	17.244	21.163	**21.452**
	SSIM	0.212	0.530	0.536	0.855	**0.882**

5 Conclusion

We formulate the probabilistic model for texture image prior by exploring the cGRBM. We propose a MAP framework using the texture prior learned from cGRBM for super resolution. We get the MAP solution by analytic formulation. Experimental results show that the result of the proposed algorithm is comparable with other approaches in terms of quantitative and qualitative comparison.

References

1. Schmidt, U., Gao, Q., Roth, S.: A Generative perspective on MRFs in low-level vision. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1751–1758 (2010)
2. Zhang, H., Zhang, Y., Li, H., Huang, T.: Generative Bayesian Image Super Resolution with Natural Image Prior. IEEE Transactions on Image Processing, 4054–4067 (2012)
3. Heess, N. Williams, C., Hinton, G.: Learning generative texture models with extended fields-of-experts. In: Proc. the British Machine Vision Conference (2009)
4. Luo, H. Carrier, P.L., Couville, A., Bengio, Y.: Texture modeling with convolutional spike-and-slab RBMs and deep extensions. In: Proc. the International Conference on Artificial Intelligence and Statistics (AISTAT), pp. 415–423 (2013)

5. Kivinen, J., Williams, C.: Multiple texture Boltzmann machines. In: Proc. the International Conference on Artificial Intelligence and Statistics (AISTAT), pp. 4054–4067 (2012)
6. Gao, Qi, Roth, Stefan: Texture synthesis: from convolutional RBMs to efficient deterministic algorithms. In: Fränti, Pasi, Brown, Gavin, Loog, Marco, Escolano, Francisco, Pelillo, Marcello (eds.) S+SSPR 2014. LNCS, vol. 8621, pp. 434–443. Springer, Heidelberg (2014)
7. Ranzato, M., Mnih, V., Susskind, J., Hinton, G.: Modeling Natural Images Using Gated MRFs. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2206–2226 (2013)
8. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proc. International Conference on Machine Learning, pp. 609–616 (2009)
9. Nair, V., Hinton, G.: Implicit mixtures of restricted Boltzmann machines. In: Proc. Advances in Neural Information Processing System (2009)
10. Welling, M., Rosen-Zvi, M., Hinton, G.: Exponential family harmoniums with an application to information retrieval. In: Proc. Advances in Neural Information Processing System (2008)
11. Takeda, H., Farisu, S., Milanfar, P.: Kernel Regression for Image Processing and Reconstruction. IEEE Transactions on Image Processing **12**(2), 349–336 (2007)
12. Glasner, D., Bagon, S., Irani, M.: Super-resoluton from a single image. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (2009)
13. Freeman, G., Fattal, R.: Image and Video Upscaling from Local Self-Examples. ACM Trans. Graph. **28**(3), 1–10 (2009)
14. Yang, J., Wang, Z., Lin, Z., Huang, T.: Coupled dictionary training for image super resolution. IEEE Transactions on Image Processing **21**(8), 3467–3478 (2012)
15. Kim, K.I., Kwon, Y.: Single-image Super-resolution Using Sparse Regression and Natural Image Prior. IEEE Transaction on Pattern Analysis and Machine Intelligence (2008)
16. Ngiam, J., Chen, Z., Koh, P.W., Ng, A.Y.: Learning deep energy models. In: Proc. International Conference on Machine Learning (2011)
17. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A New look at Signal Fidelity Measures. IEEE Signal Process. Mag. **26**(1), 98–117 (2009)
18. Tieleman, T.: Training restricted boltzmann machines using approximations to the likelihood gradient. In: Proc. International Conference on Machine Learning (2008)
19. Neal, R.M.: MCMC using Hamiltonian dynamics (2012). arXiv:1206.1901v1