IMAGE DENOISING BASED ON A STATISTICAL MODEL FOR WAVELET COEFFICIENTS

Hyung Il Koo and Nam Ik Cho

School of Electrical Engineering
Seoul National University, Seoul 151-744 Korea
email: hikoo@ispl.snu.ac.kr, nicho@snu.ac.kr

ABSTRACT

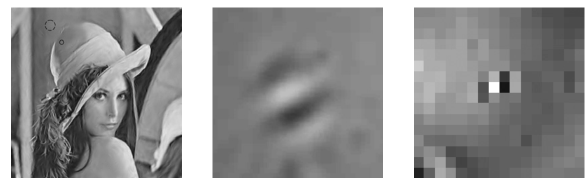
In this paper, we propose a new statistical model for the relationship of wavelet coefficients and its application to image denoising. The magnitude of a wavelet coefficient usually shows high correlations with the nearby ones. This property has been exploited in many wavelet-based image processing techniques. However, conventional works consider only the local neighborhood of a coefficient when inferring its hidden state. Consequently, the image context is not faithfully reflected and thus there are sometimes visually annoying artifacts. We attempt to alleviate this problem by developing a new statistical model for the random field that is consisted of hidden variables of the overall band and thus includes global relationship of wavelet coefficients. In this model, the image context is encoded by the relations of hidden states, and the state plane is efficiently inferred by the sum-product algorithm. In the experiment, the proposed model is incorporated with the state-of-the-art denoising algorithm, namely BLS-GSM (Bayes Least Square - Gaussian Scale Mixture). The results show that the proposed algorithm suppresses many annoying artifacts that exist in the conventional denoising methods, and thus improves the subjective quality.

***Index Terms*—** Conditional Random Fields, Bayesian Estimation, Image Denoising

1. INTRODUCTION

Wavelet transform is an efficient tool for many image processing applications such as image denoising, image compression, and image interpolation. Several properties of wavelet transform that are successfully used for these applications can be summarized as *Locality, Multiresolution, Compression, Clustering*, and *Persistence* [2]. Especially, *Clustering* and *Persistence* mean high magnitude correlation between the adjacent coefficients, which are the most important ingredients of compression and denoising algorithms.

Among many wavelet based image denoising algorithms, the BLS-GSM algorithm is known to be a state-of-the-art single image denoising algorithm [3], which shows excellent noise suppression performance while preserving discontinuities. However, the artifacts are sometimes still annoying as shown in Fig. 1. Like many other approaches, BLS-GSM modeled the wavelet coefficients with the assumption of hidden state [1, 2]. Then the probability of each state was inferred from the local observation of $9 \sim 10$ wavelet coefficients. The artifacts in Fig. 1 are usually caused by this "local observation," and thus we need to consider wider range of hidden random field. Actually, such consideration is widely used in the algorithms for stereo and segmentation problems, which are usually modeled with the local observation and also the relationship between the adjacent sites. Moreover, recently developed optimization techniques enable this inference of the complicated model to be tractable [4, 5, 6, 7]. To apply these techniques in the random fields of hidden state, we divide the hidden state so that every state has the same probability and model the hidden state random field based on the Conditional Random Field (CRF) model [8], which is a conditional version of the Markov Random Field (MRF). By applying the BLS-GSM to the observation term in the proposed model, we design a hidden state field that naturally considers the wider range relationship of wavelet coefficients. Experimental results show that the proposed algorithm shows good performance in suppressing the artifacts, while maintaining the merits of the BLS-GSM. Although the PSNR improvement is very small, the improvement of subjective quality is noticeable. The proposed algorithm reduces a large amount of the artifacts, which are the severe drawbacks of wavelet-domain image processing techniques.



(a) (b) (c)

Fig. 1. Typical artifacts of BLS-GSM. They look like wavelet basis functions and come from the incorrect inference of hidden states. (a) denoised image, (b) coarse scale error (magnification of the area in the large circle in (a)), (b) fine scale error (magnification of the area in the small circle in (a))

The rest of paper is organized as follows. Section 2 describes the image denoising process and the probabilistic model. We describe a new CRF-based model for the hidden state, and explain the inference algorithm in Section 3. Finally, we show some experimental results in Section 4, and conclude the paper in Section 5.

2. IMAGE DENOISING

In this section, we briefly review a conventional image denoising algorithm and explain the proposed wavelet coefficient estimator that extends the conventional estimator.

2.1. Image Denoising

Image denoising can be considered as an estimation problem given noisy observations. A two-step procedure for the denoising is to find a local variance from the neighborhood of a pixel, and then to apply local variance to the estimation problem. Such an approach is called *empirical Bayesian estimator* and successfully applied to many image denoising algorithms in the pixel domain. The approach has also been applied to the wavelet coefficients, i.e., the wavelet coefficients are estimated from neighboring ones. This method substantially improves the performance because of several properties of the wavelet transform. However, the two-step empirical Bayesian estimator is suboptimal although the variance estimator is optimal. To alleviate the problem caused by suboptimality, a single step optimal estimator is also proposed in [1]. When only local observation is available, the single step Bayesian estimator is the optimal. But in order to consider image context and reduce the annoying artifacts, we will introduce a single step Bayesian estimator given all observations, not just nearby ones. For the consideration of wider range of observation, we design a random field of hidden states and encode their relationship into the potential functions.

2.2. Bayes Least Square Estimator

Let us denote the true wavelet coefficient field of a given band as a random field $\mathcal{X}$, where $\mathcal{X}$ consists of N site and each random variable is denoted as $\mathcal{X}_s$, i.e., $\mathcal{X} = \{\mathcal{X}_s | s = 1, 2, \cdots, N\}$. For the statistical modeling of wavelet coefficient, we introduce the hidden state random variable $\mathcal{Z}_s$ for each site s, and denote the random field as $\mathcal{Z} = \{\mathcal{Z}_s | s = 1, 2, \cdots, N\}$ (the hidden state will be specified in the next Section). The observation is denoted as $\mathcal{Y}$, where it should be noted that $\mathcal{Y}$ is not restricted to the observation of the given band, but it can include other bands or even a lowpass band. Then the denoising of each band can be considered as the Bayes least squares (BLS) estimation of $\mathcal{X}_s$, i.e., each estimator is

$$\begin{aligned}
\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}\} &= \int \mathcal{X}_s p(\mathcal{X}_s|\mathcal{Y}) d\mathcal{X}_s \\
&= \int\int \mathcal{X}_s p(\mathcal{X}_s, \mathcal{Z}|\mathcal{Y}) d\mathcal{Z} d\mathcal{X}_s \\
&= \int\int \mathcal{X}_s p(\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}) p(\mathcal{Z}|\mathcal{Y}) d\mathcal{Z} d\mathcal{X}_s \\
&= \int \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}\} p(\mathcal{Z}|\mathcal{Y}) d\mathcal{Z}. \qquad (1)
\end{aligned}$$

Note that the difference of our least square estimator of $\mathcal{X}_s$ from the conventional ones is that the proposed estimator is conditioned on the global observation $\mathcal{Y}$, whereas the conventional methods depend on local information only. That is, the image context is naturally encoded in our model. Eq-(1) can be further reduced by using the Markovian property, $\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}\} = \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\}$:

$$\begin{aligned}
\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}\} &= \int \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\} p(\mathcal{Z}|\mathcal{Y}) d\mathcal{Z} \\
&= \int_{\mathcal{Z}_s} \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\} \int_{\mathcal{Z}_s^-} p(\mathcal{Z}|\mathcal{Y}) d\mathcal{Z}_s^- d\mathcal{Z}_s \\
&= \int_{\mathcal{Z}_s} \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\} p_s(\mathcal{Z}_s|\mathcal{Y}) d\mathcal{Z}_s \qquad (2)
\end{aligned}$$

where $\mathcal{Z}_s^-$ is $\mathcal{Z}\backslash\{\mathcal{Z}_s\}$ and the marginal distribution $p_s(\mathcal{Z}_s|\mathcal{Y})$ is defined as

$$p_s(\mathcal{Z}_s|\mathcal{Y}) = \int_{\mathcal{Z}_s^-} p(\mathcal{Z}|\mathcal{Y}) d\mathcal{Z}_s^-. \qquad (3)$$

In summary, we need to have $\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\}$ and $p_s(\mathcal{Z}_s|\mathcal{Y})$ for the denoising, where the former term will be derived in the later section from the GSM model in [1] and the latter term is followed from the CRF model for joint distribution $p(\mathcal{Z}|\mathcal{Y})$.

2.3. Gaussian Scale Mixture Model

Among many statistical models for the wavelet coefficients of natural images, the GSM is shown to provide a reasonable basis for some applications. So we use the model in [1] to characterize the local observation model for the wavelet coefficients. In the GSM of [1], the coefficients from local neighborhood form a random vector $\sqrt{z}\mathbf{u}$, where z is a *multiplier* and $\mathbf{u}$ is a zero-mean Gaussian vector. Hence, we let the hidden state $\mathcal{Z}$ as the *multiplier* in the GSM model, and compute $\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\}$ from the subset of $\mathcal{Y}$ and the state random variable $\mathcal{Z}_s$ (for details see [1]).

In contrast to the GSM model, we assume explicit finite levels (L-levels) for the *multiplier*. Assuming finite levels for the *multiplier* enables us to use combinatorial optimization techniques, which are shown to be very efficient in probabilistic inference [7, 4, 5]. The quantization of *multiplier* causes only trivial loss, because the actual implementation of BLS-GSM requires a numerical integration that needs quantization.

Another problem when applying combinatorial optimization techniques is that the density of hidden state $\mathcal{Z}_s$ is not uniform, which complicates the model. In order to cope with this problem, we use the simplest method, i.e., we find non-uniform quantization levels for the *multiplier* so that the probability of each level becomes the same. Because we use the same assumption that $\log(\mathcal{X}_s)$ has a constant prior as in [1], the discrete value of $\mathcal{X}_s$ consists of exponential scale.

3. STATISTICAL MODELING FOR HIDDEN FIELD

In this section, we derive the joint distribution $p(\mathcal{Z}|\mathcal{Y})$ and its marginal distribution $\sum_{\mathcal{Z}_s^-} p(\mathcal{Z}|\mathcal{Y})$ (discrete version of the Eq-(3)). Specifically, we use a CRF for the modeling of $p(\mathcal{Z}|\mathcal{Y})$ and a sum-product (Belief Propagation) algorithm for the inference.

3.1. Conditional Random Field

The CRF is a powerful modeling tool since they can handle overlapping and wide-range features, the definition of which is given in [8]. For two random fields $\mathcal{Y}$ and $\mathcal{Z}$, the $(\mathcal{Y}, \mathcal{Z})$ is a conditional random field if

$$p(\mathcal{Z}_s|\mathcal{Y}, \mathcal{Z}_t, t \neq s) = p(\mathcal{Z}_s|\mathcal{Y}, \mathcal{Z}_t, t \sim s) \tag{4}$$

where $t \sim s$ means t neighbors with s. Using the Hammersley-Clifford theorem and considering the first order neighborhood system, the posterior probability $p(\mathcal{Z}|\mathcal{Y})$ is given by

$$p(\mathcal{Z}|\mathcal{Y}) \propto \exp\left(-\sum_s V_s(\mathcal{Z}_s, \mathcal{Y}) - \sum_s \sum_{t \in N(s)} V_{s,t}(\mathcal{Z}_s, \mathcal{Z}_t)\right) \tag{5}$$

where $N(s)$ is the first-order neighborhood of the site s and $V_s(\mathcal{Z}_s, \mathcal{Y})$ reflects the information from the observation for a single site, and two-pixel potential $V_{s,t}(\mathcal{Z}_s, \mathcal{Z}_t)$ means spatial constraints. The energy function $V_s(\mathcal{Z}_s, \mathcal{Y})$ is defined under the assumption that it depends only on the local observation $\mathcal{Y}_s$ and thus

$$p(\mathcal{Y}_s|\mathcal{Z}_s) \propto e^{-\mathcal{V}_s(\mathcal{Z}_s, \mathcal{Y})}, \tag{6}$$

which are easily derived from the the GSM model. The relationship between the hidden states is designed according to our prior knowledge that two adjacent wavelet coefficients have large overlapping support region. That is, we can safely assume that the adjacent wavelet coefficients have smooth hidden state transition. This can be represented by the energy function

$$V_{s,t}(\mathcal{Z}_s, \mathcal{Z}_t) = \lambda(\mathcal{Z}_s - \mathcal{Z}_t)^2 \tag{7}$$

where λ is a constant.

3.2. Bayesian Inference

Finally, the joint probability of the hidden random field $\mathcal{Z}$ given a $\mathcal{Y}$ can be represented as

$$p(\mathcal{Z}|\mathcal{Y}) \propto \prod_s e^{-\mathcal{V}_s(\mathcal{Z}_s, \mathcal{Y})} \prod_{t \in N(s)} e^{-\lambda(\mathcal{Z}_s - \mathcal{Z}_t)^2}. \tag{8}$$

Also we need to compute the marginal distribution of each random variable $\mathcal{Z}_s$. For a given Markov network, the belief propagation is an iterative algorithm that estimates marginal distribution. We use loopy belief propagation, that is, the belief propagation is applied to a graph with loops. Although it does not guarantee the global optimal solution in the case of the graph with loops, loopy belief propagation has been successfully applied to many problems [6]. The standard Belief Propagation algorithm (sum-product) is defined as [7],

$$\begin{aligned} m_{st}(\mathcal{Z}_t) &\leftarrow \sum_{\mathcal{Z}_s} \psi(\mathcal{Z}_s, \mathcal{Z}_t)\psi_s(\mathcal{Z}_s, \mathcal{Y}_s) \prod_{k \in N(s)\backslash t} m_{ks}(\mathcal{Z}_s) \\ b_s(\mathcal{Z}_s) &\leftarrow \psi_s(\mathcal{Z}_s, \mathcal{Y}_s) \prod_{k \in N(s)\backslash t} m_{ks}(\mathcal{Z}_s) \end{aligned} \tag{9}$$

where m_{st} is the message that node s sends to t, $b_s(\mathcal{Z}_s)$ is the belief at the node s, and $\psi_s(\mathcal{Z}_s, \mathcal{Y}_s)$ and $\psi(\mathcal{Z}_s, \mathcal{Z}_t)$ are called local evidence and compatibility matrix respectively ($\psi(\mathcal{Z}_s, \mathcal{Z}_t) = e^{-\lambda(\mathcal{Z}_s - \mathcal{Z}_t)^2}$, $\psi_s(\mathcal{Z}_s, \mathcal{Y}_s) = e^{-\mathcal{V}_s(\mathcal{Z}_s, \mathcal{Y})}$). The iterations of the sum-product algorithm are continued until convergence. After iterations, we can get the marginal probability $p_s(\mathcal{Z}_s|\mathcal{Y})$ at the site s, and from Eq-(2) BLS estimate of the site s is given by

$$\mathbb{E}\{\mathcal{X}_s|\mathcal{Y}\} = \sum_{\mathcal{Z}_s} \mathbb{E}\{\mathcal{X}_s|\mathcal{Y}, \mathcal{Z}_s\} p_s(\mathcal{Z}_s|\mathcal{Y}). \tag{10}$$

By computing the Eq-(10) for each site, we can get the denoised band.

4. EXPERIMENTAL RESULTS

We have tested our method on a set of images (*Lena, Barbara, Peppers*) with the size of 256×256 and 512×512. The results of BLS-GSM are obtained using the software provided by the authors. The parameter λ in Eq-(7) is set to be 10, but it is found that the overall performance is not sensitive to the value of λ. In all experiments that we have performed, the BP algorithm converges within 15 iterations regardless of the size of the images and level.

The objective quality is measured in terms of PSNR with various noise levels ($\sigma = 5, 10, 15, 20, 25, 30$), and it is found that the average improvements of the PSNR are $+0.01$ dB, $+0.37$ dB, and $+0.25$ dB for the above test images respectively. Although the overall averaged PSNR improvement is small, and sometimes even below zero, the improvement in subjective quality is noticeable as demonstrated in Fig. 2. By

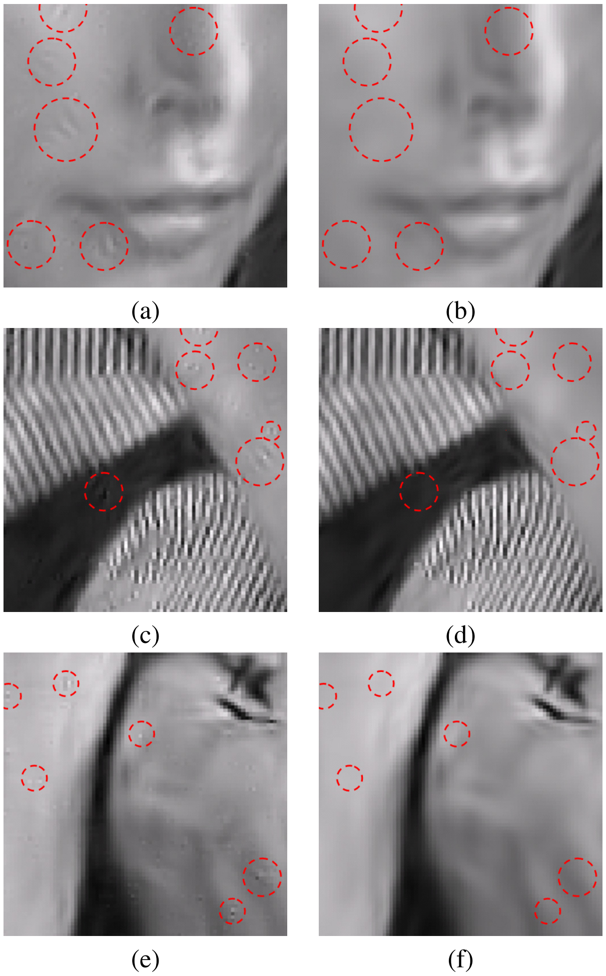


Fig. 2. Subjective comparison (Each image is cropped for visibility of the artifacts). Images in left column are the results of the BLS-GSM and images in right column are the results of the proposed one. (Please see the electronic version for the best of the view)

considering neighboring relations, the proposed model can discriminate the high state of hidden variable caused by image context from the other kinds of high state caused by local perturbation. Hence, it naturally suppresses annoying artifacts. Among many artifacts in Fig. 2, we choose 6 significant ones and draw the circles on them for easy comparison. As can be observed, the proposed algorithm reduces many artifacts without losing the merits of the BLS-GSM. Actually, there are finer scale errors (more than 30 in the case of *Peppers* image) which are corrected by the proposed algorithm.

5. CONCLUSIONS

In this paper, we have proposed a new probabilistic model for the wavelet coefficients, and developed a denoising algorithm based on the model. Differing from the conventional wavelet-based denoising algorithms using local information, the proposed method considers the correlation in larger area. Therefore the proposed algorithm can discriminate local disturbance from the edges and thus yields much less artifacts when compared with the state-of-the-art denoising algorithm. Our model is based on the CRF and the prior assumption of the smoothness of hidden state. By the appropriate quantization of the parameter in this model, the well-developed optimization techniques could have been successfully used to solve the given estimation problem.

6. REFERENCES

[1] J. Portialla, V. Sterla, M. J. Wainwright, and E. P. Simoncelli, "Image Denoising Using Scale Mixture of Gaussians in the Wavelet Domain," *IEEE Trans. Image Processing,* vol. 12. no. 11. Nov. 2003.

[2] M. Crouse, R. Nowak, and R. Baraniuk , "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models," *IEEE Trans. Signal Processing,* vol. 46, no. 4, Apr. 1998.

[3] Jia Chen, and Chi-Keung Tang, "Spatio-Temporal Markov Random Field for Video Denoisng,", In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, June. 2007.

[4] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, Nov. 2001.

[5] Y. Boykov, O. Veksler, and R. Zabih, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 9, Sep. 2004.

[6] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum, "Stereo Matching Using Belief Propagation," *IEEE Trans. Pattern Anal. Machine Intell.* , vol. 25, no. 7, July. 2003.

[7] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Bethe free energy, Kikuchi approximations and belief propagations algorithms," Technical Reprot TR-2001-16, Mitsubishi Electric Research, 2001.

[8] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Int'l Conf. Machine Learning,* pp. 282-289, 2001

[9] Y. Wang, K. F. Loe, and J. K. Wu, "A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 2, Feb. 2006.