IMAGE DENOISING BASED ON MIXTURE DISTRIBUTIONS IN WAVELET DOMAIN

Byoung Suk Bae, Jong In Lee, and Moon Gi Kang

Institute of TMS Information Technology, Yonsei University
134 Shinchon-Dong, Seodaemun-Gu, Seoul, Korea
E-mail: mkang@yonsei.ac.kr

ABSTRACT

Due to the additive white Gaussian noise (AWGN), images are often corrupted. In recent days, Bayesian estimation techniques to recover noisy images in the wavelet domain have been studied. The probability density function (PDF) of an image in wavelet domain can be described using highly-sharp head and long-tailed shapes. If a priori probability density function having the above properties would be applied well adaptively, better results could be obtained. There were some frequently proposed PDFs such as Gaussian, Laplace distributions, and so on. These functions model the wavelet coefficients satisfactorily and have its own of characteristics. In this paper, mixture distributions of Gaussian and Laplace distribution are proposed, which attempt to corporate these distributions' merits. Such mixture model will be used to remove the noise in images by adopting Maximum a Posteriori (MAP) estimation method. With respect to visual quality, numerical performance and computational complexity, the proposed technique gained better results.

Keywords: AWGN, Wavelet transform, mixture model, denoising, MAP

1. INTRODUCTION

According to the intended purposes and environments, two-dimensional signals can be acquired and transmitted through various kinds of equipments. Simultaneously unwanted signals often contaminate original information. Such polluting data are called as noise, and recently noise removal (denoising) techniques have been a major research area in image processing.

Recently digital image processing methods in transform domain are studied actively. Especially, the wavelet transforms are very attractive techniques in many fields. It is attributed to the wavelet transform's good property, that is, smart compression characteristics and clustering nature, which enables the wavelet transform to be useful in noise removal.

Because of these characteristics of wavelet, Bayesian estimation has shown good results. It is empirically known that the PDF of wavelet coefficient has a high-peak at origin and long-tail. It has been modeled as various forms, such as Gaussian, Laplace, Generalized Gaussian (GG) distributions, and so on. For example, estimator using Laplace distribution is identical to classic soft-threshold [1]. In [2], transformed signals were represented as GG distributions. However, there were some increased complexities due to additional calculations for estimating function parameters. In contrast to this, linear minimum mean square error (LLMMSE) estimator showed low complexity, however, satisfactory smoothing results [3].

The utilization of mixture model has been proposed to capture the characteristics of noise-free coefficients [4][5]. In [4], the application of Gaussian mixture model and hidden Markov model (HMM) were brought forward. Laplace mixture model also reported good denoising results [5]. That is to say, Laplace distributions can represent the coefficient's property of sharp-head and long-tail, and Gaussian also well captures the clustering characteristics of wavelet.

In this paper, the mixture model (Fig. 1), having the corporate characteristics of Gaussian and Laplace distribution, is proposed. For such purpose, the probability function of the proposed model is represented as the weighted sum of above two functions. The signal-independent AWGN is considered as noise. Using of Bayesian rule, corresponding MAP estimator will be derived as a closed-form equation. Assessments are performed with respect to visual quality, numerical efficiency, and complexity.

2. IMAGE MODEL AND CONVENTIONAL METHODS

2.1 Discrete Wavelet Transform and Two Dimensional signals

For any signal $x(t) \in L^2(R)$, the biorthogonal discrete wavelet transform (DWT) synthesis and analysis equations are written by:

$$\tilde{a}_{j,k} = \int x(t) 2^{j/2} \tilde{\phi}(2^j t - k) dt$$

$$\tilde{b}_{j,k} = \int x(t) 2^{j/2} \tilde{\psi}(2^j t - k) dt \tag{2.1}$$

$$x(t) = 2^{N/2} \sum_k \tilde{a}_{N,k} \phi(2^N t - k) + \sum_{j=N}^{M-1} 2^{j/2} \sum_k \tilde{b}_{j,k} \psi(2^j t - k) \tag{2.2}$$

Eq.(2.1) is the biorthogonal analysis (DWT) equation and Eq.(2.2) is the biorthogonal synthesis (Inverse-DWT) equation.

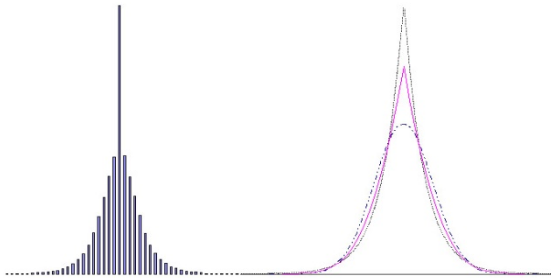

Fig. 1: Histogram of wavelet coefficients and various probability density functions.

Two-dimensional (2-D) signals, such as image, can be represented using a separable 2-D wavelet transform. A 2-D separable transform is equivalent to two 1-D transforms in series.

Noisy signal is represented by the sum of pure signal and noise. Since wavelet transform is a linear transform, transformed noisy observation can be written as:

$$\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{n} \qquad (2.3)$$

Where $\boldsymbol{y}$ is wavelet transform of noisy signal, $\boldsymbol{w}$ is wavelet coefficient by clean image, and $\boldsymbol{n}$ is the transformed noise.

2.2 Conventional Wavelet Based Denoising Algorithm

2.2.1 Wavelet Thresholding

As mentioned above, wavelet transform has smart energy compaction ability. A lot of signals have very small magnitude, simultaneously coefficients having large magnitude are sparse. Therefore small signals can be regarded as noise, and thresholding them will well preserve the important structure of original signal.

There are two categories of thresholding method. One is hard-threshold ($H(w)$) strategy. This method assumes that most of noise signal concentrate near zero point.

$$H(w) = \begin{cases} w & (if \ \ |w| > Thres.) \\ 0 & (else) \end{cases} \qquad (2.4)$$

On the other hand, soft-threshold ($S(w)$) strategy assumes that coefficients are contaminated both within threshold range and outer bound.

$$S(w) = \begin{cases} sign(w)(|w| - Thres) & (if \ \ |w| > Thres.) \\ 0 & (else) \end{cases} \qquad (2.5)$$

The hard-threshold is very sensitive to the condition of noise level. Due to this reason, in many applications the soft-threshold has been frequently used than the hard-threshold.

2.2.2 Bayesian Estimation Using Wavelet Distribution

Bayesian solutions for noise removal can be categorized into two types: one is MMSE (minimum mean squares error) estimator and the other is MAP (maximum a posteriori) estimator. First, MMSE aims to find the estimates to minimize the squared errors between original image and noisy observation. Its result can be written as following conditional mean:

$$\hat{w}_{MMSE} = \int_{-\infty}^{\infty} w P_{w|y}(w|y)dw \qquad (2.6)$$

Using Bayes rule, we can develop more above equation:

$$\hat{w}_{MMSE} = \frac{\int_{-\infty}^{\infty} w P_n(y-w) P_w(w) dw}{\int_{-\infty}^{\infty} P_n(y-w) P_w(w) dw} \qquad (2.7)$$

where $P_w(...)$ is the distribution of noise free signals, $P_n(...)$ is the distribution of AWGN, and $\hat{w}_{mmse}$ is the MMSE estimate. There is also a way to minimize the estimation error, which is maximizing the posteriori density function $P_{w|y}(w|y)$. Such MAP estimator can be formulated as:

$$\hat{w}_{MAP} = \arg\max_{w} P_{w|y}(w|y) \qquad (2.8)$$

Using Bayes rule similarly to before, following can be acquired:

$$\hat{w}_{MAP} = \arg\max_{w} P_n(y-w) P_w(w) \qquad (2.9)$$

In many applications, these two approaches result in the almost same result.

3. IMAGE DENOISING BASED ON MIXTURE DISTRIBUTIONS IN WAVELET DOMAIN

3.1 New Mixture Model

There are some limits to describe such density function using only one model. Therefore, the mixture model is a so attracting solution.

In this paper, we propose the following mixture model, which is the combination of one Gaussian distribution and one Laplace distribution with the same variance. This is given by:

$$P_w(w) = c_1 \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{w^2}{2\sigma^2}) + c_2 \frac{1}{\sqrt{2}\sigma} \exp(-\frac{\sqrt{2}|w|}{\sigma}) \qquad (3.1)$$

where $c_1 + c_2 = 1$ and $0 \le c_i \le 1$ for i=1, 2. Later, we use this density function to derive the MAP solution, and filtrate the noisy image spatial-adaptively using sample kurtosis and local variance.

3.2 MAP Estimation

From Eq.(2.3), The formulation of MAP is as following:

$$\hat{w}(y) = \arg\max_{w} P_{w|y}(w|y) \qquad (3.2)$$

By applying Bayesian rule and assuming zero-mean Gaussian noise, the following MAP estimate of w can be obtained:

$$\frac{y-w}{\sigma_n^2} - \frac{P_w'(w)}{P_w(w)} = 0 \qquad (3.3)$$

3.2.1 Single Laplace Assumption

There can be some choice of a priori knowledge $P_{\hat{w}}(\hat{w})$. First example is on single Laplace assumption.

$$\frac{y-w}{\sigma_n^2} - \frac{\sqrt{2}}{\sigma} sign(w) = 0 \qquad (3.4)$$

where ***sign*(*w*)** is either ***+1 if w > 0*** or ***-1 if w < 0.*** For most cases, it is usually accepted that ***sign*(*w*)** is the same ***sign*(*y*)**. Therefore, the following equation is obtained:

$$\hat{w} = sign(y) \cdot \max\{(|y| - \frac{\sqrt{2}\sigma_n^2}{\sigma}), 0\} \qquad (3.5)$$

where ***max{A, B}*** represents the bigger value between A and B.

3.2.2 Single Gaussian Assumption

Another example is the Gaussian case. If $P_w(w)$ is Gaussian, then Eq.(3.3) can be written as:

$$\frac{y-w}{\sigma_n^2} - \frac{w}{\sigma^2} = 0 \qquad (3.6)$$

This leads to the following equation as

$$\hat{w} = \frac{\sigma^2}{\sigma_n^2 + \sigma^2} \cdot y \qquad (3.7)$$

As seen above, MAP estimators are expressed as the closed-form, which need very simple calculations in contrast to MMSE case. That is to say, MAP is the smart estimator regarding less complexity and produces a good result similar to MMSE.

3.2.3 Proposal Model : Mixture of Laplace and Gaussian

As discussed above, the proposed model consists of two functions (Laplace and Gaussian). The new mixture model is proposed as:

$$P_w(w) = a_1 \cdot P_1(w) + a_2 \cdot P_2(w) \qquad (3.8)$$

Therefore, in this paper, Gaussian and Laplace PDF are used as,

$$P_w(w) = a \cdot G(w,\sigma) + (1-a) \cdot L(w,\sigma) \qquad (3.9)$$

where a is a constant between 0 and 1, $G(\cdot,\sigma)$ represents the Gaussian distribution with zero-mean and variance σ^2, and $L(\cdot,\sigma)$ represents the Laplace distribution with zero-mean and variance σ^2.

From Eq.(3.9), the following estimate is described:

$$\hat{w}(y) = P_G(y)\hat{w}_G(y) + P_L(y)\hat{w}_L(y) \qquad (3.10)$$

where $P_G(y)$ is the probability that w is originated from $G(\cdot,\cdot)$ *and* $P_L(y)$ is the probability that w is originated from $L(\cdot,\cdot)$. In addition, the term $\hat{w}_G(y)$ is an estimate of w based on the assumption that w is originated from $G(\cdot,\cdot)$ and $\hat{w}_L(y)$ is an estimate of w based on the assumption that w is originated from $L(\cdot,\cdot)$.

For further proceedings, consider the PDF of noisy observation y. The PDF of y is calculated by the convolution of the PDF of w and the distribution of n:

$$P_y(y) = P_w(y) * P_n(y) \qquad (3.11)$$

From Eq. (3.9), Eq. (3.11) can be written as following:

$$P_y(y) = a \cdot f_{Gn}(y) + (1-a) \cdot f_{Ln}(y) \qquad (3.12)$$

$$f_{Gn}(y) = G(y,\sigma) * P_n(y),\ f_{Ln}(y) = L(y,\sigma) * P_n(y) \qquad (3.13)$$

Now, return to Eq. (3.10). Weight probability $P_G(y)$ and $P_L(y)$ can be calculated as:

$$P_G(y) = \frac{af_{Gn}(y)}{af_{Gn}(y) + (1-a)f_{Ln}(y)} \qquad (3.14)$$

$$P_L(y) = \frac{af_{Ln}(y)}{af_{Gn}(y) + (1-a)f_{Ln}(y)} \qquad (3.15)$$

For implementation, more numerical treatments will be presented. If we re-write the above weight probability, then

$$P_G(y) = \frac{1}{1 + \frac{1-a}{a}\frac{f_{Ln}(y)}{f_{Gn}(y)}} \approx \frac{1}{1 + \frac{1-a}{a}} = a \qquad (3.16)$$

Because $P_G(y)+P_L(y)=1$, the other weight is given by:

$$P_L(y) = 1 - a \qquad (3.17)$$

Thus final estimate is written by:

$$\hat{w}(y) = a\frac{\sigma^2}{\sigma^2 + \sigma_n^2} y + (1-a) \cdot sign(y) \cdot \max\{(|y| - \frac{\sqrt{2}\sigma_n^2}{\sigma}), 0\} \qquad (3.18)$$

3.3 Kurtosis

The need of incorporating the kurtosis in image processing is owing to the special distribution of wavelet coefficients. Kurtosis can be estimated as following:

$$k_x \doteq \frac{1}{\sigma_x^4} E[(x-4)^4] = \frac{1}{\sigma_x^4} E[(x)^4] \approx \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x}{\sigma_x}\right)^4 \qquad (3.19)$$

By this relationship, the weight a can be expressed as:

$$a = \begin{cases} 0 & (if\ a < 0) \\ 1 & (if\ a > 1) \\ \frac{6\sigma^4 + 3\sigma_n^4 - \sigma_x^4 k_x}{2\sigma^4} & (else) \end{cases} \qquad (3.20)$$

4. EXPERIMENTAL RESULTS

Several conventional algorithms are applied to three standard images whose size are all 512 x 512 (“lena”, “boat”, and “babara”) to compare with proposed algorithm. BayesianShrink method and LAWML method are based on the Laplacian distribution assumption and Gaussian distribution assumption, respectively.

In order to estimate the variance (σ^2), ML (Maximum Likelihood) method is used because the results from ML are superior to those from MAP in our experiments. Also, variance of noise signal can be estimated by using information in the HH_1 band. Equations are as following:

$$\hat{\sigma}^2 = \max\{\frac{1}{N}\sum_{i=1}^{N} y(i)^2 - \sigma_n^2, 0\} \tag{4.1}$$

$$\hat{\sigma}_n^2 = \frac{Median[|x_i|]}{0.6745}, \quad x_i \in (HH_1\, band) \tag{4.2}$$

In Eq. (4.1), N is the number of neighborhood pixels.
As we can see in Table 1, proposed method works best among other conventional algorithms. Especially for “lena” case, it outperforms about +1dB than other algorithms. In view of visual aspects, proposed method removes noise effectively on the flat region so that visual quality of the proposed method is better than others. Also, proposed method is superior to the others in view of complexity. Because conventional methods using mixture model [4][5] require iterative calculation for parameters estimation, they demands relatively large loads. On the other hand, proposed method is less complex, because the method calculates using the closed-form. In our experiments, Daubechies 9/7 filter is adopted for wavelet transform.

Table 1: PSNR values [dB] from several noise removal algorithms

	BayesShrink	LAWML5 x 5	Propose
lena			
$\partial_n^2 = 100$	33.32	34.13	**34.77**
$\partial_n^2 = 400$	30.17	30.46	**31.51**
boat			
$\partial_n^2 = 100$	32.24	32.24	**32.35**
$\partial_n^2 = 400$	28.14	28.50	**28.96**
babara			
$\partial_n^2 = 100$	30.86	32.54	**32.63**
$\partial_n^2 = 400$	27.13	28.43	**28.61**

Fig. 2: Noisy image and denoising result. (a) Noisy image (σ_n^2=400) (b) 5 x 5LAWML (c) BayesShrink (d) Propose

5. CONCLUSION

In this paper, an effective noise removal algorithm is proposed. Denoising process is conducted in wavelet domain. In proposed method, its highly clustering characteristics are investigated for noise elimination in two dimensional signals. In order to utilize intra-scale dependencies, a new mixture distribution is introduced, which is combined with Bayesian estimation. The results are superior to several conventional methods in the view of performance aspects. In addition to its good performance, proposed method is very practical because it use very-customized wavelet transform and is expressed as a closed-form. These low-complexity features of proposed algorithm are expected that our method be more available to real situation.

ACKNOWLEDGEMENTS

This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul R&BD Program and by the IT R&D program of MCST/IITA (2008-F-031-01, Development of Computational Photography Technologies for Image and Video Contents).

6. REFERENCES

[1] D. L. Donoho, “Denoising by soft-thresholding”, *IEEE Trans. Inform. Theory*, 41, pp. 613-627, 1995
[2] S. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation”, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, No. 7.pp. 674-693, 1989.
[3] M.K.Mihcak, I.Kozintsev, K.Ramchandran, and P.Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients”, *IEEE Signal Processing lett.*, vol.6, pp.300-303, Dec.1999.
[4] M.S.Crouse, R.D.Nowak, and R.G.Baraniuk, “Wavelet-based signal processing using hidden Markov models”, *IEEE Trans. Signal Process.*, vol.46, pp.886-902, Apr. 1998.
[5] H.Rabbani and M.Vafadoost, “Wavelet based image denoising based on a mixture of laplace distributions”, *IJST Trans. B. Engineering*, vol.30 no.B6, pp.711-733, 2006.