Wavelet-Based Image Denoising with Mixture Laplace Model using Local Variances

Hossein Rabbani; Mansur Vafadust

ABSTRACT

The performance of various estimators, such as maximum a posteriori (MAP) is strongly dependent on correctness of the proposed model for noise-free data distribution. Therefore, the selection of a proper model for distribution of wavelet coefficients is very important in the wavelet-based image denoising. This paper presents new image denoising algorithms based on the modeling of wavelet coefficients in each subband with a mixture of Laplace probability density functions (pdfs). We also use Laplacian mixture pdf that uses local variances for the mixture model. The mixture model is able to capture the heavy-tailed nature of wavelet coefficients and the local variances can model the empirically observed correlation between the variances. Therefore, by using these relatively new models, we are able to model the statistical properties of wavelet coefficients. Within this framework, we describe a novel method for image denoising based on designing a MAP estimator, which relies on the mixture distributions. The simulation results show that our proposed technique achieves better performance than several published methods both visually and in terms of peak signal-to-noise ratio (PSNR).

KEYWORDS

Wavelet transforms, image denoising, parameter estimation, statistical modeling.

1. INTRODUCTION

Usually, noise reduction is an essential part of many image processing systems. The main sources of noise are arising from the imaging devices and from the channels during transmission [1]. In the recent years, there has been a fair amount of research on wavelet-based image denoising [1]-[8]. The motivation of denoising in the wavelet domain is that while the wavelet transform is good at energy compaction, the small coefficients are more likely caused by noise and large coefficients are caused by important signal features[4]. The small coefficients can be thresholded without affecting the significant features of the image [2]. Thresholding is a simple non-linear technique, which usually operates on one wavelet coefficient at a time [2]. In its most basic form, each coefficient is thresholded by comparing against threshold: if the coefficient is smaller than the threshold, set to zero; otherwise it is kept or modified. Replacing the small noisy coefficients by zero and applying the inverse wavelet transform on the result may lead to reconstruction with the essential signal characteristics and with less noise [2].

Many of wavelet-based denoising algorithms were developed based on soft thresholding proposed by Donoho [1]. Early methods, such as VisuShrink [2] use a universal threshold, while more recent ones, such as SureShrink [3] and BayesShrink [4] are subband adaptive algorithms and have better performance.

The problem of wavelet based image denoising can be expressed as estimation of clean coefficients from noisy data with Bayesian estimation techniques. If the MAP estimator is used for this problem, the solution requires a priori knowledge about the distribution of wavelet coefficients. Based on the distribution type, the corresponding estimator (shrinkage function) is obtained.

Various pdfs such as Gaussian, Laplace and generalized Gaussian were proposed for modeling noise-free wavelet coefficients [8]-[9]. For example, the classical soft threshold shrinkage function can be obtained by a Laplacian pdf [5]. Bayesian methods for image denoising using other distributions such as mixture models have also been proposed [10]-[15]. Because energy compactness property of the wavelet makes it reasonable to assume that essentially only a few large coefficients contain information about underlying image, the marginal distribution of wavelet coefficients is highly kurtoutic, and can be described using suitable long-tailed distributions [8]. In fact the empirical distribution is highly peaked at zero and when the distance from zero increases, it drops off more slowly than the Gaussian distribution. It has been shown that mixture pdfs are suitable models for capturing

Hossein Rabbani and Mansur Vafadust are with the Department of Bioelectrics, Biomedical Engineering Faculty, Amirkabir University of Technology (Tehran Polytechnic), Tehran, IRAN (e-mail: hrabbani@bme.aut.ac.ir, vmansur@aut.ac.ir).

this property [10]-[15]. In [11] the wavelet-based *hidden Markov model* (HMM) is proposed for statistical signal processing and a mixture of Gaussian distributions is used for modeling this heavy-tailed property of wavelet coefficients. This distribution has a closed form pdf and its parameters easily obtained using *Expectation Maximization* (EM) algorithm. Portilla *et al.* employ another class of mixture models that is called *Gaussian scale mixtures* (GSM) to model the variance dependency of wavelet coefficients [12]. This distribution also uses a type of EM algorithm to find the parameters. In [13], based on the GSM, the multivariate Laplace distribution is produced using an exponential prior for the scale factor. This pdf, can model the statistics of the discrete Fourier transform coefficients (discrete Fourier transform coefficients similar to wavelet coefficients have sparse distribution) and its parameters can easily obtained, because this pdf has a closed form. In [14], a multivariate Bernoulli-Gaussian model is proposed as prior model of noise-free wavelet coefficients in each subband that can capture jointly the inter-scale and intra-scale dependencies in the different spectral channels. Heavy-tailed distributions are also often used in various machine learning areas, e. g., *independent component analysis* (ICA). In [15], a hierarchical model based on ICA and Laplacian mixture pdf is employed for capturing the non-linear dependencies of images.

In this paper, we use a mixture of Laplace random variables to model the wavelet coefficients in each subband. Because Laplace pdf has a large peak at zero and its tails fall significantly slower than a Gaussian pdf of the same variance, a mixture of Laplace pdfs can improve modeling of wavelet coefficients distribution.

The primary properties of the wavelet transform are locality, multiresolution, and compression. The first property states that the probability structure may be defined locally, which means the intra-scale dependency of wavlet coefficients in each subband [12],[14]-[15], and the last property states that the wavelet transforms of real-world signals tend to be sparse [4]. In this paper, we also use a mixture of Laplace pdf with local variances to model these locality and heavy-tailed properties of wavelet coefficients.

The rest of this paper is organized as follows. After a brief review on the basic idea of Bayesian denoising in Section 2, we describe how the soft thresholding can be obtained using the Laplace pdf. In Section 2.2, the theoretical base of mixture models is introduced. To apply mixture model, we need to implement EM algorithm to determine model parameters. A new version of EM algorithm that finds apart parameters for each pixel is described in this section. In Section 2.3, we obtain the shrinkage function derived from our Laplacian mixture model namely, *LapMixShrink*. We also describe local version of this shrinkage function in this section. In Section 3, we use *LapMixShrink* and our new model with local variances for wavelet-based denoising of several images corrupted with additive Gaussian noise in various noise levels. The simulation results in comparison with the VisuShrink, SureShrink, BayesShrink and *hidden Markov tree* (HMT) [11], show that our algorithm achieves better performance visually and in terms of PSNR. Finally, the concluding remarks are given in Section 4.

2. Bayesian Denoising

In this section, the denoising of an image corrupted by white Gaussian noise will be considered. We observe a noisy wavelet coefficient, $y(k) = w(k) + n(k)$, where n is independent, white, zero-mean Gaussian noise and we wish to estimate the noise-free coefficient, $w(k)$, as accurately as possible according to some criteria.

The MAP estimator will be used below to estimate $w(k)$ from the noisy observation, $y(k)$ [4]. This estimator is defined as

$$\hat{w}(k) = \arg\max_{w(k)} p_{w(k)|y(k)}(w(k) \mid y(k)) \qquad (1)$$

After some manipulations, (1) can be written as

$$\hat{w}(k) = \arg\max_{w(k)} [p_n(y(k) - w(k)) p_{w(k)}(w(k))] \qquad (2)$$

We have assumed the noise is zero-mean Gaussian with variance σ_n,

$$p_n(n(k)) = (1/(\sigma_n \sqrt{2\pi})).\exp(-n(k)^2/(2\sigma_n^2))$$

By replacing (3) in (2) yields

$$\hat{w}(k) = \arg\max_{w(k)} [-(y(k) - w(k))^2/(2\sigma_n^2) + f(w(k))] \qquad (3)$$

where $f(w(k)) = \log(p_{w(k)}(w(k)))$.

Therefore, we can obtain the MAP estimate of $w(k)$ by setting the derivative to zero with respect to $\hat{w}(k)$. That gives the following equation to solve for $\hat{w}(k)$.

$$(y(k) - \hat{w}(k))/\sigma_n^2 + f'(\hat{w}(k)) = 0 \qquad (4)$$

A. Denoising Based on Local Laplace pdf

We now need a model $p_{w(k)}(w(k))$ for the distribution of wavelet coefficients. Mihcak [6] proposed a Gaussian pdf with local variance to model wavelet coefficients. We use Laplace pdf instead Gaussian pdf as

$$\begin{aligned} p_{w(k)}(w(k)) &= Laplace(w(k), \sigma(k)) \\ &= \exp(-\sqrt{2}|w(k)|/\sigma(k))/(\sigma(k)\sqrt{2}) \end{aligned} \qquad (5)$$

In this case $f(w(k)) = -\log(\sigma(k)\sqrt{2}) - \sqrt{2}|w(k)|/\sigma(k)$ thus $y(k) = \hat{w}(k) + (\sqrt{2}\sigma_n^2/\sigma(k)).sign(\hat{w}(k))$ that is often written in the following way

$$\hat{w}(k) = sign(y(k)).(|y(k)| - \sqrt{2}\sigma_n^2/\sigma(k))_+ \qquad (6)$$

Here $(a)_+$ is defined as step function with amplitude a. Let us define the *SoftL* operator as

$$SoftL(g(k),\tau(k)) := sign(g(k))(|g(k)| - \tau(k))_+ \qquad (7)$$

The shrinkage function (7) can be written as

$$\hat{w}(k) = SoftL(y(k), \sqrt{2}\sigma_n^2 / \sigma(k)) \qquad (8)$$

To apply the *SoftL* rule we need to know σ_n and $\sigma(k)$. For each data point $y(k)$, an estimate of $\sigma(k)$ is formed based on a local neighborhood $N(k)$. We use a square window $N(k)$ centered at $y(k)$. Then, we obtain an empirical estimate for $\sigma(k)$ as

$$\hat{\sigma}^2(k) = \sum_{j \in N(k)} y^2(j) / M - \sigma_n^2 \qquad (9)$$

where M is the number of coefficients in $N(k)$. When the noise variance is unknown, we can estimate it using a robust median estimator from the finest scale wavelet coefficients [3]

$$\hat{\sigma}_n^2 = \frac{median(|y_i|)}{0.6745}, \; y_i \in subband\ HH\ in\ finest\ scale \qquad (10)$$

If $\forall k, \sigma(k) = \sigma$, *SoftL* function can became as soft thresholding rule [1]. Figure 1 illustrates the histogram of HH subband of 512×512 Lena image in third scale and the best Laplace pdf and local Laplace pdf fitted to this histogram and Figure 2 shows a comparison between *SoftL* and soft thresholding function.

B. Mixture Models with Local Variances

A mixture model for a random variable has a pdf that is the sum of two simpler pdfs. In the following sections, we will use a mixture of two Laplace pdfs with local variances to model the distribution of wavelet coefficients of images as the following

$$\begin{aligned} p_{w(k)}(w(k)) &= ap_1(w(k)) + (1-a)p_2(w(k)) \\ &= a\exp(-\sqrt{2}|w(k)|/\sigma_1(k))/(\sigma_1(k)\sqrt{2}) + \\ &(1-a)\exp(-\sqrt{2}|w(k)|/\sigma_2(k))/(\sigma_2(k)\sqrt{2}) \end{aligned} \qquad (11)$$

It will be necessary to estimate the parameters $\sigma_1(k)$, $\sigma_2(k)$ and a from data. For a mixture model, an iterative numerical algorithm is required to estimate the parameters. The most frequently used algorithm to determine the parameters of a mixture model is the *Expectation Maximization* (EM) algorithm. In this paper, we use an EM algorithm with local parameters for each pixel of the subbands. The E-step calculates the responsibility factors,

$$r_1(k) \leftarrow \frac{ap_1(w(k))}{ap_1(w(k)) + (1-a)p_2(w(k))} \qquad (12)$$

$$r_2(k) \leftarrow \frac{(1-a)p_1(w(k))}{ap_1(w(k)) + (1-a)p_2(w(k))} \qquad (13)$$

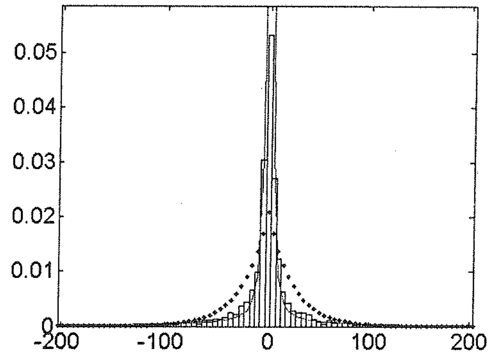

Figure 1: Histogram of HH subband of 512×512 Lena image in third scale and the best Laplace pdf (dotted line) and local Laplace pdf (solid line) fitted to this histogram.

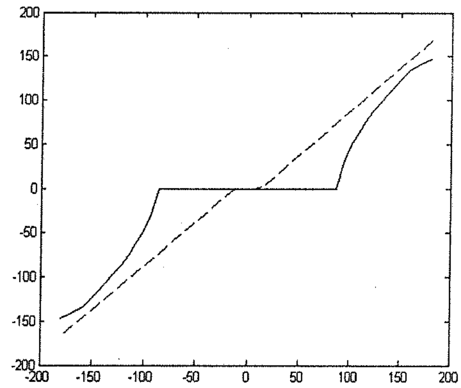

Figure 2: A comparison between *SoftL* (solid line) and soft thresholding function (dashed line).

The M-step updates the parameters $a, \sigma_1(k)$ and $\sigma_2(k)$. The mixture parameter a is computed by

$$a \leftarrow \frac{1}{N}\sum_{k=1}^{N} r_1(k) \qquad (14)$$

where N is the number of coefficients in subband. The variances $\sigma_1(k)$, $\sigma_2(k)$ are computed by

$$\sigma_i(k) \leftarrow \sqrt{2}\left(\sum_{j \in N(k)} r_i(j)|w(k)|\right) / \left(\sum_{j \in N(k)} r_i(j)\right), \quad i = 1,2 \qquad (15)$$

where M is the number of coefficients in square window $N(k)$ centered at $w(k)$. For many mixture models, a closed form for computing $\sigma_i(k)$, $i = 1,2$ does not appear. In these cases, the following formula produced from a mixture of Gaussian pdf can be used to approximate $\sigma_i(k)$:

$$\sigma_i^2(k) \leftarrow \left(\sum_{j \in N(k)} r_i(j)|w(k)|^2\right) / \left(\sum_{j \in N(k)} r_i(j)\right), \quad i = 1,2 \qquad (16)$$

If $\forall k, \sigma_1(k) = \sigma_1, \sigma_2(k) = \sigma_2$, (11) changes to

$$\begin{aligned} p_w(w) &= aLaplace(w,\sigma_1) + (1-a)Laplace(w,\sigma_2) \\ &= a\frac{\exp(-\frac{\sqrt{2}}{\sigma_1}|w|)}{\sigma_1\sqrt{2}} + (1-a)\frac{\exp(-\frac{\sqrt{2}}{\sigma_2}|w|)}{\sigma_2\sqrt{2}} \end{aligned} \qquad (17)$$

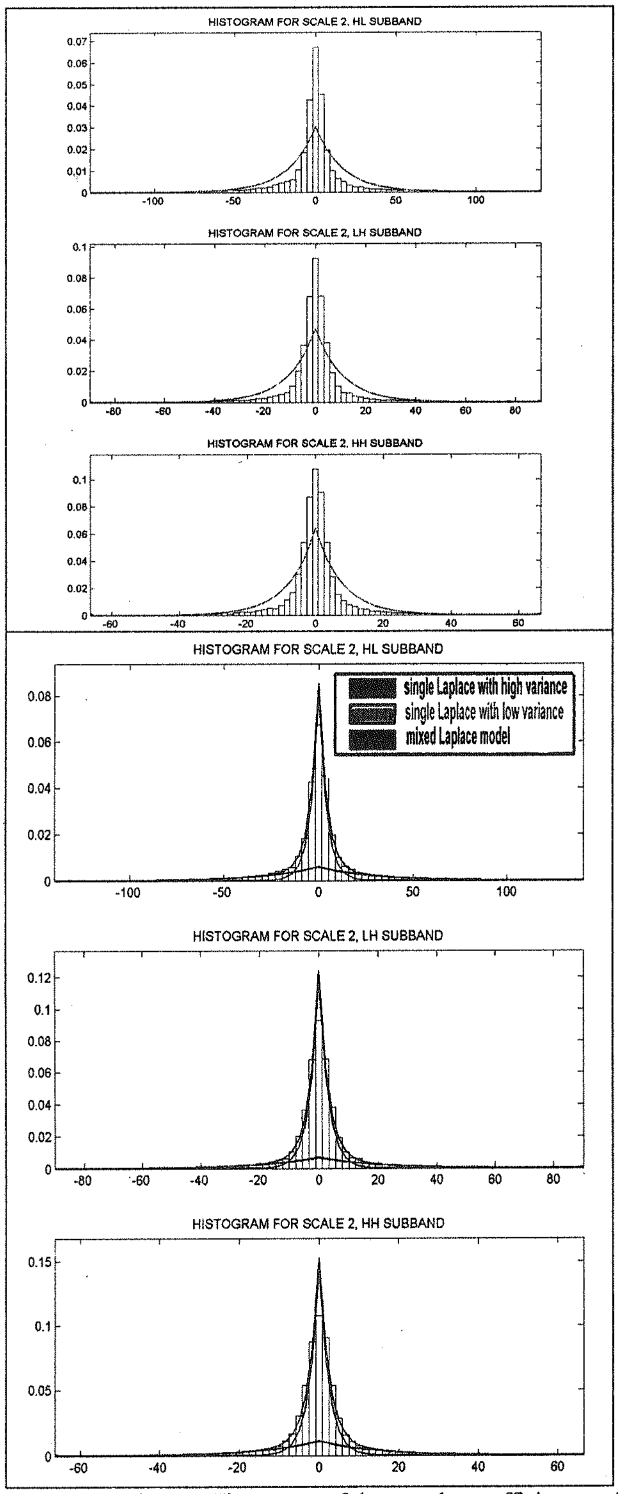

Figure 3: Top image: Histograms of the wavelet coefficients and the best fitted Laplacian mixture model in second scale of Lena image. Bottom image: Histograms of the wavelet coefficients and the best fitted Laplacian pdfin second scale of Lena image .

In this case, it will be necessary to estimate the three parameters σ_1, σ_2 and a from data. While σ_1 and σ_2 represent the standard deviation of the individual components, they are not very easily related to the standard deviation of the random variable w; for example

$$VAR[w] \neq a^2\sigma_1^2 + (1-a^2)\sigma_2^2$$

Nor are other simple relations available. The estimation of the three parameters is more difficult than it is for a single component model. For this mixture model, we use EM algorithm to estimate these parameters. A simple description of this EM algorithm can be found in Appendix.

Note that the random variable w in (16) is not the result of adding two random variables. If that were the case, then $p(w)$ would be a convolution of $p_1(w)$ and $p_2(w)$. Instead, w can be generated using a two step procedure. First, generate a binary random variable v according to

$$p(v=1) = a, \qquad p(v=2) = 1-a$$

The value of v will be either 1 or 2. For $v=1$, p_1 is used to generate w and for $v=2$, p_2 is used to generate w. Because this procedure produces a random variable w with the pdf in (16), w can be considered as being generated by either p_1 or by p_2 (even if that is not how w is physically produced).

Because Laplace pdf has a large peak at zero and tails that fall significantly slower than a Gaussian pdf of the same variance, a mixture of Laplace pdfs can improve modeling of wavelet coefficients distribution.

Figure 3 shows the best Laplacian pdf and Laplacian mixture pdf fitted to the histogram for second scale of 512×512 Lena image. For better comparison, we can see this figure in logarithmic domain in Figure 4. We see that the mixture of two Laplace pdfs follows the histogram much more closely than both Gaussian mixture model and a single Laplace pdf.

C. Denoising Based on a Mixture of Laplace pdfs with Local Variances

This section describes a non-linear shrinkage function for wavelet-based denoising that is derived by assuming that the noise-free wavelet coefficients follow a mixture model with local variances. Specifically, we assume that the noise-free wavelet coefficients are modeled as (11). We can obtain an estimate of w by the following rule

$$\hat{w}(k) = p_a(y(k))\hat{w}_1(k) + p_{1-a}(y(k))\hat{w}_2(k) \qquad (18)$$

where $p_a(y(k))$ is the probability that $w(k)$ was generated by p_1 and where similarly $p_{1-a}(y(k))$ is the probability that $w(k)$ was generated by p_2. For $i=1,2$ the expression $\hat{w}_i(k)$ is an estimate of $w(k)$ based on the assumption that $w(k)$ was generated by p_i. If p_1 and p_2 are Laplace pdfs with variances $\sigma_1(k)$ and $\sigma_2(k)$ respectively, then the *SoftL*function can be used to get $\hat{w}_1(k)$ and $\hat{w}_2(k)$. We would have

$$\begin{aligned}\hat{w}(k) &= p_a(y(k))SoftL(y(k),(\sqrt{2}\sigma_n^2)/\sigma_1(k)) \\ &+ p_{1-a}(y(k))SoftL(y(k),(\sqrt{2}\sigma_n^2)/\sigma_2(k))\end{aligned} \qquad (19)$$

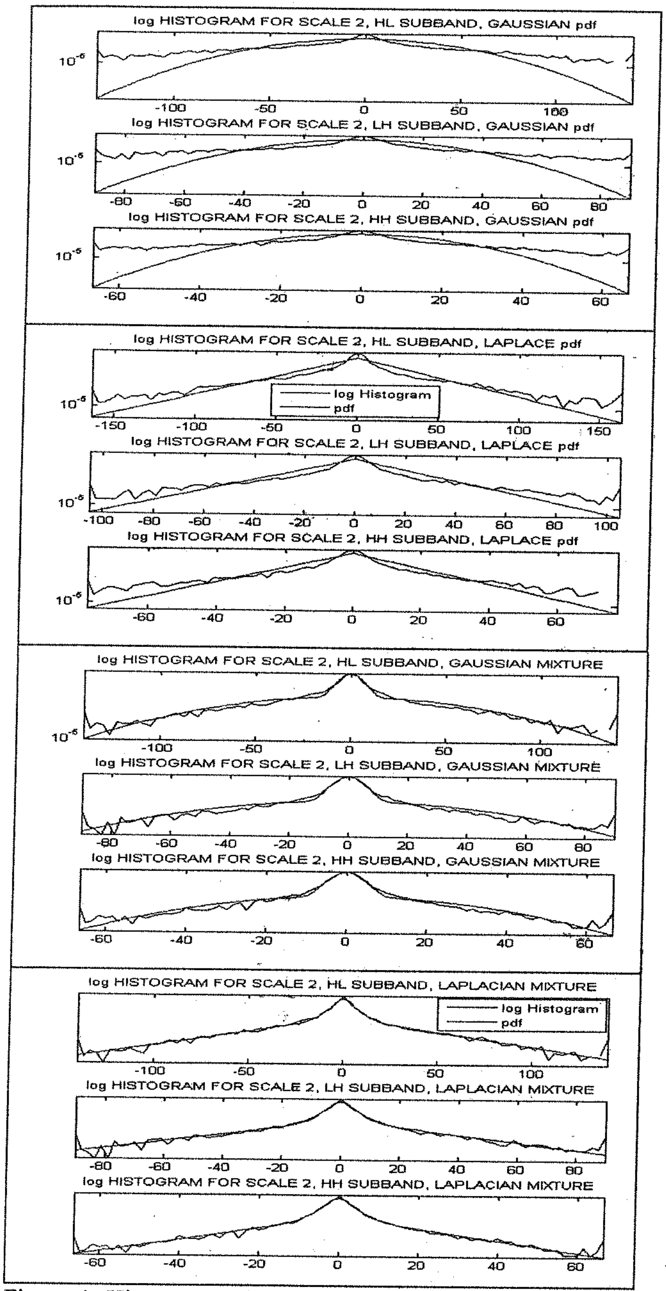

Figure 4: Histograms of the wavelet coefficients and the best fitted pdf in second scale of Lena image in the log domain. From top to bottom: Gaussian pdf, Laplace pdf, Gaussian mixture pdf and Laplacian mixture pdf.

But, we still need to determine $p_a(y(k))$ and $p_{1-a}(y(k))$. For these values, we can use the formulas based on Bayes theorem [4] as the following

$$p_a(y(k)) = \frac{a g_1(y(k))}{a g_1(y(k)) + (1-a) g_2(y(k))} \quad (20)$$

$$p_{1-a}(y(k)) = \frac{(1-a) g_2(y(k))}{a g_1(y(k)) + (1-a) g_2(y(k))} \quad (21)$$

where $g_i(y(k))$ is the pdf of $y(k)$ under the assumption that $w(k)$ was generated by p_i.

So we have

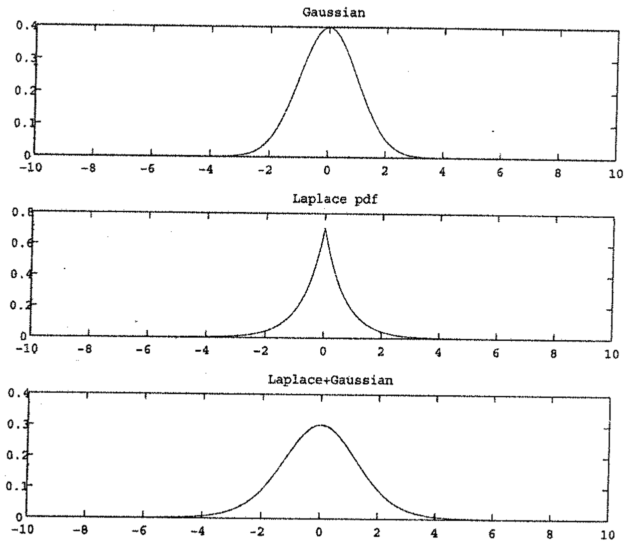

Figure 5: pdf of the sum of a Laplace and a Gaussian pdf

$$\hat{w}(k) = \frac{a g_1(y(k)) SoftL(y(k), \sqrt{2}\sigma_n^2 / \sigma_1(k))}{a g_1(y(k)) + (1-a) g_2(y(k))} + \frac{(1-a) g_2(y(k)) SoftL(y(k), \sqrt{2}\sigma_n^2 / \sigma_2(k))}{a g_1(y(k)) + (1-a) g_2(y(k))} \quad (22)$$

Because $y(k)$ is the sum of $w(k)$ and independent Gaussian noise, the pdf of $y(k)$ is the convolution of the pdf of $w(k)$ and the Gaussian pdf,

$$p_{y(k)}(y(k)) = (a p_1(y(k) + (1-a) p_2(y(k))) * p_n(y(k) = a g_1(y(k)) + (1-a) g_2(y(k))$$

where

$$g_1(y(k)) = Laplce\,(y(k), \sigma_1(k)) * Gaussian\,(y(k), \sigma_n)$$

and

$$g_2(y(k)) = Laplce\,(y(k), \sigma_2(k)) * Gaussian\,(y(k), \sigma_n)$$

$g_1(y(k))$ and $g_2(y(k))$ are not one of the standard pdfs that are commonly known. Figure 5 shows pdf of the sum of a Laplace and a Gaussian random variable. A formula for pdf of $y(k)$ that is sum of a Laplace random variable with standard deviation σ_i and a zero-mean Gaussian random variable with variance σ_n is given by [4]

$$g_i(y(k)) := LapGauss(y(k), \sigma_i(k), \sigma_n) = \exp(-\frac{y(k)^2}{2\sigma_n^2}).[erfcx(\sigma_n / \sigma_i(k) - y(k)/(\sqrt{2}\sigma_n)) + erfcx(\sigma_n / \sigma_i(k) + y(k)/(\sqrt{2}\sigma_n))]/(2\sqrt{2}\sigma_i(k)) \quad (23)$$

where $erfcx(x) = \exp(x^2)(1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt)$.

So, $y(k)$will be a mixture of two *LapGauss* pdf with the following pdf

$$p_{y(k)}(y(k)) = a LapGauss(y(k), \sigma_1(k), \sigma_n) + (1-a) LapGauss(y(k), \sigma_2(k), \sigma_n) \quad (24)$$

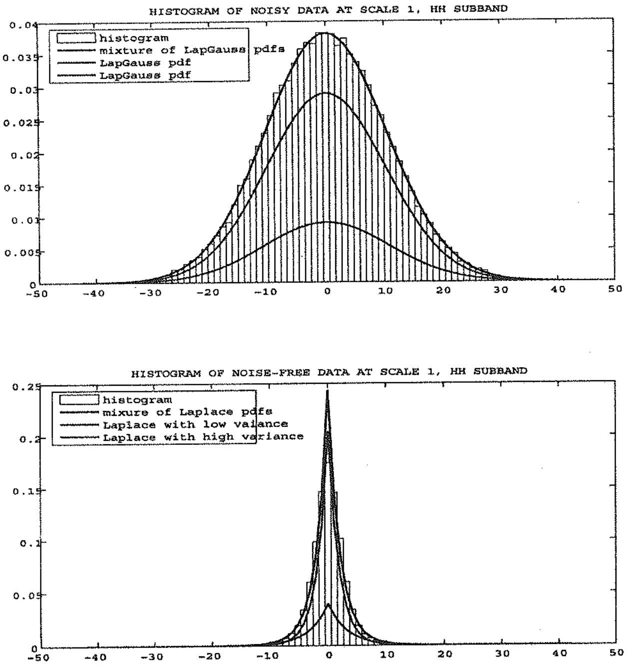

Figure 6: Histogram of the noise-free and noisy Lena image and the best fitted mixture model.

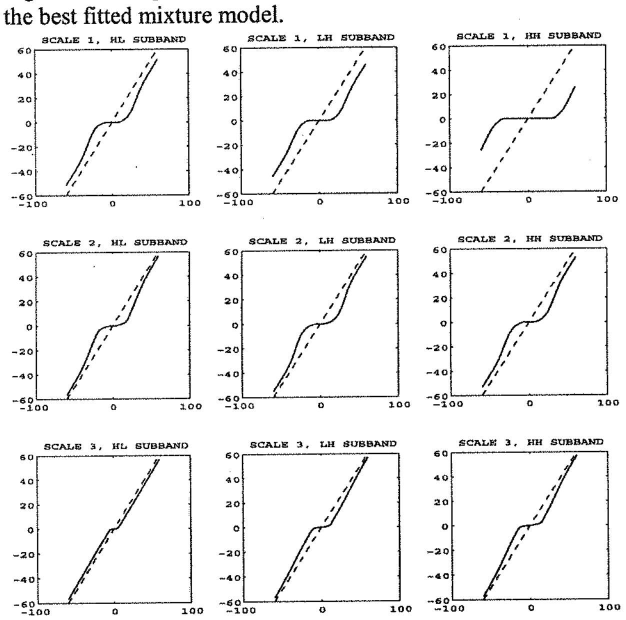

Figure 7: The threshold functions *LapMixShrink* for Lena image in each subband.

Figure 6 shows the histogram of noisy 512×512 Lena image corrupted with additive Gaussian noise with $\sigma_n=10$ and the best mixture of two *LapGauss* pdfs fitted to it.

After canceling some common terms and rearranging (20) we get $p_a(y(k)) = 1/(1+R)$ where

$$R = \frac{\frac{1-a}{\sigma_2(k)}[erfcx(\frac{\sigma_n}{\sigma_2(k)} - \frac{y(k)}{\sqrt{2}\sigma_n}) + erfcx(\frac{\sigma_n}{\sigma_2(k)} + \frac{y(k)}{\sqrt{2}\sigma_n})]}{\frac{a}{\sigma_1(k)}[erfcx(\frac{\sigma_n}{\sigma_1(k)} - \frac{y}{\sqrt{2}\sigma_n}) + erfcx(\frac{\sigma_n}{\sigma_1(k)} + \frac{y(k)}{\sqrt{2}\sigma_n})]}$$

As $p_a(y) + p_{1-a}(y) = 1$ we get $p_{1-a}(y) = R/(1+R)$ and so we can write (22) as

$$\hat{w}(k) = \frac{SoftL(y(k), \frac{\sqrt{2}\sigma_n^2}{\sigma_1(k)}) + R\,SoftL(y(k), \frac{\sqrt{2}\sigma_n^2}{\sigma_2(k)})}{1+R} \quad (25)$$

Like the soft threshold function, this shrinkage function that when $\forall k, \sigma_1(k) = \sigma_1, \sigma_2(k) = \sigma_2$ we call it *LapMixShrink* reduces (or shrinks) the value of y to estimate w. This nonlinear function does not shrink large values of y as much as the soft threshold function does. For several different values of the model parameters, some of the shrinkage functions are given in the Figure 7.

3. EXPRIMENTAL RESULTS

This section presents image denoising examples in wavelet [5] domain to show the efficiency of our new model and compare it with other methods in literature.

Figure 8 shows the denoised images obtained using Laplace pdf (soft thresholding), a mixture of 2 Laplace pdfs (*2LapMixShrink*), a mixture of 3 Laplace pdfs (*3LapMixShrink*) and a mixture of 2 Laplace pdfs with local variances (local *LapMixShrink*). The 512×512 Boat image is used for this purpose and zero mean white Gaussian with $\sigma_n = 35$ is added to the original image.

We also tested our algorithm using different additive Gaussian noise levels $\sigma_n = 10,20,30$ to three 512×512 grayscale images, namely, Lena, Barbara and Boat and compared with VisuShrink, SureShrink, BayesShrink and HMT. We implement our algorithm for known noise variance and unknown noise variance using (10). Performance analysis is done using the PSNR measure. The results can be seen in Table.1. Each PSNR value in the table is averaged over ten runs. In this table, the highest PSNR value is bolded. As seen from the results, our algorithm mostly outperforms the others.

4. CONCLUSION AND FUTURE WORKS

In this paper, we use *LapMixShrink* function based on a mixture of Laplace pdfs for modeling of wavelet coefficients in each subband. We also use local version of this algorithm. Experiments show that our model has better visual results than other methods such as soft thresholding. In order to show effectiveness of new estimator, we compared *LapMixShrink* method with effective techniques in the literature and we see that our denoising algorithm mostly outperforms the others.

Because complex wavelet transform [16] is a directional transform, it can reduce visual artifacts of denoised images. Thus if we use this algorithm in complex wavelet domain, the quality of denoised image will be improved. Instead of this shrinkage function, other nonlinear shrinkage functions can be used. For example, instead of using Laplace pdf we can use generalized Gaussian distribution or instead of using the MAP estimator we can use the *minimum mean squared error* (MMSE) estimator. Also, instead of processing each wavelet coefficient individually, better denoising results can be achieved by processing groups of wavelet coefficients together [5, 7]. Therefore, if we can use a model for wavelet coefficients that not only is a mixture but is also bivariate the performance of denoising algorithm will be improved. Because the state-of-the-art algorithms generally use local adaptive methods, using local adaptive methods in combination with mixture and bivariate models may further improve the denoising results.

5. APPENDIX: EM ALGORITHM

The Expectation-Maximization algorithm is an iterative numerical algorithm that can be used to estimate the parameters of a mixture model. Each iteration consists of an E-step and an M-step. We give here only a simple description of the EM algorithm. The mixture model is

$p(x) = ap_1(x) + bp_2(x)$

where $a+b=1$. The data is x_n for $n=1,2,\ldots N$. From the data we want to estimate the 3 parameters a, σ_1 and σ_2. The EM algorithm works by introducing an auxiliary variable that represents for each data point how likely that data point was produced by one or the other of the two components $p_1(x)$ and $p_2(x)$. This auxiliary variable is denoted by $r_1(n)$ and $r_2(n)$. $r_1(n)$ represents how responsible $p_1(x)$ is for generating data point x_n; while $r_2(n)$ represents how responsible $p_2(x)$ is for generating data point x_n.

The EM algorithm starts by initializing a, b, σ_1 and σ_2, and then proceeds with an sequence of E-M steps until the parameters satisfy some convergence condition. The initial values for a and b should satisfy $a+b=1$.

The E-step calculates the responsibility factors,

$$r_1(n) \leftarrow \frac{ap_1(x_n)}{ap_1(x_n)+bp_2(x_n)}, \; r_2(n) \leftarrow \frac{bp_2(x_n)}{ap_1(x_n)+bp_2(x_n)}$$

Note that the responsibility factors are between 0 and 1 and that $r_1(n)+r_2(n)=1$.

The M-step updates the parameters a, b, σ_1 and σ_2. In this step we maximize Q function that is defined as

$$Q = \sum_{n=1}^{N} r_1(n)\ln(a.p_1(x_n)) + r_2(n)\ln(b.p_2(x_n))$$

Based on Lagrange multiplier, the mixture parameters a and b are computed by

$$a \leftarrow \frac{1}{N}\sum_{n=1}^{N} r_1(n), \quad b \leftarrow \frac{1}{N}\sum_{n=1}^{N} r_2(n)$$

It is easy to verify that $a+b=1$ is guaranteed. One way to σ_1 and σ_2 update is to modify the basic formula for the sample variance. Instead of estimating the variance as the mean of the squares of the data values using the usual formula, $\hat{\sigma}_1^2 \leftarrow (1/N)\sum_{n=1}^{N} x_n^2$, we can estimate σ_1^2 based on the maximization of Q as a weighted sum of the data values, where the weight for x_n is the responsibility of $p_1(x)$ for the data point x_n. Because Q depend on $p_1(x_n)$ and $p_2(x_n)$, we do not have a unique formula and for each kind of pdf we would have a different formula. For example the Laplacian pdf gives the following formulas

$$\sigma_1 \leftarrow \sqrt{2}\frac{\sum_{n=1}^{N} r_1(n)|x_n|}{\sum_{n=1}^{N} r_1(n)}, \quad \sigma_2 \leftarrow \sqrt{2}\frac{\sum_{n=1}^{N} r_2(n)|x_n|}{\sum_{n=1}^{N} r_2(n)}$$

For many mixture models such as a mixture of *LapGauss* pdfs, a closed form for computing σ_1 and σ_2 does not appear. In these cases, following formulas produced from a mixture of Gaussian pdfs can be used to approximate σ_1 and σ_2.

$$\sigma_1^2 \leftarrow \frac{\sum_{n=1}^{N} r_1(n)x_n^2}{\sum_{n=1}^{N} r_1(n)}, \quad \sigma_2^2 \leftarrow \frac{\sum_{n=1}^{N} r_2(n)x_n^2}{\sum_{n=1}^{N} r_2(n)}$$

6. REFERENCES

[1] D. L. Donoho, *"De-noising by soft-thresholding,"* IEEE Trans. Inform. Theory, vol. 41, pp. 613–627, May 1995.

[2] D. L. Donoho and I. M. Johnstone, *"Ideal spatial adaptation by wavelet shrinkage,"* Biometrika, vol. 81, no 3, 1994.

[3] D. L. Donoho and I. M. Johnstone, *"Adapting to unknown smoothness via wavelet shrinkage,"* J. Amer. Statist. Assoc., vol. 90, no 432, pp. 1200–1224, 1995.

[4] M. Hansen, B. Yu, *"Wavelet Thresholding via MDL for Natural Images,"* IEEE Trans. on Information Theory, vol. 46, no 5, 2000.

[5] L. Sendur and I. W. Selesnick, *"Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,"* IEEE Trans. Signal Processing, vol. 50, no 11, pp. 2744–2756, Nov. 2002.

[6] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, *"Low complexity image denoising based on statistical modeling of wavelet coefficients,"* IEEE Signal Processing Letters, vol. 6, pp. 300–303, Dec. 1999.

[7] A. Achim, E.E. Kuruoglu, *"Image denoising using bivariate alpha-stable distributions in the complex wavelet domain,"* IEEE Signal Processing Letters, vol. 12, pp. 17-20, January 2005.

[8] E. P. Simoncelli, *"Bayesian denoising of visual images in the wavelet domain,"* Bayesian Inference in Wavelet Based Models, P. Müller and B. Vidakovic Eds. New York: Springer-Verlag, vol. 141, Lecture Notes in Statistics, March 1999.

[9] B. Vidakovic, *"Statistical Modeling by Wavelets,"* New York: Wiley, 1999.

[10] F. Abramovich, T. Sapatinas, and B. Silverman, *"Wavelet thresholding via a Bayesian approach,"* J. R. Stat., vol. 60, pp. 725–749, 1998.

[11] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, *"Wavelet-based statistical signal processing using hidden Markov models,"* IEEE Trans. Signal Processing, vol. 46, no 4, pp. 886-902, Apr. 1998.

[12] V. Strela, J. Portilla, and E. Simoncelli, *"Image denoising using a local Gaussian scale mixture model in the wavelet domain,"* in Proc. SPIE 45th Annu. Meet., 2000.

[13] Torbjørn Eltoft, Taesu Kim, and Te-Won Lee, *"On the multivariate Laplace distribution,"* IEEE Signal Processing Letters, vol. 13, no.3, pp. 300-303, May 2006.

[14] Amor Elmzoughi, Amel Benazza-Benyahia, Jean-Christophe Pesquet *"An interscale multivariate statistical model for MAP multi- component image denoising in the wavelet transform domain,"* in Proc. ICASSP 2005, pp. II-45-48, France.

[15] H. J. Park and T.-W. Lee, *"Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution,"* in Advances in Neural Information Processing Systems 17, L. K. Saul, Y.Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005.

[16] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury, *"The dual-tree complex wavelet transform - A coherent framework for multiscale signal and image processing,"* IEEE Signal Processing Magazine, 22(6):123-151, November 2005.

TABLE 1

AVERAGE PSNR VALUES OF DENOISED IMAGES OVER TEN RUNS FOR DIFFERENT TEST IMAGES AND NOISE LEVELS

NOISE VARIANCE	NOISY	VISU-SHRINK	SURE-SHRINK	BAYES-SHRINK	HMT (based on Gaussian mixture model)	2LAPMIX-SHRINK	3LAPMIX-SHRINK	LOCAL LAPMIX-SHRINK for unknown noise variance	LOCAL LAPMIX-SHRINK
LENA									
σ_n=10	28.18	28.76	33.28	33.32	33.84	33.60	33.63	34.13	**34.18**
σ_n=20	22.14	26.46	30.22	30.17	30.39	30.41	30.42	30.85	**30.88**
σ_n=30	18.62	25.14	28.38	28.48	28.35	28.67	28.75	28.96	**28.99**
BOAT									
σ_n=10	28.16	26.49	31.19	31.80	32.28	31.94	31.99	32.11	**32.34**
σ_n=20	22.15	24.43	28.14	28.48	28.54	28.59	28.63	28.88	**28.95**
σ_n=30	18.62	23.33	26.52	26.60	26.83	26.74	26.84	27.01	**27.01**
BARBARA									
σ_n=10	28.16	24.81	30.21	30.86	31.36	31.40	31.43	31.85	**32.21**
σ_n=20	22.14	22.81	25.91	27.13	27.80	27.25	27.30	27.99	**28.12**
σ_n=30	18.62	22.00	24.33	25.16	25.11	25.14	25.18	25.92	**25.99**

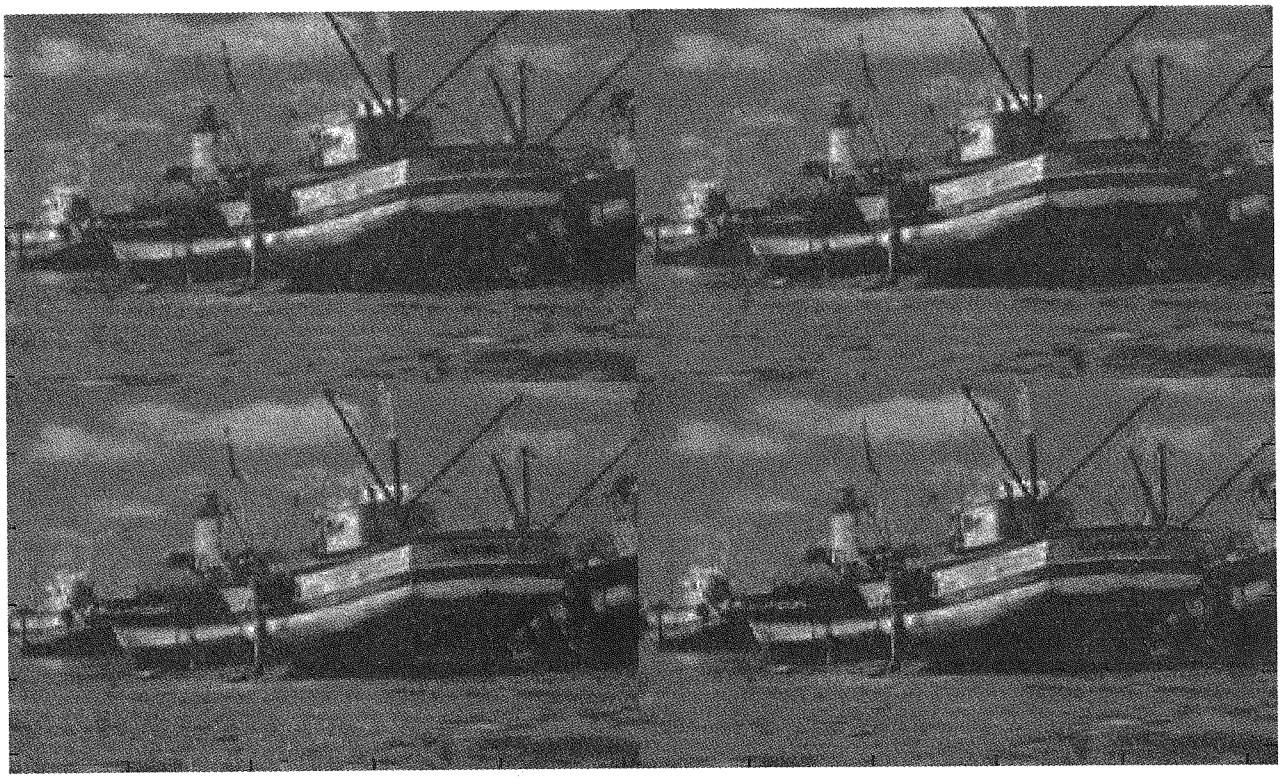

Figure 8: Denoised images. From top left, clockwise: soft thresholding, *2LapMixShrink,* local *LapMixShrink* and *3LapMixShrink*.