The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 109-115

A Bayesian Wavelet Threshold Approach for Image Denoising

Yun kee Ahn[1], Il Su Choi[2] and Sung Suk Rhee[3]

Abstract

Wavelet coefficients are known to have decorrelating properties, since wavelet is orthonormal transformation. But empirically, those wavelet coefficients of images, like edges, are not statistically independent. Jansen and Bultheel(1999) developed the empirical Bayes approach to improve the classical threshold algorithm using local characterization in Markov random field. They consider the clustering of significant wavelet coefficients with uniform distribution. In this paper, we developed wavelet thresholding algorithm using Laplacian distribution which is more realistic model.

Keywords : Bayes, image denoise, wavelet, threshold, Markov Chain Monte Carlo, Markov random field.

1. Introduction

Wavelet thresholding is a method for the reduction of noise in image. It assumes that the original image can be represented by a small number of large wavelet coefficients. In the case of an orthogonal transform, i.i.d. noise is spread out equally over all coefficients. Selecting the coefficients with the largest magnitude therefore removes most of noise, while preserving the essential image information. A number of authors have observed that wavelet coefficients have non-Gaussian distribution. The intuitive explanation for this is that images typically have spatial structure consisting of smooth areas interspersed with edges. The smooth regions lead to near-zero coefficients, and the structures give large amplitude coefficients.

The spatial structure of a wavelet representation follows from the decorrelating properties of this orthogonal transform. But this decorrelating is not complete, a wavelet transform is also a multiscale data representation and the coefficients at subsequent resolution levels tend to be

1) Professor, Department of Applied Statistics, Yonsei University, Seoul, Korea.
2) Assistant professor, Department of Applied Mathematics, Yosu National University, Yosu, Korea.
E-mail : ischoi@info.yosu.ac.kr
3) Associate professor, Department of Business Administration, Seowon University, Chongju, Korea.

correlated. Jansen and Bultheel(1999) developed the empirical Bayes approach to improve the classical threshold algorithm using local characterization in Markov random field. Even though the above method has dealt with clustering of significant Wavelet coefficients to consider decorrelating properties within the subband, it didn't take account of a realistic model.

We also developed a wavelet thresholding algorithm using Laplacian distribution which is more realistic model. This distribution has been used as a marginal statistical model to Bayesian image restoration by Simoncelli(1999). We combine this distribution with wavelet threshold for improvement. In this paper, we developed the wavelet threshold algorithm by using Laplace distribution and compared this with the existing algorithm.

2. The threshold procedure of Bayesian methods

The idea of the typical threshold algorithms is that the largest coefficients capture the essential image features. Consider an image whose pixels are contaminated with i.i.d. samples of additive Gaussian noise. Because the wavelet transform is orthonormal, the noise is also Gaussian in the wavelet domain. Thus, each coefficient in the wavelet expansion of the noise image can be written as the following type:

$$\boldsymbol{W} = \boldsymbol{V} + \boldsymbol{N}$$

where $\boldsymbol{N}$ is the noise vector, $\boldsymbol{V}$ is the uncorrupted wavelet coefficient vector, and $\boldsymbol{W}$ is the input wavelet coefficient vector.

The following wavelet threshold algorithms were proposed by Donoho and Johnstone(1995). In *hard thresholding* one replaces $W_{\lambda i}$ by

$$W_{\lambda i} = W_i X_i$$

where λ is a certain threshold and mask variable X_i such that $X_i = I\{|W_i| > \lambda\}$. It keeps the uncorrupted coefficients and replaces the noisy ones by zero.
In *soft thresholding* one replaces $W_{\lambda i}$ by

$$W_{\lambda i} = (|W_i| - \lambda)_+ sign(W_i)$$

The wavelet estimator with soft thresholding is also known as the wavelet shrinkage estimator since it is related to Stein's shrinkage estimator.

If we have a prior distribution $P(\boldsymbol{X} = \boldsymbol{x})$ and a conditional model $f_{W|X}(\boldsymbol{w}|\boldsymbol{x})$, then the Bayes' rule allow us to compute the posterior probability:

$$P(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{W} = \boldsymbol{w}) = \frac{P(\boldsymbol{X} = \boldsymbol{x}) f_{W|X}(\boldsymbol{w}|\boldsymbol{x})}{f_W(\boldsymbol{w})}$$

3. Prior and conditional model

3.1 The prior model

We are looking for a multivariate distribution for a binary image $\boldsymbol{X}$. Our model of the image is a Markov random field of binary pixels with the value of each pixel site affected by its nearest neighbors. Equivalently, it is modeled by the Gibbs distribution

$$P(X=x)=\frac{1}{Z}\exp\{-U(x)\}$$

where x is configuration of the N pixels, $x_i, i=1,\cdots,N$, the energy function U reflects the neighborhood structure, and Z is the normalizing constant(called the partition function in mathematical physics). The particular prior probability we place on the configuration is the Ising model for which

$$U(x)=-\beta\sum_{\{i,j\}\in\partial}x_i x_j$$

where the sum is taken over pairs of sites (i,j) which are nearest neighbors, and β is positive parameter.

3.2 The conditional model

We need a conditional density $f_{W|X}(\boldsymbol{w}\,|\,\boldsymbol{x})$. Since this conditional model deals with the local significance measure, we write $f_{W|X}(\boldsymbol{w}|\,\boldsymbol{x})=\prod f(w_i|x_i)$. This density expresses that if the label $X_i=1$, i.e. if the corresponding wavelet coefficient is sufficiently noise-free, a large value of V_i is probable. We now consider a model for noise-free wavelet coefficients following Laplacian distribution with a two-parameter density of the form[4]:

$$P_{V_i|X_i}(v|1)=\frac{e^{-|v/s|^p}}{Z(s,p)}$$

where the normalization constant is $Z(s,p)=2^{s/p}\Gamma(\frac{1}{p})$. The parameters $\{s,p\}$ are estimated by maximizing the likelihood of the data under the model. Typical values for p range between 0.5 and 0.8. If noise N_i has a Gaussian density $N(0,\sigma^2)$, it follows to verify that

$$f_{W_i|X_i}(w|1)=\int_{-\infty}^{w}Ze^{-s^{p}(w-n)^{p}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{n^2}{2\sigma^2}}dn+\int_{w}^{\infty}Ze^{s^{p}(w-n)^{p}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{n^2}{2\sigma^2}}dn$$

The more simple model has been studied by the assuming that noise-free coefficients are uniformly distributed on $[-\mu,-\sigma]\cup[\sigma,\mu]$, μ being the maximum coefficient magnitude, which is a parameter of the model that has to be determined[2]. If noise N_i has a Gaussian density $N(0, \sigma^2)$, it is easy to verify that

$$f_{W_i|X_i}(w|1) = \frac{1}{2(\mu - \sigma)}[\Phi(w+\mu) - \Phi(w+\sigma) + \Phi(w-\sigma) - \Phi(w-\mu)],$$

where $\Phi(z)$ is the cumulative Gaussian distribution.

4. The Bayesian algorithm

From Bayes rule, we can compute the posterior probabilities

$$P(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{W} = \boldsymbol{w}) = \frac{P(\boldsymbol{X} = \boldsymbol{x}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x})}{f_{\boldsymbol{W}}(\boldsymbol{w})}$$

It computes in each site i the marginal probabilities by the Maximal marginal posterior rule:

$$P(X_i | \boldsymbol{W} = \boldsymbol{w}) = \sum_{\boldsymbol{x}} x_i P(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{W} = \boldsymbol{w})$$

and if this probability is more than 0.5, the pixel gets value 1. Both decision rules have a binary outcome: each coefficient is classified as noise $(X=0)$ or relatively uncorrupted $(X=1)$.

The computation of $P(X_i = 1 | \boldsymbol{W})$ involves the probability of all possible configurations $\boldsymbol{x}$. Mostly, the samples are generated, not independently of each other, but in a chain, hence it is named as Markov Chain Monte Carlo(MCMC) estimation. Suppose we want a sample from the distribution with joint density $P(\boldsymbol{x}) = P(x_1, \cdots, x_N)$. An initial configuration, $\boldsymbol{x}^{(0)}$, is chosen. At each iteration, this configuration is updated at one point. The point to be updated is selected at random. The updated value is based on its conditional probability given the current value of all the other sites. These conditional probabilities are referred to as the full conditionals. For example, if site i is chosen for updating at iteration t, its new value is a sample from

$$P(x_i^t | x_{\hat{i}}^{t-1}) = P(x_i^t | x_1^{t-1}, \cdots, x_{i-1}^{t-1}, x_{i+1}^{t-1}, \cdots, x_N^{t-1})$$

where $x_{\hat{i}}$ refers to the configuration of all of the pixels except the i^{th}. They are independent of the partition function and depend only on the current values of the nearest neighbors. In the case the full conditionals are

$$P(x_i^t | x_{\hat{i}}^{t-1}) = \frac{e^{\beta \sum_{j \in \partial i} x_i x_j}}{e^{\beta \sum_{j \in \partial i} x_j} + \sum_{j \in \partial i} 1}$$

where ∂i is the set of points that are neighbors of point i. Next we can compute the probability ratio of two subsequent samples:

$$r^{(i)} = \frac{P(\,X^{(i+1)} \mid W)}{P(\,X^{(i)} \mid W)}$$

This is the only quantity needed by the algorithm, and if

$$P(X = x \mid W) = \frac{1}{Zf_w(w)} \exp[-U(x)]\, f_{w|x}(w \mid x)$$

there is no need for the enormous computation of the partition function $Zf_w(w)$.

This is the classical Metropolis sampler(Metropolis et al(1953)). The chain of states is started from an initial state $X^{(0)}$. The successive samples $X^{(i)}$ are then produced as follows: a candidate intermediate state is generated by a local random perturbation of the actual state. Then the probability ratio r of the actual state and its perturbation is computed. Since the Gibbs distribution is based on local potential functions, only positions i whose mask labels are switched by the perturbation or which have a switched label in their neighborhood ∂i are involved in the computation. If the probability ratio is larger than one, then the new state is accepted, otherwise it is accepted with probability equal to r, i.e. the acceptable probability is $\min[r^{(i)}, 1]$. To generate a completely new sample, we repeat this local switch for all locations in the grid.

5. The algorithm and its results

5.1 Algorithm summary

This is a schematic overview of the subsequent steps of the algorithm:

1. Compute the stationary wavelet transform W of the input.
2. At each level and for each component, select the appropriate threshold $\lambda = \sqrt{2\log(size)}$. This threshold generates an initial label image $X^{(0)}$.
3. Estimate the parameters $\{s, p\}$ and σ from the collection of input wavelet coefficients $\{w_i\}$. A simple solution is a maximum likelihood estimator:

$$\{\hat{s}, \hat{p}, \hat{\sigma}\} = \arg \max_{\{s, p, \sigma\}} \prod_i P_w(w_i{}' s, p, , \sigma,) = \arg \max_{\{s, p, \sigma\}} \prod_i \int e^{-|v/s|^p} e^{-(w_i - c)^2 / 2\sigma^2} dv$$

4. Run a stochastic sampler to estimate for coefficient at the given resolution level the probability $P(X_i \mid W)$. Use $X^{(0)}$ from the previous step as the starting sample. A MCMC algorithm produces the sequence of samples.
5. $\widehat{W}_{\lambda i} = W_i P(X_i \mid W)$
6. Inverse wavelet transform yields the result.

5.2 Results and discussion

In this sectin, we show examples of two empirical thresholding models for visual images. These procedures are deigned for application in image noise reduction. Jansen and Bultheel(1999) developed the Empirical Bayes Approach to improve the classical threshold algorithm using local characterization in Markov random field. Even though the above method has dealt with clustering of significant Wavelet coefficients, it was to consider the prior model taking these line singularities into account, it didn't take account of a realistic model.

We also developed wavelet thresholding algorithm using Laplacian distribution which is more realistic model. This distribution has been developed as a marginal Statistical model for Bayesian image restoration by Simoncelli(1999). We used this distribution for the wavelet threshold to improve denoising.

Table 1 shows mismatching counts($= \sum I(|\hat{Y}_i - Y_i| > difference)$ for all three algorithms, applied to a 256×256 size woman picture. Laplace algorithm outperform the other two techniques. Finally, figures 1. shows example images. Laplace model appear sharper and less noisy than the other models.

difference	noise	hard threshold	uniform	Laplace
0.20	174	291	186	174
0.25	138	90	45	27
0.3-0.7	138	51	30	15
0.75	138	48	30	15
0.8	132	36	24	15
0.85	129	33	24	15
0.9	48	15	15	9
0.95	21	9	12	9

Table 1. Denoising results for three estimators, All values indicate mismatching counts.

References

[1] Donoho, D. and Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, Vol. 90, 1200-1224.

[2] Jansen, M. and Bultheel, A. (1999). Empirical Bayes approach to improve wavelet thresholding for image noise reduction. Tech. report, Katholieke Universiteit Leuven

[3] Metropolis, N. et al.(1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, Vol. 21, 1087-1092.

[4] Simoncelli, E. P. (1999). Bayesian Denoising of Visual Images in the Wavelet Domain. In P. Müller and B. Vidakovix, editors, Bayesian Inference in Wavelet Based Models, Springer-Verlag, 141, 1-25.

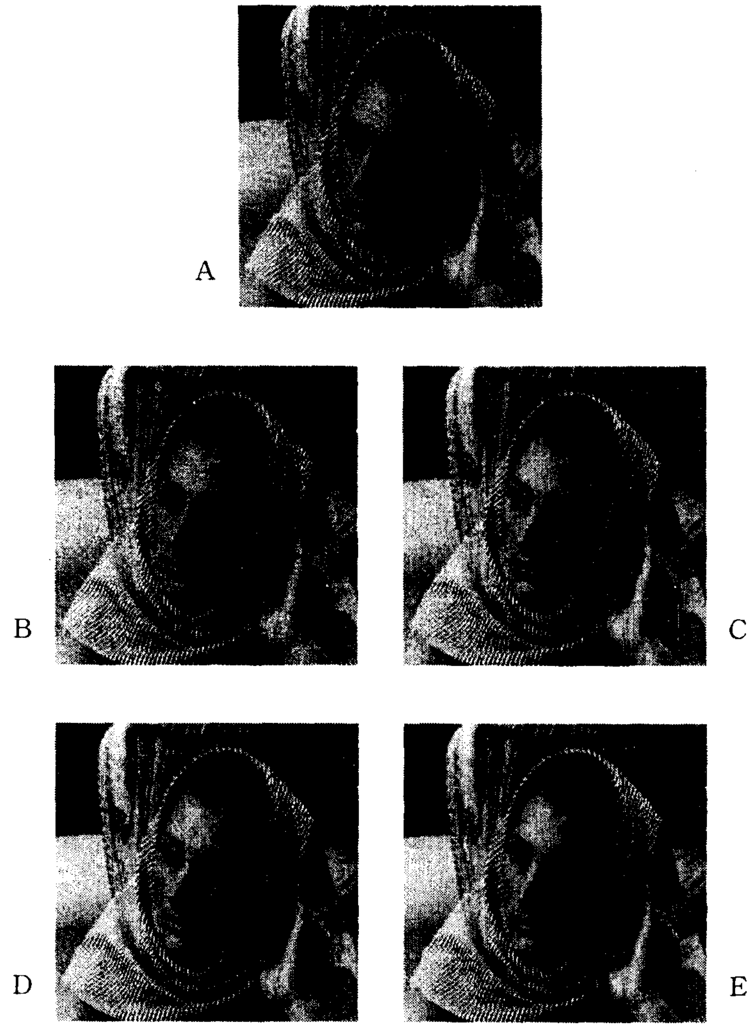

Figure 1. Denoising results A: Original "Woman" image. B: Noisy image(Gaussian noise). C: Hard threshold model. D: Uniform distribution model. E: Laplace distribution model.