Denoising of a Positive Signal with White Gaussian Noise by Using Wavelet Transform

*Ja-Yong Koo, *Hyun-Suk Park, **Kuk-Nam Park and ***Ji-sang Yoo

※This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1996.

Abstract

Given a noisy signal sampled at equispaced points with white noise, we consider problems where the signal to be recovered is known to be positive; for example, images, chemical spectra or other measurements of intensities. Shrinking noisy wavelet coefficients via thresholding offers very attractive alternatives to existing methods of recovering signals from noisy data. In this paper, we propose a method of recovering the original signal from a corrupted noisy signal, guaranteeing the recovered signal positive. We first obtain wavelet coefficients by thresholding, and use a nonlinear optimization to find the denoised signal which must be positive. Numerical examples are used to illustrate the performance of the proposed algorithm.

I. Introduction

The objectives of signal processing are to analyze accurately, code efficiently, transmit rapidly, and then to reconstruct carefully at the receiver the delicate oscillations or fluctuations of this function of time. In fields ranging from Extragalactic Astronomy to Molecular Spectroscopy to Medical Imaging to Computer Vision, the recovery of a signal or image from noisy data is imperative. Given a noisy signal, we focus only on problems where the signal to be recovered is known to be positive; think of images, chemical spectra or other measurements of intensities.

Wavelet theory has inspired the development of a powerful methodology for processing signals, images, and other types of scientific and technical data. Wavelet transform allows better resolution in time and frequency compared with the classical Fourier transform, and thus allows one to see "the forest and the trees". This feature is important for nonstationary signal analysis [1], [2].

Recently, it has been shown that wavelet shrinkage offers very attractive alternatives to existing methods of recovering signals from noisy data, where wavelet shrinkage refers to reconstructions obtained by wavelet transformation. This is then followed by shrinking the empirical wavelet coefficients towards zero, and followed by inverse transformation [3]. When the signal to be recovered is known to be positive, the prior knowledge of positivity should be built into a reconstruction algorithm [4]. If we do not incorporate such side condition into wavelet shrinkage, the reconstructed signal may be negative, which then consequently leads to an infeasible solution.

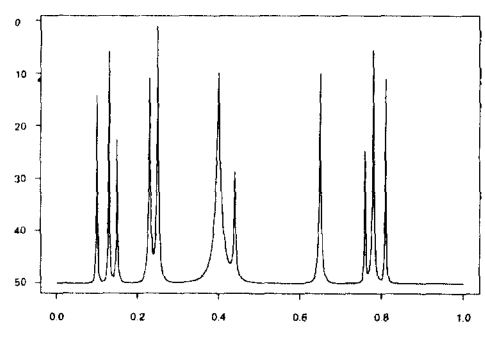


Figure 1. A Bumps function:Plot of $(t_i, f(t_i))$, $i=1, \ldots, n$.

In this paper, we propose a method of denoising a positive signal. For a function h, let $h_{j,k}(t) = 2^{j/2}h(2^j t - k)$ where $j, k \in Z$. Let us suppose that the inhomogeneous wavelet basis is derived from $\{\phi_{j_0,k} : k \in Z\}$ and $\{\psi_{j,k} : k \in Z,\ j \ge j_0\}$, where ϕ and ψ are a scaling function and a mother wavelet, respectively. Then, a signal f has a formal expansion

$$f(t) = \sum_k \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j \ge j_0}\sum_k \beta_{j,k}\psi_{j,k}(t).$$

* Department of Statistics, Hallym University

** Department of Electronic Engineering, Hallym University

*** Department of Electronic Engineering, Kwangwoon University

Manuscript Received : January 21, 1998.

The vast majority of the wavelet coefficients of the function given in Fig. 1 are zero or effectively zero and the large coefficients occur exclusively near the areas of major spatial activity [3]. Wavelet shrinkage uses this idea to select wavelet basis functions which are significant in reconstructing a signal. We use this fact to find a set of wavelet basis functions and then consider the denoiser in the form of

$$\hat{f}(t) = \exp\left(\sum_k \hat{\alpha}_{j_0,k} \phi_{j_0,k} + \sum_{j \ge j_0} \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t) \right) \quad (1)$$

where the wavelet basis functions are those wavelets with their absolute values of the corresponding wavelet coefficients greater than a threshold. The parameters $\hat{\alpha}_{j_0,k}$ and $\hat{\beta}_{j,k}$ are estimated by minimizing the residual sum of squares which will be shown in section 2.2. We consider the denoiser in the form of equation (1) since it is automatically positive. According to a simulation carried out in this paper, the proposed denoiser performs better than the wavelet shrinkage if the sought-after signal has several regions where it is close to zero such as the Bumps function shown in Fig. 1.

This paper is organized as follows. Section II describes wavelet shrinkage and the proposed denoising algorithm. Experimental results are given in Section III.

II. Recovery of Positive Signals

Shrinking noisy wavelet coefficients via thresholding offers very attractive alternatives to existing methods of recovering signals from noisy data. This new method, wavelet shrinkage, has theoretical properties that by far surpass anything previously known. This method, wavelet thresholding method, works well in problems ranging from photographic image restoration to medical imaging.

Suppose we are given a noisy signal

$$y_i = f(t_i) + \sigma\, \varepsilon_i \quad (i = 1, \ldots, n) \quad (2)$$

sampled at the equispaced points $t_i = i/n$ on [0, 1], where the ε_i is a white noise and f a function defined on the unit interval [0, 1], is an unknown signal which we would like to recover. Often we have a constraint indicating bounds on the signal; that is, the solution cannot have negative values. This constraint is extremely powerful. It is usually the easiest of the amplitude bounds to implement and is sometimes inherent in the way the solution is represented. We also consider spatial bounds; the signal is known to vanish over certain regions within well-defined limits. In this paper, we assume this spatial bounds as the unit interval [0, 1] for notational convenience. In practice, this type of constraint usually exerts a weaker influence on the solution than the amplitude bound.

2.1 Wavelet Shrinkage

The motivation of wavelet shrinkage is twofold [3]. First, for a spatially varying signal, most of the action is concentrated in a small subset of $(j, k)-$ space. Secondly, under the noise model underlying (2), the noise contaminates all wavelet coefficients equally.

Given ϕ and ψ, one may compute the wavelet coefficients by the formulae

$$u_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(t_i) y_i \quad \text{and} \quad w_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(t_i) y_i .$$

Because of the spatial localization of wavelet bases, the wavelet coefficients $w_{j,k}$ allow one to easily answer the question 'is there a significant change in the signal near t_0 ?' by looking at the wavelet coefficients at levels $j = j_0, \ldots, J$ at spatial indices k with $k2^{-j} \approx t_0$. If these coefficients are large, the answer is 'yes'. Typically the large coefficients occur exclusively near the areas of major spatial activity. This property has lead to the selective wavelet recovery [3].

Motivated by the fact that only a very few wavelet coefficients contribute to the signal, we consider the threshold rules that retain only the observed to exceed a multiple of the noise level. Define the 'hard' and 'soft' threshold nonlinearities by

$$\eta_H(w, \delta) = w 1\{|w| > \delta\} = \begin{cases} w & \text{if } |w| > \delta \\ 0 & \text{if } |w| \le \delta \end{cases} \quad (3)$$

and

$$\eta_S(w, \delta) = sgn(w)(|w| - \delta)_+ = \begin{cases} w - \delta & \text{if } w > \delta \\ 0 & \text{if } |w| \le \delta \\ -w + \delta & \text{if } w < -\delta \end{cases} \quad (4)$$

For example, if we use the hard threshold and if a wavelet coefficient is greater than δ, then it is unchanged;however, if it is less than or equal to δ, then we do not use it in the wavelet expansion.

Given a set Λ of (j, k) pairs for which the wavelet coefficients $\hat{w}_{j,k}$'s are not zero, the recovered signal via wavelet shrinkage is defined by

$$\widehat{WS}(t) = \sum_k \hat{u}_{j_0,k}\phi_{j_0,k} + \sum_{(j,k)\in\Lambda} \hat{w}_{j,k}\psi_{j,k}(t), \quad t\in[0,1]. \tag{5}$$

Here it can be seen that all of the coefficients $\hat{u}_{j_0,k}$ are used. This provides reconstructions by selecting only a subset of the empirical wavelet coefficients.

A signal f can be positive when it is expanded as in the equation (1). However, it may not be positive over the interval [0, 1] if we limit the expansion at a resolution level j_1. This argument applies to the wavelet shrinkage in that the denoiser in (5) may have negative ordinates once we apply the threshold rule (3) or (4).

2.2 Positive Denoising

Our denoising method with positivity is to select the indices (j, k) based on wavelet shrinkage and then to approximate the logarithm of f in the form of (1). Hence, our method is a postprocessing of wavelet shrinkage under the assumption that wavelet shrinkage should provide us with the important coefficients and then taking their exponential could improve $\widehat{WS}$.

Define Λ_ϕ to be the set of indices (j_0, k) for which $\phi_{j_0,k}$ does not vanish on [0, 1] and Λ_ψ to be the set of double indices (j, k) for which $\hat{w}_{j,k}$ is not zero. Let Λ be the union of Λ_ϕ and Λ_ψ. For $\lambda=(j,k)\in\Lambda$, we let $\phi_\lambda(t)=2^{j/2}\phi(2^{jt}-k)$ and let $\psi_\lambda(t)=2^{j/2}\psi(2^{jt}-k)$. We let $\{B_\lambda\}_{\lambda\in\Lambda}$ denote the set of wavelet basis functions:

$$\{B_\lambda\}_{\lambda\in\Lambda} = \{\phi_\lambda : \lambda\in\Lambda_\phi\} \cup \{\psi_\lambda : \lambda\in\Lambda_\psi\}$$

Let θ_λ be the parameter corresponding to B_λ and let $\boldsymbol{\theta}=(\theta_\lambda)_{\lambda\in\Lambda}$ denote the vector of elements θ_λ. For $t\in[0,1]$, let

$$f(t\,;\boldsymbol{\theta}) = \exp\left[\sum_\lambda \theta_\lambda B_\lambda(t)\right]. \tag{6}$$

We use $f(t\,;\boldsymbol{\theta})$ as the postprocessor of wavelet shrinkage so that it is positive. Consider the optimization problem of minimizing the following residual sum of squares

$$\Delta(\boldsymbol{\theta}) = \sum_{i=1}^{n}[y_i - f(t_i\,;\boldsymbol{\theta})]^2. \tag{7}$$

Note that since y_i and t_i are fixed observations, the residual sum of squares is a function of $\boldsymbol{\theta}$. We shall denote by $\widehat{\boldsymbol{\theta}}$, a least squares estimate (LSE) of $\boldsymbol{\theta}$, that is a value of $\boldsymbol{\theta}$ which minimizes $\Delta(\boldsymbol{\theta})$. Once we find the LSE $\widehat{\boldsymbol{\theta}}$, our reconstructed signal at each $t\in[0,1]$ is defined by

$$\widehat{PD}(t) = f(t\,;\widehat{\boldsymbol{\theta}}).$$

If the noise is a white Gaussian noise with zero-mean and the signal is represented by (6) with true parameter $\boldsymbol{\theta}^*$, then the LSE is equivalent to the maximum likelihood estimate. Assume that $y_i = f(t_i;\boldsymbol{\theta}^*)+\epsilon_i$ and $\epsilon\sim N(\mathbf{0}, I\sigma^2)$ with unknown parameter $\boldsymbol{\theta}^*$ and known σ. Since the samples are independent, we have a likelihood function of $L(\boldsymbol{\theta})$ given by

$$L(\boldsymbol{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left[-\frac{\Delta(\boldsymbol{\theta})}{2\sigma^2}\right].$$

Let us observe that maximizing $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is equivalent to minimizing $\Delta(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Hence, the LSE of $\boldsymbol{\theta}$ is also the maximum likelihood estimate of $\boldsymbol{\theta}^*$. On the other hand, we can justify our denoiser in the sense of least squares even when the noise is not Gaussian.

To find the LSE $\widehat{\boldsymbol{\theta}}$, we use a modified version of the Newton-Rapson method. Differentiating the equation (7) provides the normal equations.

$$\sum_{i=1}^{n}(y_i - f(t_i\,;\boldsymbol{\theta}))f(t_i\,;\boldsymbol{\theta})B_\lambda(t_i) = 0 \tag{8}$$

for $\lambda\in\Lambda$. Let $S(\boldsymbol{\theta})$ be the vector whose λ-th element is the left hand side of (8) and let $H(\boldsymbol{\theta})$ be the Hessian matrix of $\Delta(\boldsymbol{\theta})$ whose (λ,λ')-th element is given by

$$\frac{\partial^2\Delta(\boldsymbol{\theta})}{\partial\theta_\lambda\partial\theta_{\lambda'}} = \sum_{i=1}^{n}(2f(t_i\,;\boldsymbol{\theta}) - y_i)B_\lambda(t_i)B_{\lambda'}(t_i)$$

It can be seen that the Hessian matrix $H(\boldsymbol{\theta})$ is not necessarily positive definite.

Our method of computing $\widehat{\boldsymbol{\theta}}$ is to start with an initial guess $\boldsymbol{\theta}^0$ and iteratively determine $\boldsymbol{\theta}^{m+1}$ according to the formula

$$\boldsymbol{\theta}^{m+1} = \boldsymbol{\theta}^m + H^{-1}(\boldsymbol{\theta}^m)S(\boldsymbol{\theta}^m). \tag{9}$$

Since the LSE $\hat{\theta}$ should satisfy the normal equation $S(\hat{\theta})=0$, the Taylor expansion gives

$$0 = S(\hat{\theta}) \approx S(\tilde{\theta}) - H(\tilde{\theta})(\hat{\theta} - \tilde{\theta}),$$

where $\tilde{\theta}$ is a vector which is close to $\hat{\theta}$. The formula (9) uses this approximation iteratively. During the Newton-Raphson iteration, we use $H_0(\theta)$ instead of $H(\theta)$ for $m < M$ due to the fact that $H(\theta)$ may not be positive definite at early stages, where $H_0(\theta)$ is the matrix whose (λ, λ')-th element is given by

$$\sum_{i=1}^{n} f(t_i;\theta) B_\lambda(t_i) B_{\lambda'}(t_i).$$

The heuristic for this usage of H_0 instead of H at early stages is as follows. However, if $y_i \approx f(t_i;\theta^m)$ for some m, then $H(\theta^m)$ is approximately equal to $H_0(\theta^m)$. Since for any nonzero vector θ_0,

$$\theta_0^T H_0(\theta)\theta_0 = \sum_{i=1}^{n} s(t_i;\theta_0)^2 f(t_i;\theta) > 0,$$

the matrix $H_0(\theta)$ is positive definite for each θ.

Now we employ the step-halving, in which θ^{m+1} is determined from θ^m according to the formula

$$\theta^{m+1} = \theta^m + 2^{-r} H^{-1}(\theta^m) S(\theta^m),$$

where r is the smallest nonnegative integer such that

$$\Delta(\theta^m + 2^{-r} H^{-1}(\theta^m) S(\theta^m)) < \Delta(\theta^m).$$

We stop the iteration when $|\Delta(\theta^m) - \Delta(\theta^{m+1})| < \tau$ for a sufficiently small number τ.

III. Experimental Results

To examine the performance of our denoiser $\widehat{PD}$ and to compare it with the wavelet shrinkage denoiser $\widehat{WS}$, we carried out a small simulation study. The main purpose of this simulation is to show how much improvement can be obtained by imposing positivity when we use the same set of wavelet basis functions. As we pointed out above, our positive denoiser $\widehat{PD}$ is anticipated to perform better than the wavelet shrinkage if the sought-after signal has several regions where it is close to zero.

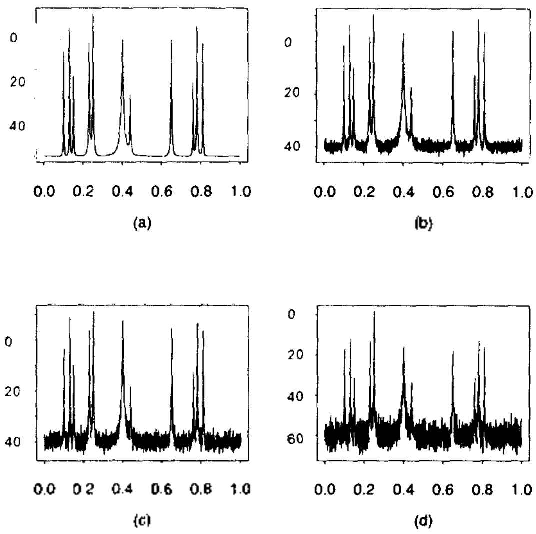


Figure 2. A Bumps function and noisy signals. (a) Plot of $(t_i, f(t_i))$, $i=1,\ldots,n$.
Plots of (t_i, y_i), $i=1,\ldots,n$: (b) when SNR=6, (c) when SNR=4, and (d) when SNR=2, where the x-axis denotes t.

In this simulation the noisy signal is in the form of the equation (2) where the sample size selected is $n=2048$; we choose $M=7$ and $\tau=10^{-10}$. We use the following Bumps function:

$$f(t) = \sum_{j=1}^{11} h_j K((t-t_j)/w_j),$$

where $K(t)=(1+|t|)^{-4}$, $(t_j)=(.1, .13, .15, .23, .15, .40, .44, .65, .76, .78, .81)$, $(h_j)=(4, 5, 3, 4, 5, 4.2, 2.1, 4.3, 3.1, 5.1, 4.2)$, and $(w_j)=(0.005, 0.005, 0.006, 0.01, 0.01, 0.03, 0.01, 0.01, 0.005, 0.008, 0.005)$. The Bumps signal was used in [3]; this signal has nonnegative values and has several peaks. Fig. 2(a) shows the shape of this Bumps function.

The Bumps signal f is discretized to n equally spaced points in the interval [0, 1], in such a way that $t_i = i/n$, $i=1,\ldots,n$ and a white Gaussian noise is added to the signal f. Let $sd(f)$ be the standard deviation of the n numbers $\{f(t_i): i=1,\ldots,n\}$. The statistical signal-to-noise ratio is defined by

$$\text{S}-\text{SNR} = \frac{sd(f)}{\sigma}.$$

We took the three values of σ determined by $S-SNR=2,4,6$. In Fig. 2(b), we present three noisy signals corresponding to three different σ's determined such that $S-SNR=2,4,$ and 6.

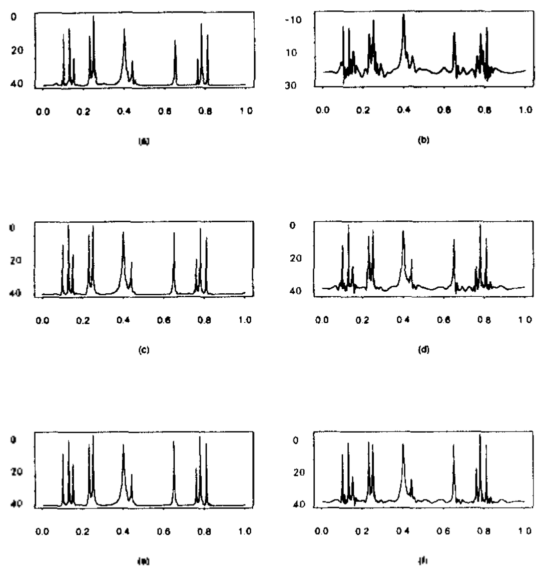


Figure 3. (a), (c) and (e) display $(t_i, \widehat{\mathrm{PD}}(t_i))$, $i=1,\ldots,n$, and (b), (d) and (f) show $(t_i, \widehat{\mathrm{WS}}(t_i))$, $i=1,\ldots,n$. S-SNR of (a) and (b) is 2, S-SNR of (c) and (d) is 4 and S-SNR of (e) and (f) is 6.

Fig. 3 shows the denoised signals $\widehat{\mathrm{PD}}$'s and $\widehat{\mathrm{WS}}$'s. Here we use ϕ and ψ with support [0, 7] which were discovered by Daubechies [1]. To compute them at a point $t\in[0,1]$, we first compute ϕ and ψ at some points by *Matlab* and use a piecewise constant interpolation to compute them at t. We use the hard thresholding rule (3) to select the important wavelet coefficients. To apply $\eta_H(w,\delta)$ the noise level σ is estimated as $\hat{\sigma}$, the median absolute deviation of the wavelet coefficients at the finest level J, divided by 0.6745 [3]. The low-resolution cutoff is $j_0=1$ and the finest level is $J=10$. The threshold is chosen by $2\sqrt{(\log n)/n}\,\hat{\sigma}$.

We can observe that $\widehat{\mathrm{PD}}$ performs better than $\widehat{\mathrm{WS}}$ when the Bumps signal f is near zero; $\widehat{\mathrm{WS}}$ appears to be more wiggly than $\widehat{\mathrm{PD}}$. Furthermore, it can be seen that $\widehat{\mathrm{PD}}$ captures the peaks better than $\widehat{\mathrm{WS}}$. As S-SNR increases, the shape of each denoiser becomes closer to that of the true Bumps function.

To carry out an object test of the behavior of $\widehat{\mathrm{PD}}$ with $\widehat{\mathrm{WS}}$, we need criteria for the performance of estimators. If $\hat{f}(t_i)$ is the denoised function value at t_i and

Table 1. This table shows RMSE, MAD, and MXDV of $\widehat{\mathrm{PD}}$ and $\widehat{\mathrm{WS}}$ as S-SNR changes.

S-SNR	RMSE		MAD		MXDV	
	PD	WS	PD	WS	PD	WS
2	1.36	2.96	0.67	1.87	11.15	23.56
4	0.67	1.80	0.34	1.18	8.96	14.46
6	0.50	1.42	0.28	0.94	5.61	13.83

n the sample size, then one can use $\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{f}(t_i)-f(t_i))^2}$. However, an estimate can have low RMSE but appear noisy to the eye, especially near the peak points. To capture this aspect we also used the mean absolute deviation (MAD) and the maximum deviation (MXDV) criterion, where $\mathrm{MAD}=\frac{1}{n}\sum_i|\hat{f}(t_i)-f(t_i)|$ and $\mathrm{MXDV}=\max_i|\hat{f}(t_i)-f(t_i)|$. As in Fig. 3, our denoiser $\widehat{\mathrm{PD}}$ performs better than $\widehat{\mathrm{WS}}$. Note that the peaks are captured better by $\widehat{\mathrm{PD}}$ than by $\widehat{\mathrm{WS}}$ on the basis of the MXDV criterion. These criteria have been used in [5] and [6].

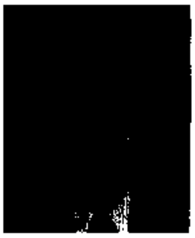
IV. Conclusion

In this paper we propose a method of denoising a noisy signal using positivity of the signal. Simulated data show that the proposed method performs better than the wavelet shrinkage method under the positivity constraint. It would be worth extending our method to two dimensional image. Since an image is positive, we anticipate that our method for two dimensional image will provide a better algorithm than the usual denoising algorithms.

Reference

1. I. Daubechies, *Ten Lectures on Wavelets*. SIAM. Philadelphia, PA, 1992.
2. M. Vetterli and J. Kovavcevic, *Wavelets and Subband Coding*. Prentice Hall, Inc. Englewood Cliffs, NJ, 1995.
3. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika* 81, pp. 425-455, 1994.
4. P. A. Jansson, *Deconvolution of images and spectra*, Academic Press, San Diego, 1997.
5. L. Breiman and S. Peters, "Comparing automatic smoothers (a public servc enterprise)," *Int. Statist. Rev.* 60, pp. 271-290, 1992.
6. J. -Y. Koo, "Spline estimation of discontinuous regression functions," *J. Comput. Graphical Statist.* 6, pp. 266-284, 1997.

▲Ja-Yong Koo



Ja-Yong Koo was born in Seoul, Korea in 1961. He received the B.S. degree from Seoul National University, Seoul, Korea in 1984, in Computer Science and Statistics, and M.S. and Ph.D. degrees from University of California, Berkeley, CA, USA, all in Statistics in 1986 and 1988, respectively.

From 1988 to 1995, he was an Assistant professor with the department of Statistics, Hallym University, Chunchon, Korea. He is currently an Associate Professor at the department of Statistics, Hallym University, Chunchon, Korea. His research interests are in function estimation and image processing based on wavelet.

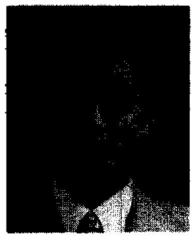
▲Hyun-Suk Park

Hyun-Suk Park received the B.S. and M.S. degree in statistics from Hallym University in 1992 and 1994, respectively. She is currently working toward the Ph.D. degree at Hallym University. She is working on PET image reconstruction based on Wavelet Transform. Her research interests include function estimation and image processing using Wavelet Transform.

▲Kuk Nam Park

Kuknam Park received the B.S. degree in electronic engineering from Hallym University in 1997. He is currently working toward the M.S. degree at Hallym University. He works on the signal denoising and the image compression using Wavelet Transform. His research interests include image enhancement, wavelet filter design for image processing, and image compression based on Wavelet Transform.

▲Ji-sang Yoo



Jisang Yoo was born in Seoul, Korea in 1962. He received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea in 1985 and 1987, all in electronics engineering, and Ph.D. degree from Purdue University, West Lafayette, IN, USA, in electrical engineering in 1993, respectively.

From September 1993 to August 1994, he worked as a senior research engineer in industrial electronics R&D center at Hyundai Electronics Industries Co., Ltd, ichon, Korea, in the area of image compression and HDTV. From 1994 to 1997, he was an Assistant professor with the department of Electronics Engineering, Hallym University, Chuncheon, Korea. He is currently an Assistant Professor with the department of electronics engineering, Kwangwoon University, Seoul, Korea. His research interests are in signal and image processing, nonlinear digital filtering, and computer vision.