A New Wavelet Denoising Method based on MDL [1)]

Gustavo Fernández, Reinhard Beichel, Horst Bischof, Franz Leberl
{gustavo, beichel, bischof, leberl}@icg.tu-graz.ac.at
Institute for Computer Graphics and Vision, TU-Graz
Infeldgasse 16 II, A-8010, Graz, Austria

Abstract:
In this paper we present a new image denoising approach based on wavelet thresholding and the Minimum Description Length (MDL) principle. Considering the probability distribution of the wavelet coefficients, we build a function based on the principles of MDL. Looking for a minimum of this function, our model allows us to automatically choose the necessary threshold value for image denoising in the wavelet domain. The method has been tested on several different images including Computer Tomography data. The results show the effectiveness of the method to different kinds of noise and different levels of it. Our method is compared with previous wavelet denoising approaches.

Keywords: *Image Denoising, Wavelet thresholding, Minimum Description Length, Filtering, Computer Tomography.*

1 Introduction

The image restoration process we consider in this work consists in the removal of noise that degrades an image. Since Donoho and Johnstone [2], [5], [6] proposed the use of wavelet thresholding for denoising 1-dimensional signals obtained with additive, white noise, different approaches and methods have been developed [1], [8], [10], [16]. Cherkassky et al. applied wavelet thresholding using Vapnik-Chervonenkis (VC) theory to select the wavelet coefficients on 1D signals and images [8], [17], [18]. This approach has the advantage that it is based on Statistical Learning Theory (SLT) [19], [20], which provides a theoretical framework for function estimation from finite samples. Recently approaches based on Minimum Description Length (MDL) principle have been developed [16], [10]. These methods are focused on wavelet denoising and efficient compression of the images. In this paper we compare several approaches and propose a novel MDL formulation with the goal of denoising. Under the MDL framework, we derive a function based on the distribution of the wavelet coefficients. The minimization of

1) This work was supported by the FWF under project number P-14897.

this function provides the coefficients of the denoised image. This paper is organized as follows: Section 2 shows the main ideas of wavelet denoising methods. In Section 3 we introduce our model. Section 4 presents some results using our method and a comparison to other wavelet denoising methods. Conclusions and discussion are presented in Section 5.

2 Wavelet denoising

In signal processing, a popular approach for approximating a function f is to estimate it as a linear combination of basis functions g_k. Popular basis functions g_k used are orthonormal basis functions as Fourier series and more recently wavelets [3], [4], [8], [10], [11], [12], [16]. Wavelets have received more attention in recent years due to excellent properties of decomposition on frequency and scale domain simultaneously which allow us to treat non-stationary signals. In recent years, several authors proposed wavelet-based methods for signal denoising and compression [1], [8], [10], [12], [16]. Basically, these methods consist of following steps:

1) Obtain the wavelet decomposition of the image,

2) select a threshold value using a specified criteria,

3) perform the inverse wavelet transformation to obtain the denoised image.

The main difference between the different approaches is the selection criteria in step 2. Recently models based on the Minimum Description Length (MDL) principle were applied to the problem of wavelet denoising. The basic idea of the MDL-criterion is to choose the model that gives the shortest description of the data.

3 Minimum Description Length (MDL) and our Method

3.1 MDL Approach

The principle of Minimum Description Length (MDL) was introduced by Rissanen [14]. This criterion has been succesfully used for wavelet denoising by DeVore et al. [1], Moulin [12], Moulin and Liu [13], Saito [16], and Hansen and Yu [10]. MDL has been used successfully in other areas such as cluster analysis by Wallace [21] and image segmentation by Leclerc [9]. The MDL criterion suggests to choose the model that gives the shortest description of the data among a given collection of models. For each model in the collection, the data can be related with codelength of encoding the data and transmit them [15]. Saito presented an algorithm for suppresing the noise component and compressing the signal component [16]. Hansen and Yu derive a criteria based on a Laplacian model for noiseless wavelet coefficients [10].

3.2 Our model

In order to use MDL for wavelet denoising we need to define a coding scheme. We use the following assumptions:
1) The coefficients describing the true signal are modelled by a Laplacian distribution (similar to [10])

$$L(w) = \frac{\lambda}{2} e^{-\lambda|w|}, \qquad w \in \Re. \tag{1}$$

2) The coefficients associated with the noise component of the signal are an i.i.d. sample from a Gaussian distribution as in [10]

$$G(w) = \frac{e^{-\frac{w-\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}, \qquad w \in \Re. \tag{2}$$

Using this distribution we need to specify for each coefficient where it comes from. This can be done by a binary mask which can be coded by

$$C_L(W_n) = \log(n), \tag{3}$$

bits, where n is the number of wavelet coefficients.
Therefore we model the cost of description the whole image by

$$C_{MDL}(w) = C_L(W_r) + C_G(W_s) + C_L(W_n). \tag{4}$$

where $C_L(W_r)$ is the cost of transmission of the r coefficients coded by the Laplacian distribution, $C_G(W_s)$ is the cost of transmission of the s coefficients coded by the Gaussian distribution ($n = r + s$). Using the optimal Shannon code, (4) can be expressed as

$$C_{MDL}(W) = \log\left(\frac{\lambda_r}{2} e^{-\lambda_r|W|}\right) + \log\left(\frac{e^{-\frac{|W-\mu|}{\sigma_s^2}}}{\sigma_s\sqrt{2\pi}}\right) + \log(n). \tag{5}$$

After further simplifications, this yields following the function to minimize

$$\arg\min_W [LGMDLWD(W)] = \arg\min_W [-\log(\lambda_r \sigma_s n)], \tag{6}$$

where λ_r is estimated under (1), σ_s is estimated using (2) and n is the total number of wavelet coefficients. To minimize this function the coefficients are ordered by

$$\left|\frac{w_1}{freq_1}\right| > \left|\frac{w_2}{freq_2}\right| > \ldots > \left|\frac{w_m}{freq_m}\right| > \ldots \tag{7}$$

This order was suggested by Cherkassky et al [8]. The idea of this order is to penalize higher-frequency wavelets. The smooth subband LL(m) (where LL is the Low-pass component of the wavelet decomposition and m indicates the highest level of decomposition) is not included in this order because these coefficients are always kept. These coefficients correspond to the parts of the image with the lowest frequencies (i.e. smooth areas), and should therefore be kept. The first r coefficients according to the order (7) correspond to the wavelet coefficients of the original signal and the remaining s coefficients ($s = n - r$) correspond to the noise component of the image. According to MDL we have to find r such that we minimize (6).

Figure 1 shows functions $LGMDLWD(W)$ of the images used in Section 4. Once the minimum of the function $LGMDLWD(W)$ has been found, the denoising procedure keeps the r most important wavelet coefficients of the order (7) ($LGMDLWD(r) = m$), sets the other s ($s = n - r$) coefficients to zero, and reconstructs the image using the inverse discrete wavelet transform.

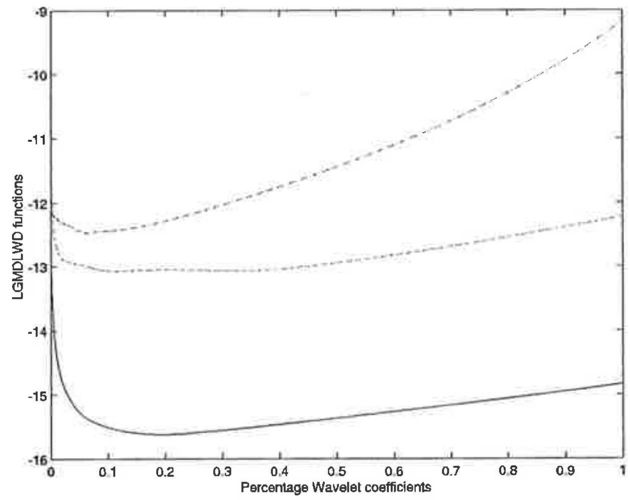

Figure 1: LGMDLWD functions for 'Lenna' (Solid line), 'Baboon' (Dash line) and 'CT' images (Dash-dot line). LGMDLWD function vs. Wavelet coefficients.

4 Results

We have tested our algorithm on different kinds of images with different types of noise at various levels. The images used were standard images such as 'Lenna' and 'Baboon', and a set of computer tomography image (named 'CT'). In the case of 'Lenna' and 'Baboon' we added additive, white Gaussian noise and multiplicative speckle noise. In both cases different noise levels were added. On the 'CT' image we did not add any kind of noise because the image contains it. Due to space reasons we include here only some results (For more of them, see [7]). All the images used have a size equal of 512×512 pixels. We compare our method against previous wavelet denoising methods. The methods used as benchmarks were:

- The SURE method proposed by Donoho and Johnstone [2] with hard thresholding (Here it will be referred as SUREHTWD).

- Cherkassky et al. [8] used the Vapnik-Chervonenkis theory (VC-theory) to select the correct threshold (We call this method CVCWD).

- Hansen and Yu applied the MDL criteria on natural images for denoising and compressing them [10] (Here it will be named HMDLWD).

In case of 'Lenna' and 'Baboon', we use the signal to noise ratio (SNR) of denoised image as a measure of the quality of restoration. Figure 2 shows the result of our method applied on 'Lenna' for different levels of additive white Gaussian noise. Figure 3 shows the result

(a) (b) (c) (d)

Figure 2: Results on 'Lenna'. Wavelet decomposition level = 9. (a) Noisy image - Gaussian noise $\sigma = 0.5$, (b) Denoised image, (c) Noisy image - Gaussian noise $\sigma = 5.0$, (d) Denoised image.

using the different methods on 'Baboon' contaminated with Gaussian noise ($\sigma = 0.5$). Figure

(a) Noisy image. (b) Denoised image. Method used: LGMDLWD. (c) Denoised image. Method used: SUREHTWD. (d) Denoised image. Method used: CVCWD.

Figure 3: Results on 'Baboon'. (a) Noisy image (Gaussian noise $\sigma = 0.5$), (b), (c), (d), Denoised images using 3 levels of wavelet decomposition.

4 shows the result of our method applied on computer tomography image for different levels of wavelet decomposition. In this case, result of CVCWD method is much worse than the results presented here, and the result of SUREHTWD is similar to the method of Hansen. Table 1 shows the SNR and the percent of wavelet coeffificients used in the restoration process,

(a) Original image.

(b) Denoised image. Decomposition level = 3. Method used: HMDLWD.

(c) Denoised image. Decomposition level = 9. Method used: LGMDLWD.

(d) Denoised image. Decomposition level = 3. Method used: LGMDLWD.

Figure 4: CT image. Original image (a) and denoised images using Hansen method with wavelet level decomposition equal to 3 (image (b)) and our method with 2 different wavelet level decomposition (images (c), and (d)).

obtained on 'Lenna' with different levels of additive, white Gaussian noise across the different levels of wavelet decomposition. Table 2 shows the SNR obtained on 'Lenna' and 'Baboon' using different wavelet denoising methods. The wavelet basis used in all the restoration processes was '*biorthogonal 6.8*' [3].

Analyzing the SNR obtained from the highest decomposition level (level 9) to the level 3, it can be seen that the SNR has the same order for a fixed level of noise added, e.g. with noise level $\sigma = 5.0$ the SNR varies in the interval $[22.28, 24.18]$, which indicates that our method works well independently of the wavelet decomposition level. Other wavelet denoising

	($\sigma = 0.5$)	($\sigma = 0.5$)	($\sigma = 5.0$)	($\sigma = 5.0$)
Decomposition level	SNR (db)	% WC	SNR (db)	% WC
1	19.28	27.55	10.42	27.09
2	22.28	11.05	14.48	9.95
3	23.84	9.65	17.60	7.53
4	24.09	9.52	18.43	7.32
5	24.18	11.14	18.48	7.31
6	24.14	12.17	18.32	7.96
7	24.14	12.17	18.21	8.36
8	24.13	12.34	18.19	8.36
9	24.13	12.34	18.19	8.36

Table 1: SNR and percent of wavelet coefficients used on the restoration. 'Lenna' with additive Gaussian noise. Image size = 512×512 pixels.

Method	'Lenna' ($\sigma = 0.5$)	'Lenna' ($\sigma = 5.0$)	'Baboon' ($\sigma = 0.5$)	'Baboon' ($\sigma = 5.0$)
SUREHTWD	22.44	19.85	18.08	15.22
CVCWD	23.91	19.58	19.30	10.74
HMDLWD	24.57	18.04	18.78	14.16
LGMDLWD	24.13	18.18	16.47	13.50

Table 2: SNR obtained using different methods. 'Lenna' and 'Baboon' with different levels of additive Gaussian noise and image size = 512×512 pixels.

methods, like the one proposed by Hansen et al. [10] do not show this behaviour, i.e., the denoising results depend strongly on the wavelet decomposition level. In general one can see that our method performs similar to other methods but unlike other methods all parameters are obtained automatically.

5 Conclusions

We introduced a novel method for denoising images based on Wavelet thresholding and Minimum Description Length. We show that our method is effective to denoise different types of images with different levels of noise introduced on them. Regarding the quality of the restoration we want to note that our method performs well, in terms of SNR, accross the different images, and levels of noise. Our method performs similar in terms of SNR levels as other methods, but our approximation needs less parameters to tune. The results of our method when it is applied to CT images are promising. The method has eliminated much of the noise present in smooth regions, while the borders have been preserved.

6 Acknowledgments

The authors would like to thank Dr. V. Cherkassky and Dr. M. Hansen for the code of their algorithms and the exchange of ideas.

References

[1] R. A. DeVore N. Lee A, Chambolle and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Processing*, Vol. 7:319–335, 1998.

[2] I. M. Johnstone D. Donoho. Ideal spatial adaptation via wavelet shrinkage. 81:425–455, 1994.

[3] I. Daubechies. *Ten Lectures on Wavelets.* 1992.

[4] R. DeVore and B. J. Lucier. Fast wavelet techniques for near-optimal image processing. *IEEE Military Communications Conference Record*, pages 1129–1135, 1992.

[5] D. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, pages 613–627, 1995.

[6] D. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, Vol. 90, no. 432:1200–1224, 1995.

[7] H. Bischof G. Fernandez, R. Beichel and F. Leberl. Comparison of image wavelet denoising methods. *Technical Report - ICG, TU-Graz*, In preparing, 2002.

[8] V. Cherkassky J. Shao. Improved VC-based signal denoising. *IEEE Trans. Inform. Theory*, pages 2439–2444, 2001.

[9] Y. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. Jou. Comp. Vis.*, Vol. 3:73–102, 1989.

[10] B. Yu M. Hansen. Wavelet thresholding via MDL for natural images. Technical report, 2000.

[11] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI*, Vol. 11:674–692, 1989.

[12] P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Inf. Theory*, Vol. 3:909–919, 1994.

[13] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors. *IEEE Trans. Information Theory*, Vol. 45:909–919, 1999.

[14] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Stat.*, Vol. 11:416–431, 1983.

[15] J. Rissanen. Universal coding, information, prediction and estimation. *IEEE Trans. Inf. Theory*, 1984.

[16] N. Saito. Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion. 1994.

[17] X. Shao V. Cherkassky. Model selection for wavelet-based signal estimation. *Proc. IJCNN-98*, 1998.

[18] X. Shao V. Cherkassky. Signal estimation and denoising using VC-theory. *Neural Networks*, 14, 2001.

[19] V. Vapnik. *Statistical Learning Theory and its applications.* 1982.

[20] V. Vapnik. *The Nature of Statistical Learning Theory.* 1995.

[21] R. Wallace. *Finding natural clusters through entropy minimization.* PhD thesis, 1989.