14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A GEOMETRICAL WAVELET SHRINKAGE APPROACH FOR IMAGE DENOISING

Bruno Huysmans, Aleksandra Pižurica and Wilfried Philips

TELIN Department, Ghent University
Sint-Pietersnieuwstraat 41, 9000, Ghent, Belgium
phone: + (32)92647966, fax: + (32)92644295 , email: bruno.huysmans@telin.ugent.be
web: http://telin.ugent.be/ipi

ABSTRACT

In this paper a denoising technique for digital gray value images corrupted with additive Gaussian noise is presented. We studied a recently proposed hard thresholding technique which uses a two stage selection procedure in which coefficients are selected based on their magnitude, spatial connectedness and interscale dependencies. We construct a shrinkage version of the algorithm which outperforms the original one. We also present a new hard thresholding algorithm which incorporates the spatial connectivity information in a more simple and efficient way and construct a shrinkage version of it. The new algorithms are faster and lead to better denoising performances compared to the original one, both visually and quantitatively.

1. INTRODUCTION

Due to its energy compaction property, the wavelet transform is a practical tool for image denoising. Denoising is usually done by shrinking the wavelet coefficients: coefficients that contain primarily noise should be reduced to negligable values, while the ones containing a significant noise-free component should be reduced less. In early methods only the coefficient magnitude was used to predict whether a coefficient represents useful signal or mainly noise [2, 11]. More recent methods also include intra- and interscale coefficient dependencies [4].

Another kind of information that can be integrated in the denoising scheme is *geometrical* information, where one assumes that image features (lines, edges, ...) show a certain directional continuity. The methods of [7] , [6] and [10] combine the intra- and interscale dependencies with a bilevel Markov Random Field (MRF) model, which encodes the prior knowledge about the spatial clustering of wavelet coefficients, i.e., which encodes the 'geometrical properties' of image details. We recently proposed another technique which exploits geometrical information [5]: first the wavelet coefficients belonging to image features are detected, then an averaging step is performed. Although the results are good, the method has many (image dependent) thresholds which have to be chosen. The method also fails at high noise levels.

In this paper, we statistically analyse a recently proposed feature-based hard thresholding algorithm [1]: only those coefficients with a high magnitude and a sufficiently large spatial support (the number of large coefficients connected to them) are retained. In [1] the authors claim that "wavelet shrinkage methods which either select or reject wavelet coefficients (i.e. hard thresholding algorithms) are statistically better than the probabilistic methods". However, in this paper we present a shrinkage version of their algorithm and prove that it results in a better denoising performance. We also analysed the optimality of their coefficient classification in terms of its closeness to the ideal, oracle thresholding. Based on our findings we propose a new, better measure to describe the spatial surrounding and we integrate the measure into a hard thresholding and a shrinkage denoising scheme. Both schemes outperform the original denoising method, both visually as in terms of PSNR. We name the proposed methods 'geometrical' because they employ the geometry of wavelet coefficient clusters.

The method from [1] is explained in Section 2. In Section 3 we describe a novel rule for describing the spatial surrounding of the wavelet coefficients. This rule is also integrated into a hard thresholding framework. In Section 4 we construct shrinkage versions of both the original algorithm from [1] and our new hard thresholding algorithm. Section 5 describes the experiments we have done and in Section 6 a conclusion is given.

2. FEATURE BASED HARD THRESHOLDING

First we will explain the notation used in the rest of the paper. We restrict ourself to images corrupted with additive white Gaussian noise. Due to the linearity of the wavelet transform the additive noise model in the image domain remains additive in the transform domain as well:

$$w_{k,d}(x,y) = y_{k,d}(x,y) + n_{k,d}(x,y) \tag{1}$$

In this expression $w_{k,d}(x,y)$ and $y_{k,d}(x,y)$ are the noisy, resp. noisefree, wavelet coefficients of scale k and orientation d and $n_{k,d}(x,y)$ is the noise component.

The method from [1] uses a redundant wavelet transform with the Haar wavelet and five resolution scales. A first threshold τ is used to determine a binary label $I_{k,d}(x,y)$ for each coefficient:

$$I_{k,d}(x,y) = \begin{cases} 1, & \text{if } |w_{k,d}(x,y)| > \tau \\ 0, & \text{else} \end{cases} \tag{2}$$

The coefficients with $I_k(x,y) = 1$ are those with a sufficiently large magnitude and are called the *valid* coefficients. Now for each valid coefficient the support value $S_{k,d}(x,y)$ is calculated: $S_{k,d}(x,y)$ is the total number of valid coefficients which are spatially connected to the valid coefficient, see Fig. 1. $S_{k,d}(x,y)$ is used to refine the original binary map

THIS RESEARCH WAS FINANCED WITH A SPECIALIZATION SCHOLARSHIP OF THE "FLEMISH INSTITUTE FOR THE PROMOTION OF INNOVATION THROUGH SCIENCE AND TECHNOLOGY IN FLANDERS (IWT-VLAANDEREN)". A. PIŽURICA IS A POSTDOCTORAL RESEARCH FELLOW OF FWO FLANDERS, BELGIUM.

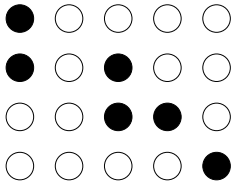


Figure 1: The white coefficients are the invalid ones. The support value for the two black coefficients on the left side is 2, for the four other black coefficients the support size is 4.

$I_{k,d}(x,y)$ to a new map $J_{k,d}(x,y)$:

$$J_{k,d}(x,y) = \begin{cases} 1, & \text{if } S_{k,d}(x,y) > s, \\ & \text{or } J_{k+1,d}(x,y) I_{k,d}(x,y) = 1 \\ 0, & \text{else} \end{cases} \quad (3)$$

The parameter s is the required number of support coefficients needed for selection. Coefficients with $S_{k,d}(x,y) > s$ are called *spatially supported*, coefficients with $J_{k+1,d}(x,y) = 1$ are called *supported by scale*. $J_{k,d}(x,y)$ is equal to one when there exist enough wavelet coefficients of large magnitude around the current coefficient, i.e., when the coefficient is likely to belong to a geometrical feature, like an edge or a line. However, it is also one when the magnitude of the coefficient is large ($I_{k,d}(x,y) = 1$) and the coefficient on the same position but a coarser scale is large and locally supported ($J_{k+1,d}(x,y) = 1$). Thus the method selects coefficients which are sufficiently large and locally supported as well as isolated coefficients which are sufficiently large and supported by scale. $J_{k,d}(x,y)$ is calculated recursively, starting from the coarsest resolution scale. Good choices for the parameters τ and s are [1]:

$$\tau(\sigma_n) = 2.37\sigma_n - 2.30 \quad (4)$$

$$s(\sigma_n) = \lfloor 0.24\sigma_n + 4.21 \rfloor \quad (5)$$

with σ_n the standard deviation of the noise, which is estimated with the method from [3]. Those threshold values were obtained experimentally by evaluating their performance on different test images. The estimation $\widehat{y}_{k,d}(x,y)$ of the noise free coefficient $y_{k,d}(x,y)$ becomes:

$$\widehat{y}_{k,d}(x,y) = \begin{cases} w_{k,d}(x,y), & \text{if } J_{k,d}(x,y) = 1 \\ 0, & \text{if } J_{k,d}(x,y) = 0 \end{cases} \quad (6)$$

3. NEW COEFFICIENT SELECTION PROCEDURE

We statistically analysed the selection criterium (3). In order to do this we defined an optimal binary mask $J^{opt}_{k,d}(x,y)$ for each detail image. This optimal mask uses information from the noise-free wavelet coefficients. Recall that the ideal coefficient selection in terms of mean squared error is [8]:

$$J^{opt}_{k,d}(x,y) = \begin{cases} 1, & \text{if } \mid y_{k,d}(x,y) \mid \geq \sigma_n \\ 0, & \text{else} \end{cases} \quad (7)$$

We analyse statistically the classification rule (3) by comparing the resulting masks with the optimal classification from (7). The coefficients $J_{k,d}(x,y)$ for which $J^{opt}_{k,d}(x,y) = 1$ consist of true positives (TP's) and false negatives (FN's):

$$TP_{k,d}(x,y) = \begin{cases} 1, & \text{if } (J^{opt}_{k,d}(x,y) = 1) \text{ AND } (J_{k,d}(x,y) = 1) \\ 0, & \text{else} \end{cases} \quad (8)$$

$$FN_{k,d}(x,y) = \begin{cases} 1, & \text{if } (J^{opt}_{k,d}(x,y) = 1) \text{ AND } (J_{k,d}(x,y) = 0) \\ 0, & \text{else} \end{cases} \quad (9)$$

The denoising result can be improved when we succeed in reducing the number of FN's, while retaining the TP's.

3.1 Support value

The method from [1] uses the magnitude of $w_{k,d}(x,y)$ and the support value $S_{k,d}(x,y)$ to distinguish between noise coefficients on the one hand and useful coefficients on the other hand. However, this approach has two disadvantages:

A first disadvantage is that the calculation of the support value $S_{k,d}(x,y)$ is a time consuming operation. A second drawback is that the threshold values $\tau(\sigma_n)$ on the magnitude $w_{k,d}(x,y)$ are high, see (4). Due to this a lot of useful edge coefficients are falsely removed from the mask. Fig. 2 shows the number of TP's and FN's obtained with the classification rule (3). We present results for the two finest resolution scales of the Lena test image. On Fig. 2 we see that there are more FN's than TP's, which means that less than 50% of the coefficients for which $J^{opt}_{k,d}(x,y) = 1$ are detected. We can conclude that the coefficient selection method is mainly focused on avoiding false positives and because of that excludes too many real edge coefficients.

3.2 Local mean coefficient magnitude

To overcome the abovementioned drawbacks of the coefficient selection procedure from [1] we analysed some other significance measures. Instead of calculating the support value $S_{k,d}(x,y)$, we did some experiments with counting the number of valid coefficients $I_k(x,y) = 1$ inside a small window around each coefficient. However, the best result was obtained by replacing the support value $S_{k,d}(x,y)$ by the mean coefficient magnitude $m_{k,d}(x,y)$ calculated in a local window centered around each coefficient:

$$m_{k,d}(x,y) = \frac{1}{(2W+1)^2} \sum_{i=x-W}^{x+W} \sum_{j=y-W}^{y+W} \left| w_{k,d}(i,j) \right| \quad (10)$$

with W the half window size. Using this measure has the advantage that the preliminary classfication step (2) can be omitted. This measure was also used in [9], where it was integrated in a Bayesian statistical framework. Our new proposed hard thresholding algorithm becomes:

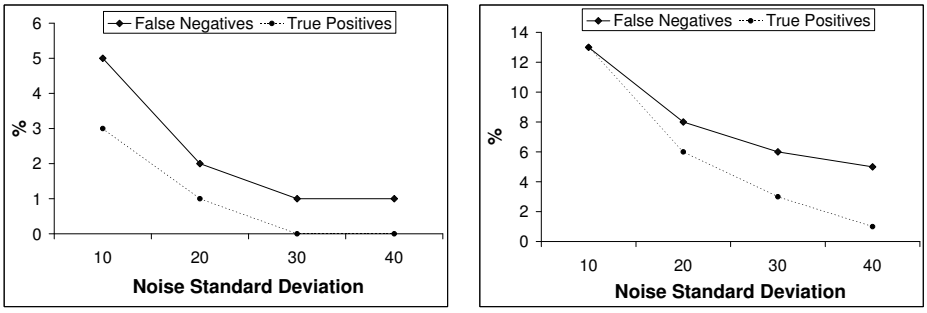
$$\widehat{y}_{k,d}(x,y) = \begin{cases} w_{k,d}(x,y), & \text{if } m_{k,d}(x,y) > \tau_m \\ 0, & \text{if } m_{k,d}(x,y) \leq \tau_m \end{cases} \quad (11)$$

The optimal value for τ_m was experimentally determined as:

$$\tau_m(\sigma_n) = 1.5\sigma_n \quad (12)$$

4. FROM HARD THRESHOLDING TO SHRINKING

In [1] the authors claim that hard thresholding methods should perform better than (probabilistic) shrinkage methods. In this section we convert their original hard thresholding algorithm (described in Section 2) and our proposed hard



(a) Scale 1 (b) Scale 2

Figure 2: The percentage of TP's and FN's for the first two scales of the Lena image, obtained with the coefficient selection method from (3).

thresholding algorithm (described in Section 3.2) to *shrinkage* based approaches. For the shrinkage functions we use simple linear models. In the next section we show that significant improvements are obtained by going to a shrinkage approach, even with our simple linear models.

4.1 Shrinkage version of the original algorithm

We adapted the original hard thresholding algorithm in two ways. A first modification aims at obtaining better masks. The preliminary classification $I_{k,d}(x,y)$, see (2), is replaced by $\widetilde{I}_{k,d}(x,y)$:

$$\widetilde{I}_{k,d}(x,y) = \begin{cases} 1, & \text{if } |w_{k,d}(x,y)| > \tau_2 \text{ AND } |w_{k+1,d}(x,y)| > \tau_3 \\ 0, & \text{else} \end{cases} \tag{13}$$

with a lower threshold τ_2 on the coefficient magnitude ($\tau_2 < \tau$) and an additional interscale criterion. The lower value for τ_2 aims at reducing the number of FN's, while the extra interscale criterion removes FP's from the mask. Good choices for τ_2 and τ_3 (experimentally determined) are:

$$\tau_2 = 2\sigma_n\,,\ \tau_3 = \sigma_n \tag{14}$$

Instead of the hard threshold imposed on the spatial support, we attach a shrinkage function to it. The form of the shrinkage function is given in Fig. 3. The value of s_{shrink} was experimentally determined as:

$$s_{shrink}(\sigma_n) = \lfloor 0.24\sigma_n + 4.21 \rfloor \tag{15}$$

In order to estimate the noise free coefficient value $\widehat{y}_{k,d}(x,y)$, each noisy coefficient $w_{k,d}(x,y)$ is multiplied by the shrinkage factor obtained for that coefficient.

4.2 Shrinkage version of the proposed algorithm

We have also developed a shrinkage version of our hard thresholding algorithm proposed in section 3.2. The shrinkage factor for each coefficient depends on the mean coefficient magnitude $m_{k,d}(x,y)$ calculated around that coefficient. The form of the shrinkage function is given in Fig. 4. The values of $m_{1,shrink}$ and $m_{2,shrink}$ are experimentally determined as:

$$m_{1,shrink}(\sigma_n) = \sigma_n\,,\ m_{2,shrink}(\sigma_n) = 2\sigma_n \tag{16}$$

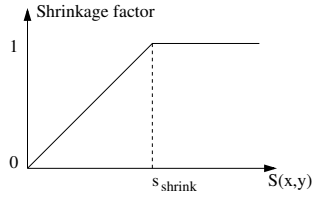


Figure 3: Proposed shrinkage function for the waveletcoefficients. The shrinkage function is based on the support value $S_{k,d}(x,y)$.

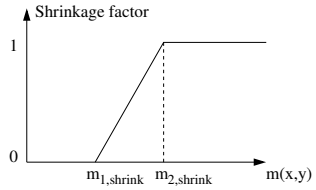


Figure 4: Proposed shrinkage function for the waveletcoefficients. The shrinkage function is based on the mean value $m_{k,d}(x,y)$ of the coefficients in a small local window around each coefficient.

Again, the noise free coefficient value $y_{k,d}(x,y)$ is estimated by multiplying each noisy coefficient $w_{k,d}(x,y)$ with the shrinkage factor obtained for that coefficient.

5. RESULTS AND DISCUSSION

Table 1 lists peak signal to noise ratio (PSNR) values for three well known graylevel images: Lena (512x512), Barbara (512x512) and House (256x256). For each noise level, we considered 10 noisy versions of each image. The PSNR values listed in Table 1 are the mean values of the 10 individual denoising results. The results are obtained with the Haar wavelet and 4 resolution scales. For the calculation of (10) we used $W = 2$.

From Table 1 we can conclude that the new hard thresholding algorithm, which thresholds on the local magnitude around each coefficient, is not only faster than the original one, but also yields better denoising results. For images with

Table 1: Denoising results [PSNR] of the proposed methods compared to the original method, for different values of σ_n. All results are obtained with the Haar wavelet and 4 resolution scales.

		Standard deviation of noise			
		10	20	30	40
Lena (512x512)	Hard thresholding from [1]	34.4	31.5	29.8	28.5
	Proposed shrinkage version of [1]	34.7	31.8	30.1	28.8
	Proposed hard thresholding method	34.8	31.9	30.2	28.9
	Proposed shrinkage method	35.1	32.1	30.2	28.9
Barbara (512x512)	Hard thresholding from [1]	31.8	27.6	25.4	24.1
	Proposed shrinkage version of [1]	32.1	28.2	26.1	24.8
	Proposed hard thresholding method	32.6	28.5	26.3	24.8
	Proposed shrinkage method	33.0	28.9	26.6	25.2
House (256x256)	Hard thresholding from [1]	34.8	31.8	30.0	28.7
	Proposed shrinkage version of [1]	35.1	32.1	30.4	29.1
	Proposed hard thresholding method	34.9	31.9	30.2	29.1
	Proposed shrinkage method	35.2	32.2	30.4	29.1

little texture, like the House image, the improvements are small (about $0.1dB$), but for more detailed images like Lena or Barbara the PSNR gain is significant ($0.4dB$ for Lena, $0.8dB$ for Barbara).

We can also conclude that the shrinkage variants of the algorithms outperform their hard thresholding versions. For both algorithms a PSNR gain of at least $0.3dB$ can be observed. When we compare our new shrinkage approach with the original hard thresholding technique from [1], we notice a PSNR gain which goes from $0.4dB$ for low textured images (House) to more than $1dB$ for highly textured images (Barbara).

A visual comparison of the original thresholding algorithm and our new shrinkage approach is given in Fig. 5. On top we see a denoised version of a part of the Barbara image (which was corrupted with Gaussian noise of $\sigma_n = 10$) using the hard thresholding technique from [1]. On the bottom we see a denoised version of the same noisy image, using our new method. When comparing the two results one can clearly see that our shrinkage method results in a much smoother and less blocky image. Also the texture is restored in a better way.

6. CONCLUSION

We studied a hard wavelet thresholding algorithm for image denoising in which coefficients are chosen based on their magnitude, spatial connectedness and interscale dependencies. We developed a simple shrinkage version of the algorithm and showed that it outperforms the hard thresholding version (PSNR improvements of about $0.3dB$). We also replaced the spatial connectedness measure by a more simple one, which is based on the mean magnitude of the surrounding wavelet coefficients. We integrated the new measure into a hard thresholding and a shrinkage denoising algorithm. The new shrinkage algorithm outperforms the original one both visually and in terms of PSNR. For low textured images improvements of about 0.4 dB are obtained, for highly textured images the improvements climb to 1 dB and more. Future work will include a study of other wavelets than the Haar wavelet. We will also investigate the possibilities to integrate interscale measures into our shrinkage approach.

REFERENCES

[1] E. Balster, Y. Zheng, and R. Ewing. Feature-based wavelet shrinkage algorithm for image denoising. *IEEE Transactions on Image Processing*, 14(12):2024–2039, December 2005.

[2] D. Donoho. Denoising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(5):613–627, 1995.

[3] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 8:425–455, 1994.

[4] T.C. Hsung, P.K. Lun, and W.C. Siu. Denoising by singularity detection. *IEEE Transactions on Signal Processing*, 47(11):3139–3144, 1999.

[5] B. Huysmans, A. Pižurica, and W. Philips. Image denoising by directional averaging of wavelet coefficients. In *Proceedings of SPIE Wavelet Applications in Industrial Processing III*, volume 6001, Boston, USA, 2005.

[6] M. Jansen and A. Bultheel. Empirical bayes approach to improve wavelet thresholding for image noise reduction. *J. of the American Statistical Association*, 96(454):629–639, 2001.

[7] M. Malfait and D. Roose. Wavelet-based image denoising using a markov random field a priori model. *IEEE Transactions on Image Processing*, 6(4):549–565, 1997.

[8] S. Mallat. *A wavelet tour of signal processing*. Academic Press, London, 1998.

[9] A. Pižurica and W. Philips. Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Transactions on Image Processing*, 15(3):654–665, 2006.

[10] A. Pižurica, W. Philips, I. Lemahieu, and Acheroy M. A joint inter- and intrascale statistical model for wavelet based bayesian image denoising. *IEEE Transactions on Image Processing*, 11(545-557), 2002.

[11] E. Simoncelli and E. Adelson. Noise removal via bayesian wavelet coring. In *Proc. IEEE Int. Conf. on Image Proc. (ICIP'96)*, pages 379–382, Lausanne, Switserland, October 1996.

Figure 5: Visual comparison of the original hard thresholding algorithm [1] (top) and our new shrinkage algorithm (bottom). The results are obtained for the Barbara image with Gaussian noise of $\sigma_n = 10$ added.