Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-2-2004

Shift-Invariant Wavelet Denoising Using Interscale Dependency

P. Chen and D. Suter

MONASH
UNIVERSITY

Shift-Invariant Wavelet Denoising Using Interscale Dependency

Pei Chen and David Suter
Dept ECSE

ABSTRACT

Statistical modeling in the wavelet domain has proven its usefulness, as can be seen in image denoising. The index of PSNR and the visual quality are both improved, compared with Donoho and Johnstone's wavelet threshold method. In [3], a pixel-adaptive Bayesian (PAB) denoising approach in the wavelet domain was proposed, which favorably compares with hidden Markov model (HMM) based approaches. However, the denoised images also suffer from the Gibbs-like artifacts, like ringing around the edges and speckles in the smooth regions. In this paper, we extend the PAB approach [3] to shift-invariant (SI) wavelet denoising in order to reduce these unpleasant artifacts. The experimental result shows that not only is the visual quality greatly improved but also a PSNR gain of about 0.7~0.9 dB is obtained. The proposed approach, called siPAB, outperforms siHMT, which is a competitive SI wavelet denoising approach, by 0.1~0.5 dB.

Keywords: Shift-invariant wavelet transform, image denoising, interscale statistics.

I. INTRODUCTION

Recently, much effort [1-4, 6, 8-10] has been devoted to signal or image Bayesian denoising in the wavelet domain. Wavelet-based image denoising typically consists of three steps: discrete wavelet transform (DWT), the actual denoising process of the noisy wavelet coefficients, and inverse discrete wavelet transform (IDWT). Irrespective of what wavelet is employed, the first and the last steps are 1-1 mappings, and are of no interest in this paper. Only the processing of the noisy wavelet coefficients will be studied in the following sections. In a general Bayesian wavelet denoising approach, as followed here, a prior is first specified for the wavelet coefficients of the unknown image, and then the Bayesian estimate is computed.

Consider a signal x, which is contaminated with noise n, so that one observes $y = x + n$. Assume $x \sim N(0, \sigma_x^2)$, $n \sim N(0, \sigma_n^2)$. The well-known minimal mean squared error (MMSE) estimate is as follows [8]:

$$\hat{x}(y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y \qquad (1)$$

However, the simple Gaussian prior cannot fit well the marginal density of the wavelet coefficients, which tend to be dominated by a few large ones. The actual density of the wavelet coefficients usually has a marked peak at zero and heavy tails, and it is strongly non-Gaussian. The Gaussian mixture model (GMM) [4] and the generalized Gaussian distribution (GGD) [9] are the common tools for modeling this non-Gaussian property of the wavelet coefficients. Although GGD is more accurate in fitting the actual density of the wavelet coefficients than GMM, the latter is preferred in this paper due to its simple form.

Moreover, wavelet coefficients can not be independently treated, although the orthogonal wavelet transform can be viewed as an approximate Karhunen-Loeve transform and tends to approximately decorrelate the image. The reason is that other joint statistics, such as clustering, persistency, decay property across scale, and strong persistence at finer scales [10], exist between the wavelet coefficients of real-life images. These statistical properties in the wavelet domain, like interscale dependency [6, 10] and intrascale dependency [8], have been successfully applied in image denoising. In particular, the hidden Markov tree (HMT) was employed by Crowse *et al.* [6] to capture the interscale dependency. A disadvantage of HMT is the computational burden in the training stage. In order to overcome this computational problem, a simplified universal HMT, named as uHMT [10], was proposed by introducing nine meta-parameters. Although the training stage is not needed in uHMT, its performance is close to that of the more complicated HMT. In [3], a new property of parent/children-type statistics in the wavelet domain was analyzed and a new approach (PAB) was proposed. Noting that the density of the wavelet coefficients can be well fitted by a 3-mode GMM, where the variances for the 3 modes are linearly dependent on their parent, the PAB approach dispenses with the HMM and yet the performance favorably compares with HMM based approaches, like HMT and uHMT [3]. The reason is that the 3-mode GMM captures most of the interscale dependency of real-life images. Secondary properties (non-Gaussianality and persistency) [10], and the tertiary properties (decay property across scale and strong persistence at finer scales) [10], are also explicitly considered in [3].

The images denoised by the PAB approach suffer from some visual artifacts, usually in the form of Gibbs-like ringing around the edges and speckles in smooth regions, as is true of other traditional (maximally decimated) wavelet denoising approaches. The reason for this unpleasant phenomenon lies in the lack of shift invariance (SI) in the traditional wavelet transform. To overcome this drawback, Coifman and Donoho [5] proposed SI wavelet denoising. The strategy for this SI wavelet denoising is cycle-spinning [5]. By cycle-spinning, the ringing and speckle effects can be visibly reduced and the visual quality can be greatly improved, especially when the noise is strong. In terms of PSNR, an improvement of about 0.8~1 dB [2, 10] has been reported by employing SI wavelet denoising.

In this paper, we extend the PAB approach [3] to SI wavelet denoising, which we call siPAB. In section II, we review the modeling of the parent-children statistical property in the PAB approach [3]. In section III, we extend the PAB approach to SI wavelet denoising. In section IV, we analyze the performance of this SI wavelet denoising approach, by the comparison with the PAB approach [3] and with siHMT [10].

II. A review of the PAB approach

A four-level decomposition of the image Lena is shown in Fig.1, where white pixels denote large-magnitude coefficients. Interscale and intrascale dependencies can be easily observed in fig. 1, where large magnitude coefficients cluster near the edges in all bands and in all levels.

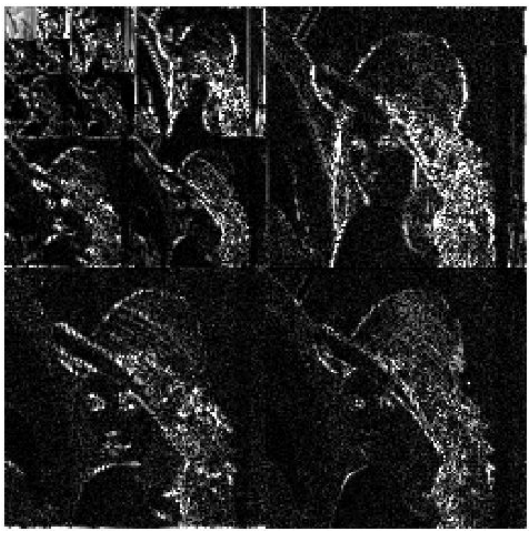

Fig. 1 Four-level wavelet quad-tree decomposition of the Lena image.

In [3], only interscale dependency was employed. Given their parent, wavelet coefficients are modeled as a 3-mode GMM. The variances for these 3 modes in the GMM are linearly dependent on their parent, as is different from HMM-based approaches, where the variances, predefined in uHMT [10] or obtained by expectation-maximization algorithm in HMT [6], are constant. Fig. 2 is a typical conditional density of the child in the vertical band of level 1. The crux of the method in [3] lies in obtaining this conditional density, which shows the statistics of the interscale dependency between children and their parent. It consists of two steps: variance estimation and Gaussian mixture modeling.

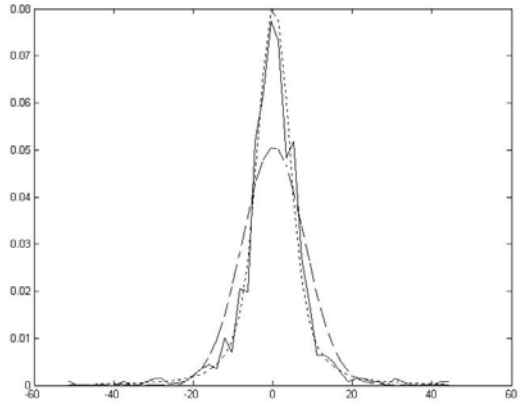

Fig. 2 The conditional density of the child in the vertical band of level 1, the magnitude of whose parent coefficient lies between 10 and 11. The solid line denotes the conditional density, the dotted one for the fixed mixture model, and the dashed one for the Gaussian model.

A. Variance estimation

In PBA [3], the following linear formula is employed to estimate the variance field: $\sigma_c = Am_p + B$, where σ_c is the variance of the child c and m_p is the magnitude of the corresponding parent p. The rationale comes from the intuitive observation that large coefficients persist across scales. So, it can be assumed that the children are of large/small variance if their parent has a large/small magnitude, since the wavelet coefficient can be regarded as zero-mean. Fig. 3 shows this kind of dependency of level 1 on level 2. Other levels share this near-linear property. Usually, in the same level, the horizontal band and the vertical band show similar statistics, while the diagonal band has a smaller variance assuming the magnitude of the parent is same, as it can be seen from fig. 3. So, the true model becomes $\sigma_{c,level,band} = A_{level,band} m_p + B_{level,band}$. For the coarsest level, no parent exists. The parent-on-child dependency is utilized to estimate the variance field for the coarsest level. A similar

linear relationship is specified for the parent-on-child dependency, with different *A*'s and *B*'s. Table 1 lists the experimental parameters for *A* and *B* [3].

Table 1: Parameters for interscale child-on-parent/parent-on-child dependency.

Level	LH		HL		HH	
	A	B	A	B	A	B
1	3.5	0.26	3.7	0.26	2.3	0.15
2	8.5	0.38	10	0.41	6.5	0.29
3	24	0.35	30	0.35	13.5	0.6
4	60	1.9	62	2.4	37	1.1

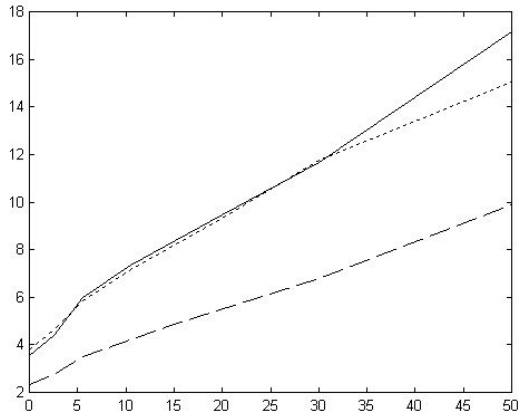

Fig. 3 Variance field estimation in level 1. The solid line denotes the horizontal band of level 1, the dotted one for the vertical band and the dashed one for the diagonal band.

B. Gaussian mixture model

Like the marginal density of the wavelet coefficients in the whole band, the conditional densities of $p(x_c \mid x_p)$ and $p(x_p \mid x_c)$ also have the non-Gaussian property, which shows a marked peak at zero and heavy tails. In PAB, as in Chipman *et al.* [4], a GMM is used to fit this type of non-Gaussian property. If the variance σ for x is known, the following mixture model is specified to fit this non-Gaussian property [3],

$$x \sim a_1 N(0,\sigma_1^2) + a_2 N(0,\sigma_2^2) + a_3 N(0,\sigma_3^2) \quad (2)$$

where $\sigma_1 = \sigma / n_1$, $\sigma_2 = \sigma$ and $\sigma_3 = n_3 \sigma$. $a_1 = 0.6$, $a_2 = 0.3$, $a_3 = 0.1$ and $n_1 = n_3 = 2.5$ were experimentally determined. They work well in [3] and the relation $a_1 / n_1^2 + a_2 + a_3 n_3^2 = 1$ approximately holds. Fig. 2 shows an example and its fitted mixture model.

For a noisy image, the noise-free wavelet coefficients are unknown and it is impossible to obtain the true variances σ in the model mentioned above. In [3], a substitute $\hat{\sigma}$ for σ was obtained from the denoised parent coefficient for all except the coarsest level or from its four noisy children for the coarsest level. With σ_m known, the following MMSE estimator is used to estimate the noisy coefficients

$$\hat{x}(y) = \sum_{m=1}^{3} p(m \mid y) \frac{\sigma_m^2}{\sigma_m^2 + \sigma_n^2} y \quad (3)$$

where $$p(m \mid y) = \frac{a_m g(y, \sigma_m)}{\sum_{i=1}^{3} a_i g(y, \sigma_i)} ,$$

$$g(y, \sigma_m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp(-\frac{y^2}{\sigma_m^2})$$

III. Extension to SI wavelet transform

The images denoised by the PAB approach [3] suffer from some visual artifacts, usually in the forms of ringing around the edges and speckles in smooth regions, as in other traditional wavelet-based denoising approaches. Here, we extend the PAB approach [3] to SI wavelet denoising in order to reduce the ringing and speckle effects. The idea of SI wavelet denoising [5] is simple: first, to circularly shift the image; second, to denoise all the shifted images; last, to align and average over the denoised images. This strategy aims to "average out" the translation dependency in maximally decimated wavelet transform, and was coined as cycle-spinning [5]. However, the direct implementation of Average [Shift-Denoise-Unshift] will have computational complexity $O(n^2)$. In fact, cycle-spinning can be implemented in an undecimated wavelet transform, where the complexity reduces to $O(n \log n)$.

The PAB approach [3] can be easily extended to the undecimated wavelet transform, because the undecimated representation is consistent with the decimated version. The consistency of the undecimated representation means that all the coefficients in the decimated wavelet transform reappear in the new representation. However, the relationship between parent and child is a little different in the two representations. In the decimated wavelet transform, each parent has four children, while the wavelet trees in undecimated representation overlap---the same coefficients appear in more than one tree. This redundancy introduces a 1-1 parent-child relationship. The change from a 1-4 parent-child relationship to a 1-1 relationship makes only a very small change in the top-down procedure to estimate variances. In the traditional decimated 1-4 scheme, the variance estimate for a node, in all except the coarsest level, is

obtained from its denoised parent, and therefore the 1-1 scheme can retain the same character. However, at the coarsest level, since the node has no parent, its variance is estimated from its four noisy children in the decimated 1-4 scheme. Now, with a 1-1 relationship, only one noisy child node is available for each parent at the coarsest level. In order to reduce the noise effect, the variance for a node p in the coarsest level is estimated from the average of the magnitudes of five nodes, which consist of its child c and its four nearest neighbors (denoted by ▲) in the finer level, as in fig. 4.

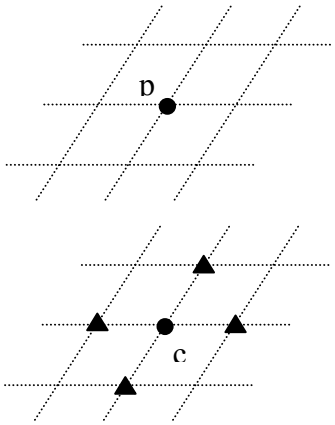

Fig. 4 Five coefficients for the estimation of variance σ_p in the coarsest level.

We summarize the approach below:

1. Implement the wavelet transform, without downsampling.
2. From coarsest to finest, compute the denoised coefficients
   a. Linearly estimate the variance based on its denoised parent (five noisy children for the coarsest level) according to table 1.
   b. Compute denoised coefficients from (3).
3. Reconstruct the denoised image by the undecimated IDWT.

IV. EXPERIMENTAL RESULTS AND CONCLUSION

As in [3, 10], the Daubechies' length-8 wavelet D4 [7] is employed to decompose the images into four levels. To evaluate the proposed denoising scheme, we compare it with the PAB approach [3], and with uHMT [10] and siHMT [10]. In table 2, the results of PSNR for 11 images [10] of size 256×256 are listed. Compared with the PAB approach, this SI-wavelet denoising approach gains an improvement of 0.7-0.9 dB, similar to the gains in other SI wavelet approach [2, 10]. It also outperforms siHMT with a gain of 0.1~0.5 dB.

A visual display of the image "bridge" can be seen in fig. 5. A noticeable improvement of visual quality in the denoised images by SI can be easily observed, over those denoised by PAB, because both the rings and the speckles are greatly eliminated. To compare siPAB and siHMT, we particularly focus on the bridge image to compare the visual effect because the PNSR indexes for two approaches, used to denoise this image, are almost same. Note: it is a little difficult to compare the visual effects, especially in the printed images, because of the similarity as reflected in the near identical PSNR indexes for two approaches. By taking a detailed look at the images, we can see from fig. 6 that siPAB performs slightly better than siHMT in preserving the straight lines, while the siHMT works slightly better than siPAB on the texture.

Table 2: Comparison of PSNR for different approaches with $\sigma_n = 0.05 / 0.1 / 0.2$

05/10/20	siPAB	si-HMT [10]	PAB [3]	uHMT [10]
Baby	33.0/30.0/26.9	33.1/29.6/26.3	32.0/28.8/25.9	32.4/28.9/25.8
Birthday	30.9/28.1/25.6	29.6/26.4/23.7	30.3/27.4/24.9	28.9/25.8/23.1
Boats	31.8/28.2/25.0	31.4/27.4/24.1	31.0/27.3/24.1	30.4/26.4/23.3
Bridge	28.8/25.4/22.7	28.9/25.3/22.7	28.1/24.8/22.0	28.1/24.6/22.0
Buck	33.6/29.8/26.4	33.7/29.6/25.8	32.8/28.8/25.2	32.5/28.4/24.7
Building	30.5/27.2/24.0	30.4/26.6/23.5	29.7/26.3/23.0	29.7/25.9/22.8
Camera	31.0/27.4/24.2	31.1/27.0/23.7	30.2/26.5/23.3	30.3/26.2/23.1
Clown	31.5/28.0/24.6	31.7/27.8/24.5	30.7/27.0/23.6	30.6/26.8/23.7
Fruit	33.1/29.8/26.5	33.3/29.7/26.4	32.4/28.8/25.5	32.2/28.5/25.3
Kgirl	32.4/29.3/26.3	32.6/29.3/26.4	31.8/28.8/25.4	31.6/28.3/25.4
Lena	31.2/27.7/24.9	31.1/27.6/24.5	30.4/26.9/24.1	30.4/26.7/23.8
average	31.41/28.05/25.01	31.28/27.61/24.51	30.66/27.22/24.11	30.43/26.76/23.76

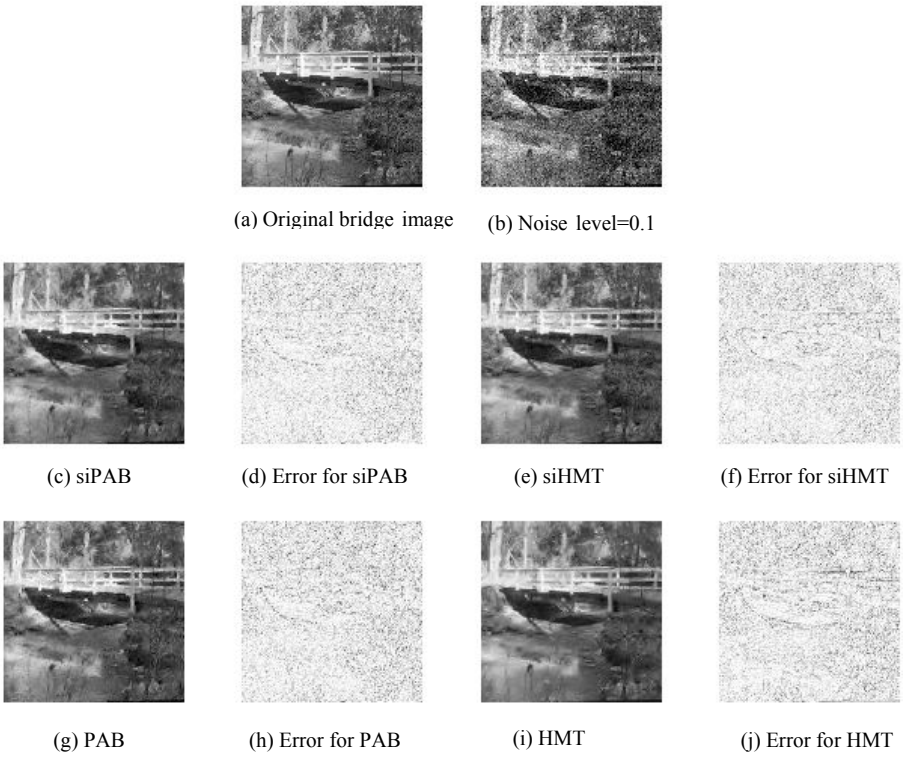

Fig. 5, Image bridge (a), its noised copy (b) with level of 0.1, its denoised copies by (c) siPAB, (e) siHMT, (g) PAB, and (j) uHMT, and their error images. In order to have a visible scene, the error images have been scaled at a same ratio. The darker the pixel, the bigger the error magnitude. The SI wavelet denoising has an improved visual effect, which can be seen by comparing siPAB with PAB and by comparing siHMT with HMT. As siPAB and siHMT are concerned, it is a little difficult to compare the visual effect, especially for the printed images, because the PSNR indexes are almost same, where PSNR for siPAB is 25.42 and siHMT has an index of 25.36.

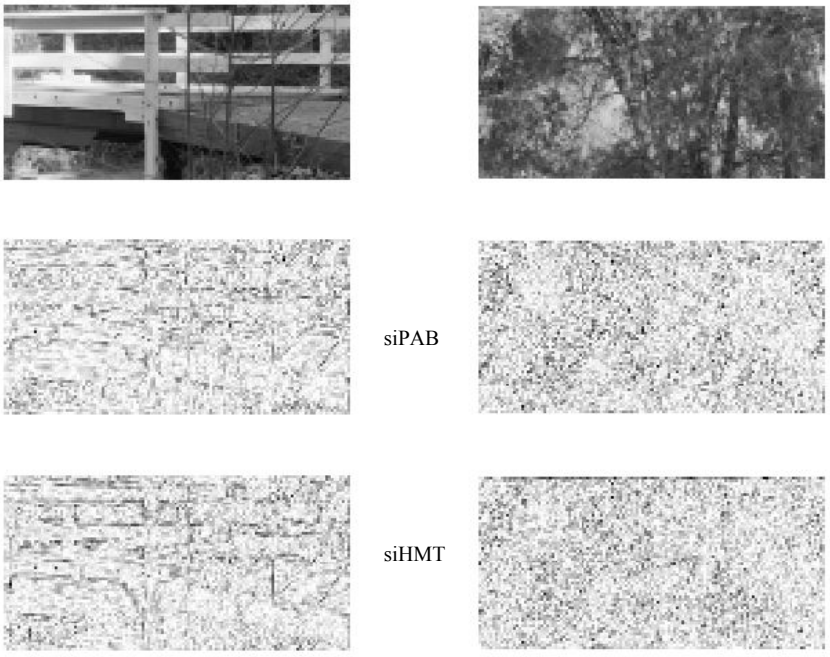

Fig. 6, Two different portions of the bridge image and their error images for siPAB (the second row) and siHMT (the third row). (a) siPAB performs a little better in preserving the structures, like the straight lines of the balustrade of the bridge, than siHMT. (b) siHMT works slightly better than siPAB on the textured regions.

REFERENCES

1. R. W. Buccigrossi, and E. P. Simoncelli, 'Image compression via joint statistical characterization in the wavelet domain', IEEE Image Process., Vol. 8, No 12, Dec. 1999, pp. 1688-1701.
2. S. G. Chang, B. Yu, and M. Vetterli: 'Spatially adaptive wavelet thresholding with context modeling for image denoising', IEEE Image Process., Vol. 9, No 9, Sep. 2000, pp. 1522-1531.
3. P. Chen and D. Suter, A simple pixel-adaptive Bayesian approach to image denoising using wavelet interscale dependency, Technical Report MECSE-1-2002, Monash University, Clayton 3800, Australia.
4. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch: 'Adaptive Bayesian wavelet shrinkage', J. Amer. Stat. Assoc., Vol. 92, No 440, Dec. 1997, pp. 1413-1421.
5. R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," in Wavelets and Statistics, Springer Lecture Notes in Statistics 103. New York: Springer-Verlag, 1994, pp. 125–150.
6. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk: 'Wavelet-based statistical signal processing using hidden Markov Models', IEEE Sig. Process., Vol. 46, No 4, Apr. 1998, pp. 886-902.
7. I. Daubechies: 'Ten lectures on wavelet', Philadelphia, SIAM, 1992.
8. M. K. Mihcak, I. K. Kozintsev, K. Ramchandran, and P. Moulin: 'Low-complexity image denoising based on statistical modeling of wavelet coefficients', IEEE Sig. Process. Lett., Vol. 6, No 12, Dec. 1999, pp. 300-303.
9. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors", IEEE Infor. Theory, Vol. 45, No 3, Apr. 1999, pp. 909-919.
10. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models", IEEE Image Process., Vol. 10, No 7, Jul. 2001, pp. 1056-1068.