Wavelet-based image denoising using non-stationary stochastic geometrical image priors

Sviatoslav Voloshynovskiy*, Oleksiy Koval, and Thierry Pun
University of Geneva, Department of Computer Science, 24 rue Général-Dufour,

ABSTRACT

In this paper a novel stochastic image model in the transform domain is presented and its superior performance in image denoising applications is demonstrated. The proposed model exploits local subband image statistics and is based on geometrical priors. Contrarily to complex models based on local correlations, or to mixture models, the proposed model performs a partition of the image into non-overlapping regions with distinctive statistics. A close form analytical solution of the image denoising problem for AWGN is derived and its performance bounds are analyzed. Despite being very simple, the proposed stochastic image model provides a number of advantages in comparison to the existing approaches: (a) simplicity of stochastic image modeling; (b) completeness of the model, taking into account multiresolution, non-stationary image behavior, geometrical priors and providing an excellent fit to the global image statistics; (c) very low complexity of the algorithm; (d) tractability of the model and of the obtained results due to the closed-form solution and to the existence of analytical performance bounds; (e) extensibility to different transform domains, such as orthogonal, biorthogonal and overcomplete data representations. The results of benchmarking with the state-of-the-art image denoising methods demonstrate the superior performance of the proposed approach.

Keywords: denoising, stochastic modeling, ML, MAP, variance estimation, wavelets, overcomplete expansions.

1. INTRODUCTION

Due to various factors during acquisition and transmission, an image might be degraded by noise leading to a significant reduction of its quality. Therefore, it is necessary to apply an efficient denoising technique to compensate for such annoying effect. To select the proper denoising method, correct assumptions about signal and noise statistics should be made as well as about their interaction mechanism.

In the scope of this paper we wish to find the most reliable estimate of the original image data given its degraded version obtained as an additive mixture with a noise component:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}\,, \tag{1}$$

where $\mathbf{y}$ is the degraded noisy data; $\mathbf{x}$ is the original unknown image; $\mathbf{n} \sim N(0, \sigma_n^2)$ is an independently identical distributed additive white Gaussian noise (AWGN) with zero-mean and known variance σ_n^2. The dimensionality of $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{n}$ is $\mathbf{N}\times\mathbf{N}$.

To solve this AWGN removal problem in the above formulation a number of effective techniques are known, that could be classified based on the underlying stochastic image model and on the type of the transformation used.

According to the first classification criterion we should first mention the denoising algorithms that are based on the global i.i.d. assumption about image data, which are modeled using a single marginal distribution. Introduced by Nikolova in 1996 [1], this approach became very popular. The main issue to be solved according to this concept is to find the best estimator for data with i.i.d. Generalized Gaussian distribution (GGd) [2, 3] or Student law [4]. Following this direction we are coming to different types of thresholding and shrinking techniques that are optimal from the information-theoretic approach point of view as was shown by Moulin and Liu [5]. Another approach proposes to divide an image into two regions of edges and flat areas and to apply different models to describe these regions (GGd and Gaussian distribution correspondingly) [6]. The main drawback of these models is their inability to capture local statistics that play a crucial role in denoising applications. Moreover, by applying global permutation of image data we do not change their marginal properties and therefore denoising based on global statistical models does not reflect the local image structure.

* Correspondance: Email: svolos@cui.unige.ch; WWW: http://sip.unige.ch/~vision

Another possible solution is inspired by lossy wavelet data compression where the so-called *estimation-quantization* (EQ) model was developed and demonstrated to have superior performance in denoising application [7]. The main idea of the EQ model consists in representing the image data in wavelet subbands as an i.i.d. Gaussian field with zero mean and slowly varying variance. Under such an assumption the best estimator will coincide with a classical scalar form of the Wiener filter. This strategy has given significantly better denoising performance (approximately 2 dB) than for the previous case [8].

An attempt to combine the EQ strategy with the GGd prior image model was reported in [3]. However, the performance of this algorithm was worse than for the locally Gaussian prior case.

The second important classification criteria is the domain where the denoising procedure is applied. The simplest solution is to denoise the data in the coordinate domain directly. If data are assumed to be i.i.d. stationary Gaussian, the denoising will coincide with a classical scalar Wiener filter [9]. This solution is the simplest one but it is not the best according to the output image quality.

Probably, the most popular choice towards the end of 90's for denoising applications was to use an orthogonal critically sampled wavelet transform (Figure 1.a) [2,3,4]. The wavelet transform captures most of the information about the image in the low frequency part where noise presence is not so significant, and represents the data in a pyramidal and a sparse way. These give new possibilities for effective processing (for instance, only a few coefficients in high frequency wavelet subbands significantly contribute to the image quality). The efficiency of the algorithms is higher in this case in comparison with the coordinate domain choice but in edges area the Gibbs phenomenon was observed in the reconstructed image as well as near image borders [10].

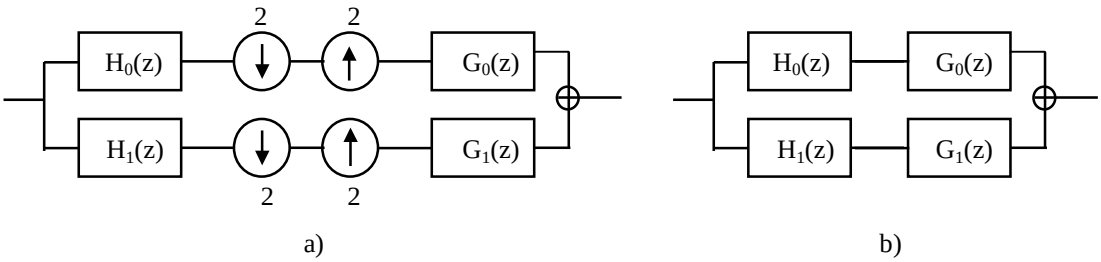


Figure 1. (a) Structure of 1-level critically sampled and (b) non-decimated wavelet decomposition; the filters for the analysis stage are denoted by H(z) and for the synthesis stage by G(z); index "0" corresponds to the low pass filters and index "1" corresponds to the high pass filters.

First introduced by Coifman and Donoho [11], non-decimated or stationary wavelet transform overcomes this Gibbs artifacts problem by eliminating its down- and upsampling stages (Figure 1.b). Moreover, using biorthogonal filters [12], boundary problems could be also solved. A denoising technique that exploits an improved version of the EQ model and this type of transformation demonstrated a very good performance [12].

The recent work of Sterla et al [13] presented a more powerful transform with a larger amount of orientations (steerable pyramid [14]) that could bring additional benefit for the performance of the algorithm. But the main advantage is coming from the correct assumption made about the image statistics or the stochastic model of the data that are supposed to be correlated in the transform domain. Each sample in this case could be completely determined according to the Gaussian scale mixture model, as the product of a zero-mean unit variance stationary Gaussian random variable with some positive log-normally distributed scalar. In this case, vector processing is applied to denoise the image. Experimental results presented in [13] show that this method provides the best quality of the denoised image among currently known techniques.

Therefore, the main focus in this paper is to develop a stochastic image model that will allow to increase the denoising performance of the algorithm according to the peak signal-to-noise ratio (PSNR) criterion without this vector processing and without the need for a more complex transform.

The structure of the paper is the following. In Section 2 a stochastic approach to removal of AWGN is reviewed and analyzed analytically. As a result, some problems when maximum likelihood (ML) strategy is used for data variance estimation are pointed out. In Section 3, a new stochastic image model based on geometrical prior information about an image structure will then be introduced, as well as the solution to this problem. In Section 4, three versions of the denoising algorithm based on geometrical priors will be proposed to investigate the model effectiveness in the coordinate, orthogonal critically sampled wavelet and biorthogonal non-decimated wavelet transform domains. In Section 5, a new stochastic image model, the so-called edge process model, will be finally introduced and investigated in the wavelet and in the overcomplete transform domains.

2. MAP IMAGE DENOISING

2.1. Stochastic image priors

In order to find the most accurate estimate of an original image $\mathbf{x}$ based on its degraded version $\mathbf{y}$ by taking into account any available proper prior information, a maximum a posteriori probability (MAP) estimate was selected to perform denoising:

$$\hat{\mathbf{x}} = \arg\max_{\tilde{\mathbf{x}}} \mathbf{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \tilde{\mathbf{x}})\, \mathbf{p}_{\mathbf{X}}(\tilde{\mathbf{x}}) \tag{2}$$

where $\hat{\mathbf{x}}$ denotes the estimate of the original image, $\mathbf{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \tilde{\mathbf{x}})$ is the likelihood function for the AWGN in our case, $\mathbf{p}_{\mathbf{X}}(\tilde{\mathbf{x}})$ is the prior distribution of the original image.

Assuming that both data and noise are Gaussian processes with respective means $\bar{\mathbf{x}}$ and 0 and respective covariance matrixes $\mathbf{C_x}$ and $\mathbf{C_n}$, i.e. $\mathbf{x} \sim N(\bar{\mathbf{x}}, \mathbf{C_x})$ and $\mathbf{n} \sim N(\mathbf{0}, \mathbf{C_n})$, the solution to problem (2) becomes:

$$\hat{\mathbf{x}} = \mathbf{C_n}(\mathbf{C_x} + \mathbf{C_n})^{-1}\bar{\mathbf{x}} + \mathbf{C_x}(\mathbf{C_x} + \mathbf{C_n})^{-1}\mathbf{y}\,. \tag{3}$$

It is well known that after application of wavelet or overcomplete transforms the global mean value of the data in subbands $\bar{\mathbf{x}}_{\mathbf{tr}} = \mathbf{0}$. Therefore, expression (3) becomes:

$$\hat{\mathbf{x}}_{\mathbf{tr}} = \mathbf{C}_{\mathbf{x}}^{\mathbf{tr}}(\mathbf{C}_{\mathbf{x}}^{\mathbf{tr}} + \mathbf{C}_{\mathbf{n}}^{\mathbf{tr}})^{-1}\mathbf{y}_{\mathbf{tr}}\,, \tag{4}$$

where index **tr** denotes the transform domain. In the case of a non-stationarity assumption (4) could be rewritten as:

$$\hat{x}_{tr}[k] = \frac{(\sigma_x^{tr}[k])^2}{(\sigma_x^{tr}[k])^2 + (\sigma_n^{tr})^2}\, y[k]\,. \tag{5}$$

In following, we refer to the transform domain omitting index **tr**.

2.2. MAP estimator performance analysis

One of the conditions required to have an optimal estimator is to guarantee a minimum variance of the estimation error. For the above MAP estimator the variance of the estimation error can be expressed as:

$$Var(e) = \sigma_e^2 = \mathbf{E}\left(\left|x - \hat{x}\right|^2\right), \tag{6}$$

where $\mathbf{E}$ denotes the expectation operator.

In the case of a Wiener filter applied to stationary Gaussian data, expression (6) has the following form:

$$\sigma_e^2 = \frac{\sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2}\,. \tag{7}$$

Therefore, the higher is the variance of the image, the lower is the accuracy of the estimation. It is clear that for stationary data the estimation performance is fixed. To analyze the non-stationary Gaussian data case, the local variance estimate should be obtained. Assuming local stationarity this can be performed using a maximum likelihood (ML) strategy:

$$\sigma_x^2 = \frac{1}{(2L+1)^2} \sum_{k=-L}^{L} \sum_{l=-L}^{L} (x(i+k,\, j+l) - \bar{x}(i,\, j))^2, \tag{8}$$

where $\bar{x}(i, j)$ is the local mean in a (2L+1)× (2L+1) sampling window:

$$\bar{x}(i, j) = \frac{1}{(2L+1)^2} \sum_{k=-L}^{L} \sum_{l=-L}^{L} x(i+k, j+l). \tag{9}$$

It is known that this variance estimate is only asymptotically unbiased. For a 1D sequence of $\mathbf{x}:[x(0), x(1),..., x(N-1)]$ we have instead of (9) and (8) the following equations:

$$\bar{x} = \frac{1}{N} \sum_{k=0}^{N-1} x[k], \tag{10}$$

$$\sigma_x^2 = \frac{1}{N} \sum_{k=0}^{N-1} (x[k] - \bar{x})^2 . \tag{11}$$

The biases of the local mean and variance estimates are:

$$\mathbf{E}(\bar{x}) = \mathbf{E}\left(\frac{1}{N} \sum_{k=0}^{N-1} x[k] \right) = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{E}\big(x[k]\big) = \bar{x}; \tag{12}$$

$$\mathbf{E}(\sigma_x^2) = \mathbf{E}\left[\frac{1}{N} \sum_{k=0}^{N-1} (x[k] - \bar{x})^2 \right] = \frac{1}{N^2} \sum_{k=0}^{N-1} \mathbf{E}\left[x[k]^2\right] - \mathbf{E}[\bar{x}]^2 = \frac{1}{N^2} \left[N(\sigma_x^2 + \bar{x}^2) - N\left(\frac{\sigma_x^2}{N} + \bar{x}^2 \right) \right] = \sigma_x^2 - \frac{1}{N} \sigma_x^2 . \tag{13}$$

Although, the estimate of the mean is unbiased (eq. (12)), the variance estimate is only asymptotically unbiased (eq.(13)), i.e., $\lim_{n \to \infty} \mathbf{E}(\sigma_x^2) = \sigma_x^2$, and to decrease the bias one needs to enlarge the sampling space N.

3. A NEW STOCHASTIC NON-STATIONARY IMAGE MODEL BASED ON GEOMETRICAL PRIOR INFORMATION

3.1. Coordinate domain setup

The following example should help link the results of Section 2 with the foregoing issues of stochastic image modeling. It is well known that real images have a non-stationary nature. Let us assume that an image can be represented as a union of a number of statistically homogeneous regions of different intensity levels (see Figure 2.a) we can then model the data inside each region as a stationary Gaussian process with some variance (for instance, $\sigma_x^2 = 100$) (see Figure 2.b and Figure 2.c). In this case the application of the classical ML estimators based on the sampling space located in the boundary regions (Figure 2.d), to estimate the local variance will lead to completely incorrect local variance estimates. Assume that in the sampling window there are pixels from the two-component i.i.d. stationary Gaussian mixture of mean values $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ and variances σ_1^2, σ_2^2, respectively, then the joint probability density function (pdf) is:

$$p_{\mathbf{x}}(\mathbf{x}) = \alpha N(\bar{\mathbf{x}}_1, \sigma_1^2 \mathbf{I}) + (1-\alpha) N(\bar{\mathbf{x}}_2, \sigma_2^2 \mathbf{I}), \tag{14}$$

where α is the fraction of the data in the mixture from the first component. For the mixture distribution then we have the corresponding mean and variance:

$$\bar{x} = E[\mathbf{x}] = \alpha \bar{\mathbf{x}}_1 + (1-\alpha) \bar{\mathbf{x}}_2, \tag{15}$$

$$\sigma^2 = E[(\mathbf{x} - \bar{x})^2] = \alpha \sigma_1^2 + (1-\alpha) \sigma_2^2 + \alpha(\bar{x} - \bar{x}_1)^2 + (1-\alpha)(\bar{x} - \bar{x}_2)^2 . \tag{16}$$

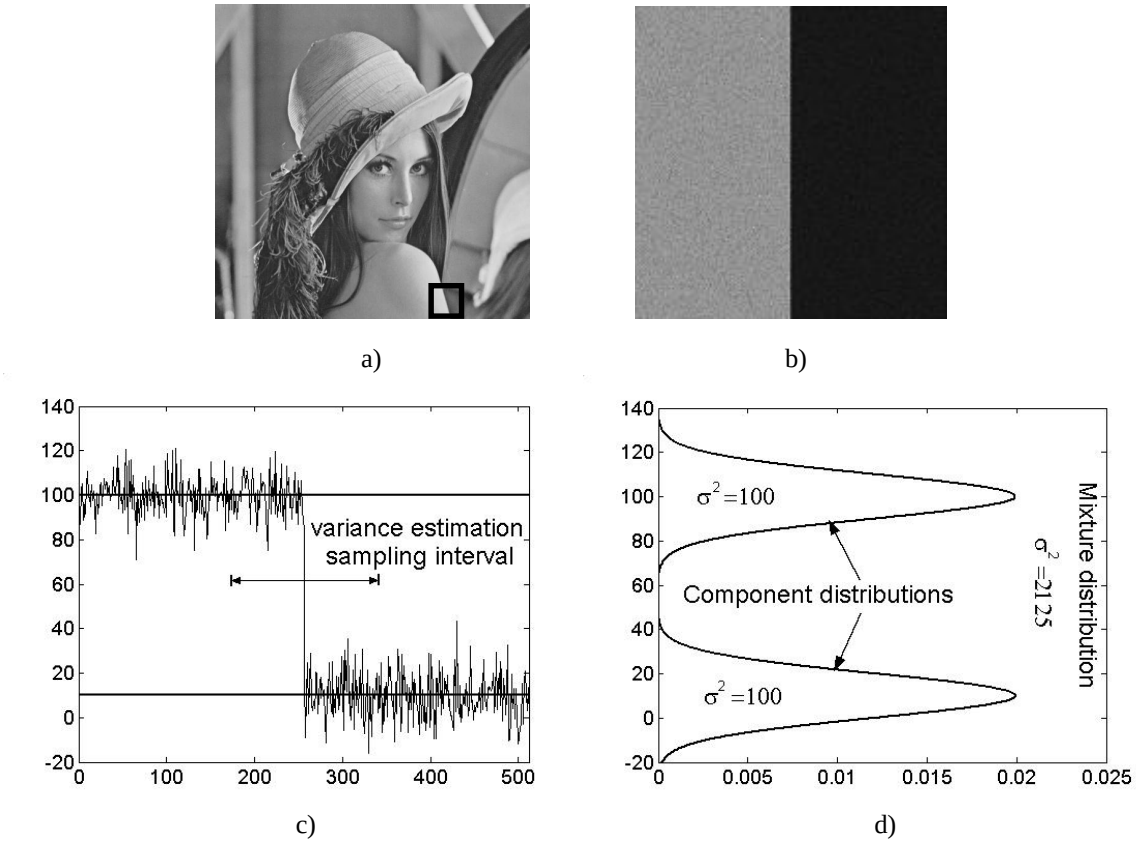


Figure 2. Variance estimation using standard ML strategy for the edge region:(a) "Lena" test image and its fragment (marked by the square); (b) two-region modeling of Lena's fragment; (c) modeling example for 1D edge profile; (d) individual components and resulting pdf (with estimated variances).

If for instance α =0.5, then the overestimation ratio (estimated-to-correct variance ratio), for the case of above selected variances, is more than 20 times. The main issue to be resolved in order to decide a proper stochastic image model is, therefore, to resolve the trade-off between reducing the estimate bias while keeping into account image non-stationarity.

As the solution to this problem we propose to model an image using the following partition of its support S as a set of non-overlapping regions (Figure 3):

$$\bigcup_i R_i = S; \quad R_i \cap R_j = 0; \quad i = 1,2,...,k\,. \tag{17}$$

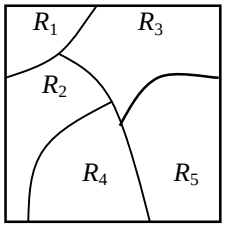


Figure 3. Image partition as a set of regions with i.i.d. homogeneous statistics.

Let $\mathbf{x}_i$ denotes the subset of image pixels supported by the region $\boldsymbol{R}_i$. In our model we assume that each region is fully covered by the probabilistic model $\theta_i \in \{\Theta_1, \Theta_2, ..., \Theta_K\}$ and no two neighboring regions are described by the same model. In particular, we assume that the pixels in the image subregion $\mathbf{x}_i$ are distributed according to the joint pdf $p_{\mathbf{x}}(\mathbf{x}_i|\theta_i)$. The

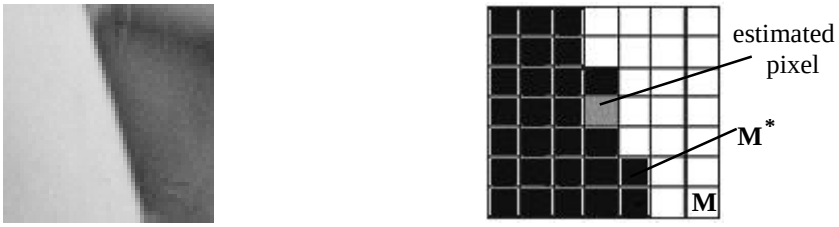
class of such models is very broad and depends on the used domain (coordinate or transformed). In the following, we only concentrate on Gaussian models, meaning $p_x(\mathbf{x}_i|\theta_i) = N(\bar{\mathbf{x}}_i, \sigma_i^2\mathbf{I})$. This model is different from the classical Gaussian mixture model used for example as the background assumption for the EQ model, since it assumes the homogeneity of the data within the same region, while the Gaussian mixture model allows the presence of samples with different statistics. This explains the fact that "permutation" of the coefficients is not reflected on the marginal statistics.

According to our model, to perform a correct ML-based local data variance estimation one needs to take into account only those coefficients of the transformed domain that belong to the subset $\mathbf{M}^*$ of the local sampling space **M.** $\mathbf{M}^*$ contains the elements from the corresponding subregion R_i, to which the estimated sample belongs (Figure 4). In this case the classical ML estimate should be replaced by a restricted support ML local variance estimate:

$$\hat{\sigma}_x^2[k] = \frac{1}{\left|\mathbf{M}^*\right|} \sum_{k \in M} x[k]^2 m_k, \tag{18}$$

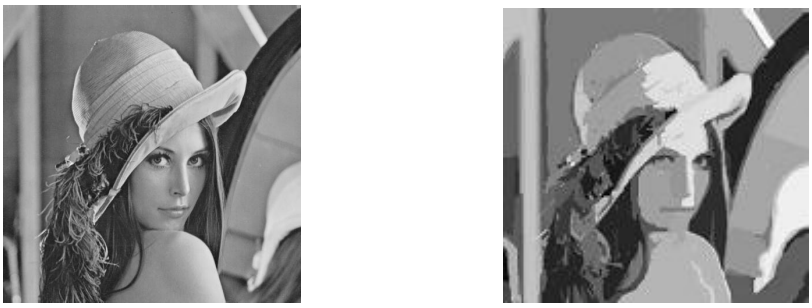
where m_k is a subset indicator function, $m_k = \begin{cases} 1, m_k \in \mathbf{M}^* \\ 0, m_k \notin \mathbf{M}^* \end{cases}$; $\left|\mathbf{M}^*\right|$ is the cardinality of the subset $\mathbf{M}^*$, and the dimensionality of $\mathbf{M}$ is selected to be large enough to guarantee compensation of the bias.

Such partition can be performed using any known segmentation software; in our case we used the software [15] developed by Cornell University (Figure 5.b).



a) b)

Figure 4. (a) Fragment of test image Lena; (b) local estimation window $\mathbf{M}$ with indication of the subset $\mathbf{M}^*$ (black pixels) and of the estimated pixel (gray) for the ML local variance estimation.



a) b)

Figure 5. Segmentation results: (a) test image Lena; (b) segmented image Lena using the Cornell University segmentation software.

3.2. Transform domain setup

If one transforms the data to a wavelet or overcomplete domain, the bias compensation problem still exists but in a different

set-up: since the data mean value in the transform domain is equal to zero. Therefore, instead of using a step edge mean and an equal component variance test signal to justify the introduced stochastic model, we propose to study the following two-texture example (Figure 6.a and Figure 6.b). In the sampling space there are at least two components with zero mean and different variances (100 and 25). According to (16) it is easy to verify that the calculated variance estimate will be also incorrect. Therefore, the region partition technique (Figures 3-5) and the reduced support ML procedure (18) are needed to eliminate the estimation error.

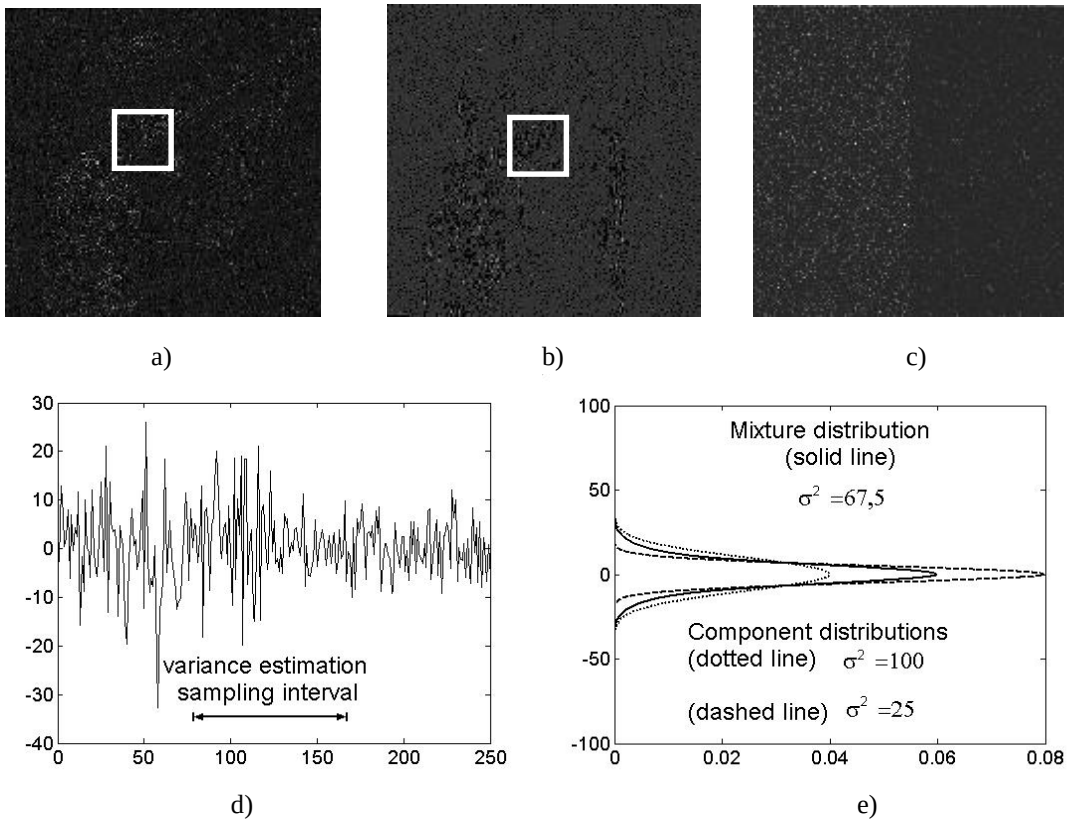


Figure 6. Variance estimation using classical ML strategy for the edge image region: (a) horizontal subband of the first level wavelet; (b) non-decimated transform of the test image Lena and its fragment (marked by the square); two-region modeling example: (c) two-dimensional and (d) one-dimensional case; (e) individual component and resulting pdf with estimated variances (e).

Therefore, the only difference with the model definition in the coordinate domain is that the data in the regions R_i covered by the model $\theta_i \in \{\Theta_1, \Theta_2, ..., \Theta_K\}$ are distributed according to a joint zero-mean Gaussian pdf $p_x(\mathbf{x}_i|\theta_i) = N(0, \sigma_i^2\mathbf{I})$.

4. ALGORITHM IMPLEMENTATION AND BENCHMARKING

According to the benchmarking results summarized in [16], one can see that denoising in the coordinate domain can not be as efficient as in the transform domain. However, to estimate the benefit we can expect from the proposed underlying stochastic model it could be really indicative. In this case, assuming locally stationary Gaussian image behavior, i.e., $x \sim N(\mu_x[k], \sigma_x^2[k])$, denoising will consist in the application of the scalar Wiener filtering based on the ML local variance estimate, using the region partition map (Figure 5) as side information to form the correct sampling space (Figure 4,b). The Cornell University segmentation software was used to generate this map. The PSNR performance improvement for the case of Lena image and the developed model is 0.11 dB contrary to the wiener2 procedure from Matlab with known variance of the noise (correspondently, 29.41dB and 29.30 dB).

To prove the effectiveness of the proposed model in practice we developed two versions of the AWGN denoising algorithm for the critically sampled and the non-decimated wavelet transform domains. The structure of the algorithm is

shown in Figure 7 and includes three main blocks: direct transform block (DT), Wiener filter block (WF) that receives side information (geometrical prior (GP) information) about regions R_i, and inverse transform block (IT).

The critically sampled wavelet transform is implemented using the orthogonal Beylkin filter pair [17] (5 decomposition levels were selected as in [8]). The noise reduction method is the same as the one used for the coordinate domain. The side information for all decomposition levels except the fifth one is obtained from the reconstructed lowpass subband. It is the same for all highpass subbands from all levels. The data on the fifth level are denoised using the classical ML local variance estimate. The local variance estimation is performed in a local window of size from 5×5 to 11×11 depending on the decomposition level.

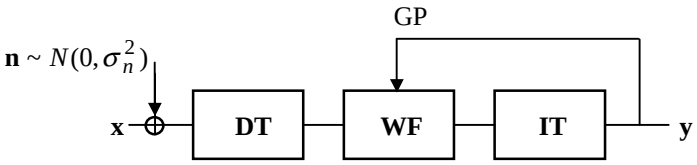


Figure 7. Structure of the proposed denoising algorithm in the transform domain.

A practical problem that was encountered stemmed from the computational complexity of the segmentation software to obtain the partition. This caused a significant increase in execution time for transform domain denoising, where the segmentation needs to be performed repeatedly in the various subbands. To overcome this problem, we propose to use a quantization-based segmentation, which is performed using a simple uniform quantization of the image dynamic range (Figure 8). For optimization, we performed a number of tests to establish the optimal number of quantization bins, from the point of view of quality of the denoised image 4 bins was the best value. Using this simpler segmentation, we obtained denoising results similar to those attained using the Cornell University segmentation software, while reducing segmentation time by a factor of 2.5. Therefore, all following experiments were obtained using this quantization-based segmentation.

Figure 8. Quantization based segmentation results for the 4-bin uniform quantizer case.

The algorithm has a similar structure in the non-decimated biorthogonal transform domain. According to the schema presented in Figure 1.b, the transform was accomplished using 9/7 CDF [18] biorthogonal filter pair and 4 levels of decomposition. In addition, we took into account the fact that coefficient variances in high frequency subbands are distributed according to a marginal distribution that can be very closely approximated by the Rayleigh law (Figure 9):

$$p(x) = \frac{x}{s^2} \exp\left(-x^2 \Big/ 2s^2 \right), \tag{19}$$

where s is a scale parameter. This knowledge of the prior variance distribution allows us to apply a MAP estimate rather than a ML estimate for the local variance:

$$\sigma^2_{MAP} = \frac{s^2(n-1)}{2}\left[-1 + \sqrt{1 + \frac{4}{s^2(n-1)} \sum_{i=1}^{n} x_i^2} \right]. \tag{20}$$

The scale parameter s is estimated using an ML estimate in a 3×3 sampling window:

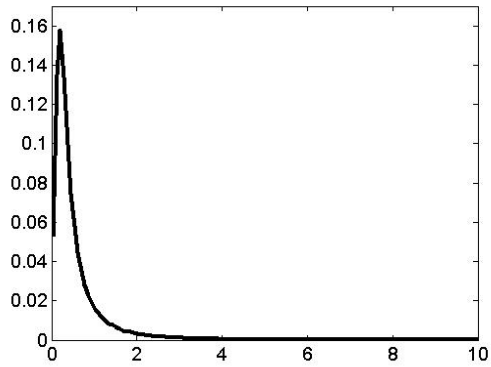


Figure 9. Histogram of the variance of the high frequency subband (first decomposition level diagonal orientation) of test image Lena in the overcomplete transform domain.

$$s^2 = \frac{1}{2n}\sum_{i=1}^{n} x_i^2. \tag{21}$$

The proper sampling space is again formed based on the denoised lowpass subband geometrical image prior information, using the proposed quantization based segmentation (except for the fourth decomposition level subbands). A complete sampling space dimensionality of size of 15×15 was found to be optimal from the output image PSNR point of view.

To verify the performance of the developed algorithms, we applied them to the 8 bit 512×512 test images Lena and Barbara for 100, 225, 400, 625 noise variances of the AWGN, and compared it with the best denoising techniques. Due to the fact that none of the candidates simultaneously is the best for the case of two test images, the benchmarking was performed using the average PSNR for these images for a particular noise variance value (Table 1). The average PSNR results prove that for the critically sampled transform the performance of the proposed algorithm is the best, but for the case of the overcomplete domain the method proposed in [13] provides better results. The first explanation for this comes from the low robustness to noise of the segmentation, which leads to a bias during the partitioning process. The second reason is connected to the more sophisticated transform type used in the best algorithm, which decorrelates the data in more than 3 directions (steerable pyramid transform [14]), and to the assumption that data in the linear transform domain are still correlated. A more complex vector form of the Wiener filter (4) should be used, but this would not be attractive for many practical applications. Moreover, without this correlation assumption this algorithm [13] performs worse the proposed in this paper.

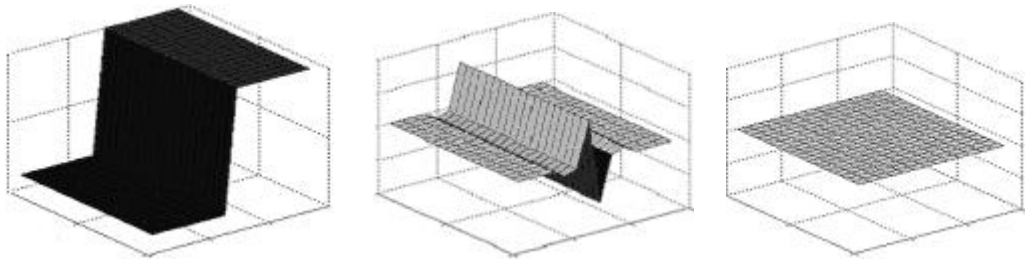
Table 1. Comparison of average PSNR [dB] results for several methods and both test images.

Denoising method	**Standard deviation of the noise**				**Transform type**
	10	**15**	**20**	**25**	**Wavelets**
Noisy image	28.13	24.63	22.10	20.17	
[2]	33.61	31.30	-	-	*Db8*
[8]	33.52	31.32	29.81	28.64	*Db8*
[3]	-	30.51	29.07	28.01	*Symmlet 8*
[19]	32.89	30.58	29.09	28.01	*Db8*
[20] (1D HMM)	33.49	31.30	29.80	28.68	*Db8*
[20] (AHMF)	33.58	31.41	29.92	28.80	*Db8*
[21]	33.55	31.35	29.90	28.75	*Db8*
Proposed method	**33.68**	**31.51**	**29.92**	**28.84**	*Beylkin*
					Overcomplete
[3]	-	31.82	30.31	29.12	*Symmlet 8*
[12]	34.12	32.04	30.55	29.40	*10/18*
[21]	34.30	32.20	30.70	29.55	*Symmlet 8*
[13]	34.38	32.39	31.01	29.95	*steerable*
Proposed method	**34.41**	**32.37**	**30.98**	**29.86**	*9/7*

5. DENOISING BASED ON THE EDGE PROCESS MODEL

Based on the experimental results of Section 4, one can conclude that to maximize the denoising performance it is necessary to take into account local data correlation in the stochastic image model. Since this could lead to an increase in computational complexity of the algorithm, the natural question arises: whether it is possible to enhance the algorithm performance without significantly increasing its complexity?

The residual correlation of the data in the high frequency subbands exists because no linear transform is able to completely decorrelate the edges. Therefore, if one finds a way to completely remove the edges from the subband data, this will allow an increase in performance without the need for any vector processing application (Figure 10).



a) b) c)

Figure 10. Edge subtraction from the high frequency subband data: (a) original data representing a test step edge in the coordinate domain; (b) non-decimated transform of the edge data; (c) subtracted edge data.

As a practical example, let us consider the fragment of Lena test image (Figure 2.a) in the overcomplete transform domain. If there is a way to correctly estimate and subtract the edges from the data (Figure 11.a), then three main results can be obtained (Figure 11.b-d):

1) the variances within the subbands are significantly reduced;
2) the signal becomes more decorrelated;
3) marginal highly non-Gaussian distribution are transformed into the distribution very close to pure Gaussian one.

It is therefore not any necessity to perform vector processing, and consequently the computational complexity of the denoising algorithm remains the same.

Assuming that the variation of the data along the edge is close to zero and that the global flat information can be precisely modeled using a non-stationary Gaussian distribution we propose a new so-called *edge process* (EP) image model for each single subband level. Keeping the same notation as in Section 3.1 we assume that each subband $S_j, l = 3J$, where J is a number of transform decomposition levels, could be represented as a union of two main non-overlapping regions R_1 and R_2:

$$\bigcup_{i=1}^{2} R_i = S_l; \quad R_1 \cap R_2 = \varnothing. \tag{22}$$

The first region R_1 represents all flat image areas while the second one R_2 is again a union of non-overlapping subregions R_2^i. The data inside R_1 is assumed to be non-stationary i.i.d. zero mean Gaussian. Each of the R_2^i subregions is represented by a stationary non-zero mean and small variance Gaussian model, i.e.:

$$R_1 = \{\mathbf{x} : x[k] \sim N(0, \sigma_x^2[k])\}; \tag{23}$$

$$R_2 = \bigcup_i R_2^i; \forall\, j \neq i : R_2^i \bigcap R_2^j = \varnothing; \tag{24}$$

$$R_2^i = \{\mathbf{x} : x[k] \sim N(\mu_i, \sigma^2)\}. \tag{25}$$

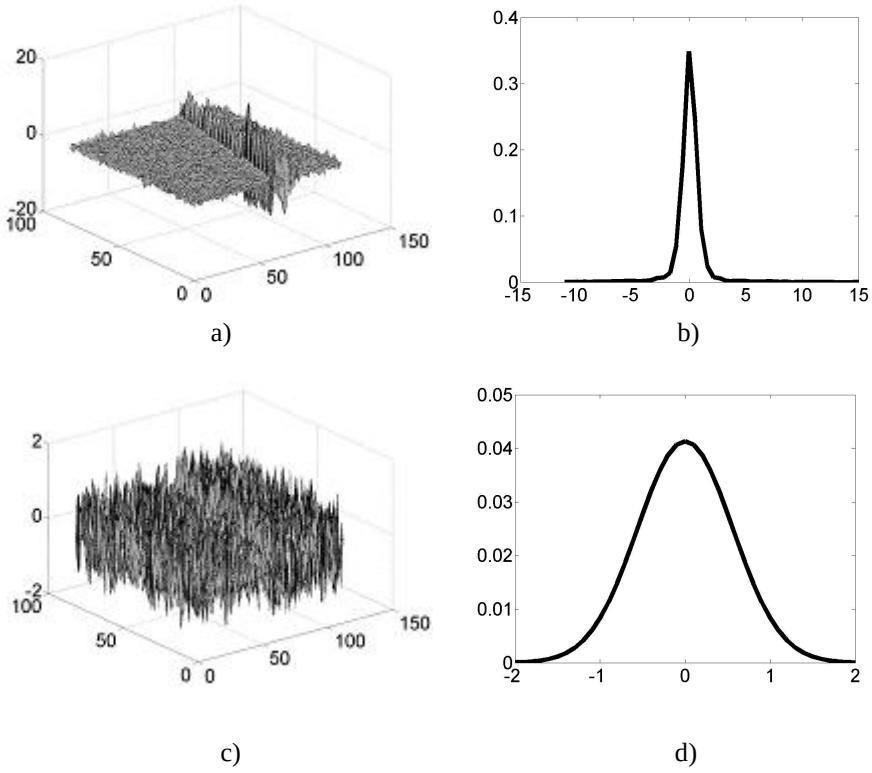


Figure 11. Edge subtraction from the high frequency subband data: (a)original data with variance $\sigma^2 = 1.81$; (b)corresponding histogram; (c) subtracted edge data with variance $\sigma^2 = 0.31$; (d) corresponding histogram.

The expressions (22)-(25) define the proposed EP stochastic image model. In order to have a tractable model, we assume that it is possible to disregard the variance of the data along the edge (Figure 12). In other words, we consider that edge amplitudes are slowly varying; each edge fragment R_2^i is consequently represented by its mean value μ_i. Therefore, instead of equation (25) we finally obtain:

$$R_2^i = \{\mathbf{x} : x[k] \sim N(\mu_i, 0)\}. \tag{26}$$

To verify the EP model effectiveness we incorporated it into the proposed non-decimated biorthogonal wavelet transform denoising algorithm (Figure 7). Due to the fact that the segmentation process is very sensitive to noise presence, we assuming that the positions of the R_2 subband samples are known. This test gives us the EP model *theoretical achievable upper bound* in the AWGN image denoising application for the given type of transform (like in the case of the EQ model [8]). It should be noted that the amount of side information in the case of the EP model is much lower than for the EQ upper bound.

The average PSNR experimental results for Lena and Barbara test images are presented in Table 2. Based on them it is possible to conclude that the denoising performance enhancement over the algorithm described in Section 4 (the last line in Table 1) is more than 1 dB on average for all tested noise variances.

6. CONLUSIONS AND FUTURE PERSPECTIVES

We have presented in this paper a new stochastic image model based on geometrical image prior information about the local image structure in the critically sampled and non-decimated wavelet transform domains. Data in high frequency subbands are modeled as locally stationary in some non-overlapping subregions. The region partition is performed using the quantization-based segmentation of the corresponding lowpass subband data. We have developed two versions of the denoising

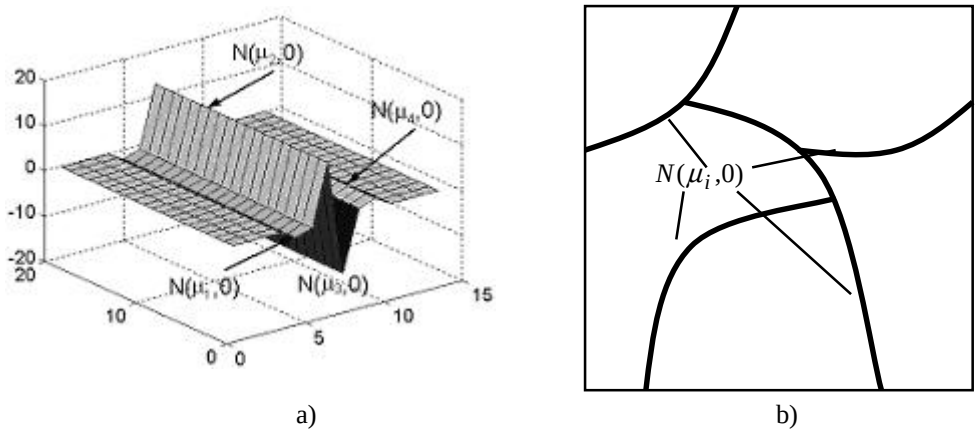


Figure 12. Graphical interpretation of the edge process model: (a) edge representation in the highpass subband; (b) correspondence to the model subband partition map.

Table 2. Upper theoretical bound for the EP model based image denoising (average PSNR, [dB]).

Denoising method	**Standard deviation of the noise**				**Transform type**
	10	15	20	25	**Overcomplete**
EP model performance upper bound	35.52	33.64	32.21	31.03	*9/7*
Performance enhancement in comparison to the proposed method from Section 4 (last line in Table 1)	1.11	1.27	1.23	1.17	*9/7*

algorithms, that exploit this model in critically sampled and overcomplete transform domains. Benchmarking has demonstrated that our approach is the best in the case of the critically sampled transform, and the second best in the non-decimated transform domain. The ranking in the latter case can be explained by low robustness of the segmentation to noise and to a more sophisticated structure of the best technique.

Aiming in enhancing denoising performance without increasing algorithmic computational complexity, the edge process stochastic image model was proposed as a way to decrease the residual correlation in the high frequency subbands. In the case of the EP model we treat data in the flat areas and in the edge regions in different ways: as non-stationary zero-mean Gaussian for flat areas, and as locally stationary non-zero mean Gaussian with very low variance for edges. To demonstrate the possible benefit obtained from using the EP model we have performed tests assuming that the edge region spatial position was available. In this case a gain of more than 1 dB in average PSNR was obtained for all tested AWGN variances.

As was mentioned, the main open issue of the EP model is the reliable estimation of the model parameters in the presence of noise. Therefore, we will concentrate on the solution to this problem in our ongoing research, and will exploit it for other applications such as image compression and watermarking [22] where noise is not a critical issue.

7. ACKNOWLEDGEMENTS

This work was partially supported by the Swiss SNF Grant No 21-064837.01 “Generalized stochastic image modelling for image denoising, restoration and compression” and by the Swiss Interactive Multimodal Information Management (IM2) project. The authors are thankful to Frederic Deguillaume and Yuriy Rytsar for helpful and interesting discussions.

8. REFERENCES

1. M.Nikolova, “Regularization functions and estimators”, in *Proc. IEEE Int. Conf. On Image Proc*, Lausanne, Switzerland, pp. 457-460, Sept. 1996.
2. J. Liu and P. Moulin, “Image denoising based on scale-space mixture modeling of wavelet coefficients,” in *Proc. IEEE Int.Conf. on Image Proc*., Kobe, Japan, Oct. 1999.
3. S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image de-noising,” in *Proc. IEEE Int. Conf. on Image Proc*., Chicago, IL, Oct. 1998.
4. A. Synyavskyy, S.Voloshynovskiy and I. Prudyus, “Wavelet-Based MAP Image Denoising Using Provably Better Class of Stochastic I.I.D. Image Models”, *Facta Universitatis* (Nis), 14, 2001.

5. P. Moulin and J. Liu, “Analysis of Multiresolution Image Denoising Schemes Using Generalized - Gaussian and Complexity Priors”, *IEEE Trans. Info. Theory*, Special Issue on Multiscale Analysis, v**ol. 45**, No. 3, pp. 909-919, Apr. 1999.
6. J. Liu and P. Moulin, “Information-Theoretic Analysis of Interscale and Intrascale Dependencies Between Image Wavelet Coefficients,” *IEEE Trans. on Image Processing*, Vol. 10, No. 10, pp. 1647-1658, Nov. 2001.
7. S. M. LoPresto, K. Ramchandran, and M. T. Orchard, “Image coding based on mixture modeling of wavelet coefficients and a fast estimation- quantization framework,” in *Proc. Data Compression Conf.*, Mar. 1997.
8. M. K. Mihcak, I. Kozintsev, and K. Ramchandran, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Signal Processing Letter*s, **vol. 6**, No. 12, pp. 300–303, Dec. 1999.
9. S.V. Vaseghi, *Advanced Signal Processing and Noise Reduction*, John Wiley & Sons, Inc., 2nd Edition, 1998, 498 p.
10. C. Chrysafis, Wavelet Image Compression Rate Distortion Optimizations and Complexity Reductions, PhD Thesis, University of Southern California, 2000.
11. R.R. Coifman and D.L.Donoho, “Translation invariant denoising”, In *Wavelets and Statistics, Springer Lecture notes in Statistics 103*, pp. 125-150, New York: Springer-Verlag.
12. Xin Li and Michael T. Orchard, “Spatially adaptive image denoising under overcomplete expansion”, in *Proc. of ICIP*, Vancouver, Canada, Sep. 2000.
13. J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, “Adaptive Wiener Denoising Using A Gaussian Scale Mixture Model In The Wavelet Domain”, in *Proc. 8th Intern. Conf. on Image Proc*., Thessaloniki, Greece. Oct. 7-10 2001.
14. E.P. Simoncelli, W.T. Freeman, E.H. Adelson and D.J. Heeger. “Shiftable Multi-Scale Transforms” [or, "What's Wrong with Orthonormal Wavelets"], *IEEE Trans. Information Theory*, *Special Issue on Wavelets*, **vol. 38**, No. 2, pp. 587-607, Mar. 1992.
15. http://www.ai.mit.edu/people/pff/seg/seg.html.
16. V. Strela, “Denoising via block Wiener filtering in wavelet domain”, in *3rd European Conference of Mathematics*, Birkhauser Verlag, Barcelona, July, 2000.
17. M.V. Wickerhauser, *Adapted wavelet transform: from theory to software*, A.K. Press, 1994, 504 p.
18. M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, “Image coding using wavelet transform”, *IEEE Trans. Image Proc.*, **vol. IP-1**, pp.205-220, Apr. 1992.
19. J. K. Romberg, H. Choi, and R. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden Markov models,” in *Proc. of SPIE*, Denver, CO, vol. 3816, pp. 31–44, July 1999.
20. S. Xiao, I. Kozintsev, and K. Ramchandran, “Stochastic wavelet-based image modeling using factor graphs and its application to denoising”, in *Proc. SPIE Conf. On Vis. Commun. and Image Proc.*, Jan. 2000.
21. G. Fan and X.-G. Xia, “Wavelet-based statistical image processing using Hidden Markov Tree model,” in *Proc. 34th Annual Conference on Information Sciences and System*s, Princeton, NJ, USA, March 2000.
22. S.Voloshynovskiy, O.Koval, F.Deguillaume, and T. Pun, , “Data hiding capacity analysis for real images based on stochastic non-stationary geometrical models”, in *IS&T/SPIE’s annual Symposium, Electronic Imaging 2003: Security and Watermarking of Multimedia Content*, Santa Clara, California USA, 20-24 Jan. 2003.