

# Image denoising based on the edge-process model\*

S. Voloshynovskiy, O. Koval, and T. Pun<sup>1</sup>

*University of Geneva - CUI, 24 rue du Général Dufour, CH-1211 Geneva 4,  
Switzerland*

*<http://watermarking.unige.ch/>*

*E-mail: {svolos, Oleksiy.Koval, Thierry.Pun}@cui.unige.ch*

---

## Abstract

In this paper a novel stochastic image model in the transform domain is presented and its performance in image denoising application is experimentally validated. The proposed model exploits local subband image statistics and is based on geometrical priors. Contrarily to models based on local correlations, or 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 an additive white Gaussian noise (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, spatially adaptive image behavior, geometrical priors and providing an accurate 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.

---

<sup>1</sup> Further information: (Send correspondence to S. Voloshynovskiy): E-mail: [svolos@cui.unige.ch](mailto:svolos@cui.unige.ch), <http://sip.unige.ch>.

\*Parts of this paper appeared as S. Voloshynovskiy, O. Koval and T. Pun, "Wavelet-based image denoising using non-stationary stochastic geometrical image priors", *In SPIE Photonics West, Electronic Imaging 2003, Image and Video Communications and Processing V*, Santa Clara, CA, USA, January 20-24 2003, and S. Voloshynovskiy, O. Koval, F. Deguillaume and T. Pun, "Data hiding capacity-security analysis for real images based on stochastic non-stationary geometrical models", *In SPIE Photonics West, Electronic Imaging 2003, Security and Watermarking of Multimedia Contents V*, Santa Clara, CA, USA, January 20-24 2003.

---

# Debruitage d'images base sur le modèle de traitement des contours

## Résumé

Dans cet article, un nouveau modèle stochastique pour les images dans le domaine transformée est présenté et ses performances pour les applications de débruitage sont expérimentalement validées. Le modèle proposé exploite les statistiques locales de l'image décomposée en sous-bandes et il est basé sur des primitives géométriques. Contrairement aux modèles basés sur des corrélations locales ou aux modèles mixtes, le modèle proposé effectue une partition de l'image en des régions disjointes avec des statistiques distinctes. Une proche solution analytique au problème de débruitage d'image pour un Bruit Blanc Gaussien Additif (BBGA) est obtenue et ses performance limites sont analysées. Malgré sa grande simplicité, le modèle stochastique proposé fournit de nombreux avantages en comparaison des approches existantes : (a) simplicité de la modélisation stochastique des images ; (b) complétude du modèle, offrant la multirésolution, un comportement spatialement adaptif, des primitives géométriques, un ajustement précis aux statistiques globales de l'image ; (c) très petite complexité de l'algorithme; (d) tractabilité du modèle et des résultats obtenus dus à une solution analytique et à l'existence de limites de performance analytiques; (e) extensibilité à différents domaines transformée tels que les représentations orthogonales, biorthogonales et redondantes.

---

*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. The artifacts arising due to imperfectness of these processes create obstacles to the perception of visual information by an observer. Thus, for image quality improvement, efficient denoising technique should be applied to compensate such annoying effects.

Development of such image denoising methods remains a valid challenge at the crossing of functional analysis and statistics. Depending on the underlying assumptions about the properties of the image two main research lines have

been developed targeting the solution to this problem. According to the first one [1], the image is mostly considered as a smooth (piecewise smooth) function both in the coordinate and different transform domains. Following this line, significant amount of deterministic denoising algorithms has been developed using various filtering techniques and variational calculus during last fifty years. For further details concerning development and performance analysis of the deterministic denoising techniques the interested reader is pointed to [2–8].

Within the stochastic framework, it is assumed that both image and noise are realizations of some random variables distributed according to some prior probability density functions (pdf). In this case, given contaminated (noisy) version of the original image, it is necessary to find its accurate estimate that is usually done using Bayesian approach.

Assuming independent identically distributed (i.i.d.) Gaussian statistics of the noise, it is possible to classify the developed Bayesian denoising approaches based on the underlying stochastic image model and on the type of the transformation used.

According to the first classification criterion, one can 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 [9], 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) [10,11] or Student law [12]. Following this direction we are coming to various types of thresholding and shrinking techniques that are optimal from the information-theoretic approach point of view as was shown by Moulin and Liu [13]. It is important to underline at the moment [8] that one of the best deterministic denoising techniques known as the Osher-Rudin-Fatemi algorithm [6] that is based on the  $(u + v)$  or “cartoon + texture” image model is similar to one of such shrinking techniques [14].

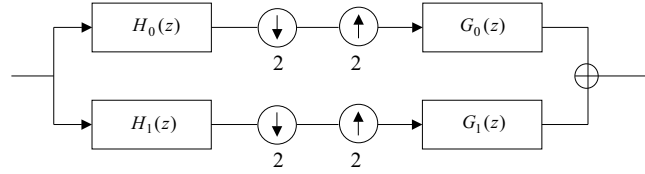
Another approach proposes to divide an image into two regions of edges and flat areas and to apply different models to describe these regions (i.i.d. GGd and Gaussian distribution correspondingly) [15]. The main drawback of these models is their inability to capture local statistics that play a crucial role in denoising applications. Moreover, by applying global permutation of image data we do not change their marginal properties under the i.i.d. assumption and therefore denoising based on global statistical models does not reflect the local image structure.

Another possible solution is inspired by lossy wavelet data compression where the so-called estimation-quantization (EQ) model was developed [16] and demonstrated to have superior performance in denoising application [17]. The main idea of the EQ model consists in representing the image data in wavelet

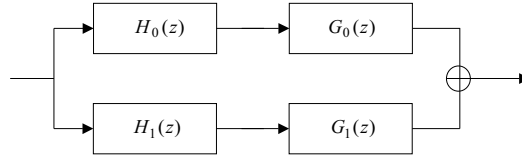
subbands as an i.i.d. Gaussian field with zero mean and slowly varying variance. Under such an assumption the best estimator will coincide with a classical scalar form of the Wiener filter. This strategy has given significantly better denoising performance (approximately 2 dB) than the previous class of models [17]. Here and in the following we use peak-signal-to-noise ratio (PSNR) for quantitative analysis of denoising algorithm performance. An attempt to combine the EQ strategy with the GGd prior image model was reported in [11]. However, the performance of this algorithm was worse than for the locally Gaussian prior case.

The second important classification criterion is the domain where the denoising procedure is applied. The simplest solution is to process the data in the coordinate domain directly. If the data are assumed to be i.i.d. stationary Gaussian, the denoising will coincide with a classical Wiener filter [18]. 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 in denoising applications was to use an orthogonal critically sampled wavelet transform (Figure 1,a) [10–12]. 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 efficient processing (for instance, only few coefficients in the 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 but the Gibbs phenomenon is observed in the reconstructed image near image borders (edges) [19].

A solution that overcomes this Gibbs artifacts problem of the critically sampled wavelet transform by eliminating its down- and upsampling stages, the so-called non-decimated or stationary wavelet transform, was firstly introduced by Coifman and Donoho [20] (Figure 1,b). Moreover, using biorthogonal filters [21], 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 [21]. The recent work of Sterla *et al.* [22] presented a more powerful transform with a larger amount of spatial orientations (steerable pyramid [23]) 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 about the stochastic model of the data that are supposed to be correlated in the transform domain. The data in this transform domain is characterized by a Gaussian mixture model [22], where a random vector  $\mathbf{a}$  can be expressed as the following product:  $\mathbf{a} \stackrel{d}{=} \mathbf{b}\sqrt{c}$  of a zero-mean Gaussian vector  $\mathbf{b}$  and an independent positive scalar  $\sqrt{c}$  and equality is with respect to the correspondent pdfs. In this case, a processing applied to denoise the image operates with covariance matrices rather than exploits correspondent local variances. Experimental results presented in [22] show that this method provides the best quality of the



(a)



(b)

Fig. 1. (a) Structure of 1-level critically sampled and (b) non-decimated wavelet transform; 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.

denoised image among existing Bayesian techniques in terms of PSNR.

Other authors have developed representations with similar properties. The details can be found in [24–27]. Examples of bidimensional multiresolution transforms in image denoising applications can be found in [28,29].

Another possibility of ringing artifacts reduction was recently proposed in [30], where a new geometrical image representation was introduced and applied to the denoising problem in the translation invariant set-up. This representation, known as bandlets, is based on the orthonormal wavelet basis function adaptation to the directions in which the image grey levels have regular variations.

Based on the above discussion, the main goal of this paper can be formulated as follows: to develop a stochastic image model that will allow to increase the Bayesian denoising algorithm performance in terms of the PSNR assuming i.i.d. statistics of the data and without using more complex transforms.

The structure of the paper is the following. In Section 2 a stochastic approach for the removal of the AWGN is reviewed and analyzed analytically. As a result, some problems when a 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 is then introduced, as well as the solution to these problems. In Section 4, three versions of the denoising algorithm based on geometrical priors are proposed to investigate the model effectiveness in the coordinate, orthogonal critically

sampled wavelet and biorthogonal non-decimated wavelet transform domains. Finally in Section 5, a new stochastic image model, the so-called *edge process model*, is introduced and investigated in the wavelet and in the overcomplete transform domains.

**Notations.** We use capital letters to denote scalar random variables  $X$ , bold capital letters  $\mathbf{X}$  to denote vector random variables, regular letters  $x$  and bold regular letters  $\mathbf{x}$  to designate the realizations of scalar and vector random variables, respectively. We use  $X \sim f_X(x)$  or simply  $X \sim f(x)$  to indicate that a continuous random variable  $X$  is distributed according to  $f_X(x)$ . The mean value of scalar and vector random variables are denoted  $\bar{x}$  and  $\bar{\mathbf{x}}$ , respectively. The covariance matrix of  $\mathbf{X}$  and variance of  $X$  are denoted as  $\mathbf{C}_{\mathbf{X}}$  and  $\sigma_X^2$ . We use  $\mathbf{I}_N$  to designate the identity operator of  $N \times N$  dimensionality. Calligraphic characters  $\mathcal{R}$  are used to indicate data set cardinalities. Also,  $\mathbf{x}$  is used to denote a two-dimensional sequence representing the luminance of the original image. We use in our notations the so-called lexicographical ordering  $\mathbf{x} = \{x[1], x[2], \dots, x[N]\}$ , where  $N = M_1 \times M_2$  is the size of the image. The  $i^{th}$  element of  $\mathbf{x}$  is designated as  $x[i]$  where  $i = M_1 \cdot n_1 + n_2$ ,  $1 \leq i \leq N$  and  $\mathbf{x} \in \mathbb{R}^N$ .

## 2 MAXIMUM A POSTERIORI PROBABILITY IMAGE DENOISING

### 2.1 Problem statement

In the scope of this paper we assume that the original image  $\mathbf{X}$  is corrupted by an additive noise  $\mathbf{Z}$ :

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}, \quad (1)$$

where  $\mathbf{Y}$  is the degraded noisy data and  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_Z)$  is an i.i.d. AWGN with zero-mean and known covariance matrix  $\mathbf{C}_Z = \sigma_Z^2 \mathbf{I}_N$ .

Therefore, the main task of this paper consists in the accurate recovery of  $\mathbf{X}$  from its noisy version  $\mathbf{Y}$ .

### 2.2 Maximum a Posteriori Probability estimator for the Gaussian model

It is known that the most accurate estimate of an original data based on its degraded version can be obtained taking into account available proper prior information. Therefore, a maximum a posteriori probability (MAP) estimate

was selected to perform denoising:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathbb{R}^N} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})p_{\mathbf{X}}(\mathbf{x}), \quad (2)$$

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

Assuming that both data and noise are multivariate Gaussian with respective means  $\bar{\mathbf{x}}$  and  $\mathbf{0}$  and respective covariance matrices  $\mathbf{C}_X$  and  $\mathbf{C}_Z$ , i.e.  $\mathbf{X} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{C}_X)$  and  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_Z)$  the solution to the problem (2) becomes:

$$\hat{\mathbf{x}} = \mathbf{C}_Z(\mathbf{C}_Z + \mathbf{C}_X)^{-1}\bar{\mathbf{x}} + \mathbf{C}_X(\mathbf{C}_Z + \mathbf{C}_X)^{-1}\mathbf{y}. \quad (3)$$

The MAP estimator produces some estimation error  $\tilde{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{x}$  that will also be Gaussian  $\tilde{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\tilde{X}})$ , since the MAP estimator is linear. The covariance matrix of the estimation error is defined as:

$$\mathbf{C}_{\tilde{X}} = E[(\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})^T] = \mathbf{C}_X(\mathbf{C}_X + \mathbf{C}_Z)^{-1}\mathbf{C}_Z, \quad (4)$$

and the variance of the MAP estimator is:

$$\sigma_{\tilde{X}}^2 = \frac{1}{N} \text{tr}[\mathbf{C}_{\tilde{X}}] \quad (5)$$

that is one of possible criteria for evaluating estimator performance.

### 2.3 MAP estimator performance analysis

If the original signal is uncorrelated but non-stationary  $\mathbf{C}_X = \text{diag}[\sigma_X^2[1], \sigma_X^2[2], \dots, \sigma_X^2[N]]$  and the noise is the AWGN,  $\mathbf{C}_Z = \sigma_Z^2 \mathbf{I}_N$ , the MAP estimator of  $\mathbf{x}$  is reduced to:

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{C}_X(\mathbf{C}_X + \sigma_Z^2 \mathbf{I}_N)^{-1}(\mathbf{y} - \bar{\mathbf{x}}), \quad (6)$$

or simply to:

$$\hat{x}[i] = \bar{x}[i] + \frac{\sigma_X^2[i]}{\sigma_X^2[i] + \sigma_Z^2}(y[i] - \bar{x}[i]), \quad (7)$$

$1 \leq i \leq N$ . In this case the variance of the MAP estimator is found as:

$$\sigma_{\tilde{X}}^2 = \frac{1}{N} \sum_{i=1}^N \frac{\sigma_X^2[i]\sigma_Z^2}{\sigma_X^2[i] + \sigma_Z^2}. \quad (8)$$

If the signal is additionally white Gaussian,  $\mathbf{C}_X = \sigma_X^2 \mathbf{I}_N$ , and the noise is the AWGN,  $\mathbf{C}_Z = \sigma_Z^2 \mathbf{I}_N$ , the MAP estimator of  $\mathbf{x}$  is simply one-to-one mapping:

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \sigma_X^2 \mathbf{I}_N (\sigma_X^2 \mathbf{I}_N + \sigma_Z^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \bar{\mathbf{x}}) \quad (9)$$

or simply:

$$\hat{x}[i] = \bar{x}[i] + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Z^2} (y[i] - \bar{x}[i]), \quad (10)$$

$1 \leq i \leq N$ . In this case the variance of the MAP estimator is found as:

$$\sigma_{\hat{X}}^2 = \frac{\sigma_X^2 \sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}. \quad (11)$$

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, for one dimensional signal this can be performed using a maximum likelihood (ML) strategy:

$$\sigma_X^2[i] = \frac{1}{|\Omega|} \sum_{k \in \Omega(i)} (x[k] - \bar{x}[i])^2, \quad (12)$$

where we use a square window  $\Omega(i)$  centered at  $x[i]$  with  $|\Omega|$  to be the number of coefficients in  $\Omega(i)$  and  $\bar{x}[i]$  is the local mean in  $\Omega(i)$ :

$$\bar{x}[i] = \frac{1}{|\Omega|} \sum_{k \in \Omega(i)} x[k]. \quad (13)$$

### 3 A NEW STOCHASTIC 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. One can 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 with some variance [31] (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  $\bar{x}_1, \bar{x}_2$  and variances  $\sigma_1^2, \sigma_2^2$ , respectively, then the joint probability density function (pdf) is:

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

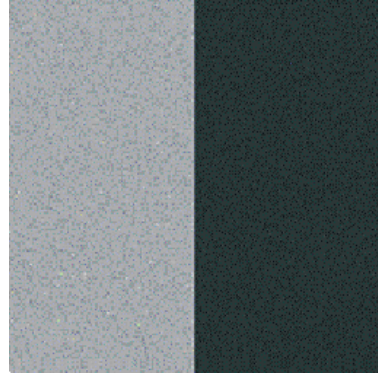
where  $\alpha$  is the first component fraction in the mixture. Then, for the mixture distribution we have the corresponding mean and variance:

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

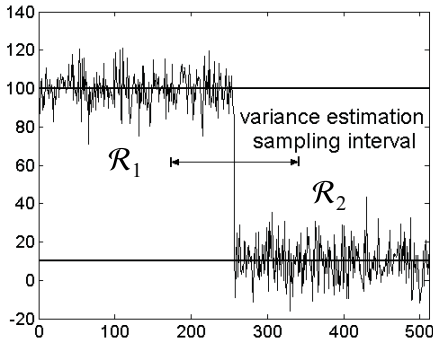
$$\sigma^2 = E[X - \bar{x}]^2 = \alpha \sigma_1^2 + (1 - \alpha) \sigma_2^2 + \alpha (\bar{x} - \bar{x}_1)^2 + (1 - \alpha) (\bar{x} - \bar{x}_2)^2. \quad (16)$$



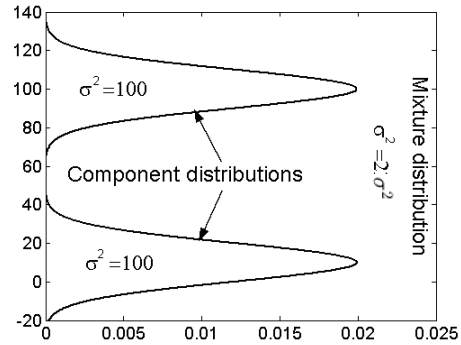
(a)



(b)



(c)



(d)

Fig. 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,  $\alpha = 0.5$ , then the over-estimation ratio (estimated-to-correct variance ratio), for the case of above selected variances, is more than 20 times.

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

$$\cup_i \mathcal{R}_i = S; \mathcal{R}_i \cap \mathcal{R}_j = \emptyset; i = 1, 2, \dots, v. \quad (17)$$

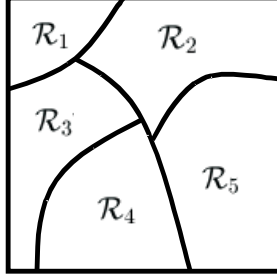


Fig. 3. Image partition as a set of regions with i.i.d. homogeneous statistics.

Let  $\mathbf{x}_i$  denotes the subset of image pixels supported by the region  $\mathcal{R}_i$ . In our model we assume that each region is fully covered by the probabilistic model  $\theta \in \{\Theta_1, \Theta_2, \dots, \Theta_v\}$  and no two neighboring regions are described by the same model. In particular, we assume that the pixels in the image sub-region  $\mathbf{x}_i$  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 transform). In the following, we only concentrate on Gaussian models, meaning  $p_{\mathbf{X}}(\mathbf{x}_i|\theta_i) = \mathcal{N}(\bar{\mathbf{x}}_i, \sigma_i^2 \mathbf{I})$ . This model is different from the classical Gaussian mixture model used for example as the background assumption for the EQ model, since it supposes the homogeneity of the data within the same region, while the Gaussian mixture model allows the presence of samples with different statistics as well as from the Osher-Rudin-Fatemi model [6] in such a way that no smoothness assumption within  $\mathcal{R}_j, j \in 1, 2, \dots, v$ , is required.

According to our model, to perform a correct ML-based local variance estimation one needs to take into account only those coefficients in the transform domain that belong to the subset  $\Omega^*$  of the local sampling space  $\Omega$  that contains the elements from the corresponding subregion  $\mathcal{R}_i$ , 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:

$$\hat{\sigma}_X^2[i] = \frac{1}{|\Omega^*|} \sum_{k \in \Omega(i)} x[k]^2 m_k, \quad (18)$$

where  $m_k$  is a subset indicator function,  $m_k = \begin{cases} 1 & \text{if } m_k \in \Omega^*, \\ 0 & \text{otherwise} \end{cases}$ ;  $|\Omega^*|$  is the

cardinality of the subset  $\Omega^*$ , and the dimensionality of  $\Omega$  (dimensionality of a sampling window centered at  $x[i]$ ) is selected to be large enough to guarantee reliable estimation.

Such a partition is targeting region mean separation and satisfaction of the assumed Gaussian statistics within the defined regions. For this purpose we exploited segmentation software [32] developed by Cornell University with empirically adjusted parameters (Figure 5,b,c).



Fig. 4. (a) Fragment of test image Lena; (b) local estimation window with indication of the subset (black pixels) and of the estimated pixel (gray) for the restricted support ML local variance estimation.

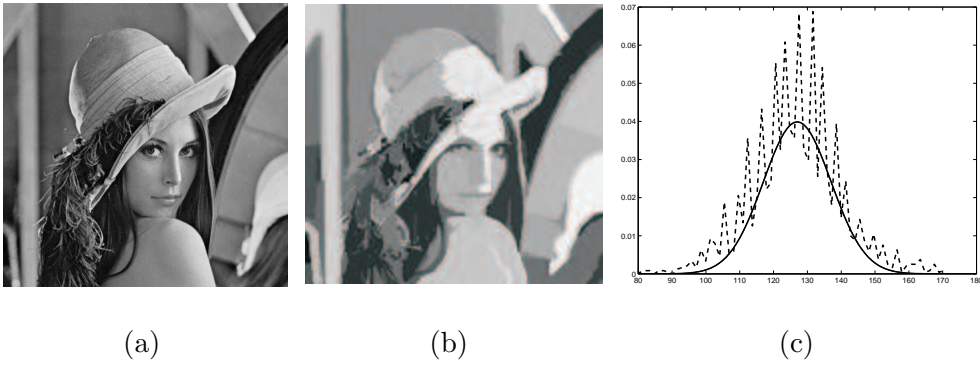


Fig. 5. Segmentation results: (a) test image Lena; (b) segmented image Lena using Cornell University segmentation software; (c) statistics of image data in one of the selected regions: real histogram (dashed line) and Gaussian pdf (solid line).

### 3.2 Transform domain setup

If one transforms the data to the wavelet or overcomplete domain, the variance over-estimation compensation problem still exists. Since in this case the data mean value in the transform domain is equal to zero, the current set-up can be modeled using (Figure 4,a). To justify the introduced stochastic model, we propose to study the following two-texture example (Figure 6,a). In this case there are two components with zero mean and different variances (100 and 25) in the sampling space. 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 adaptive ML procedure (18) are needed to eliminate variance over-estimation.

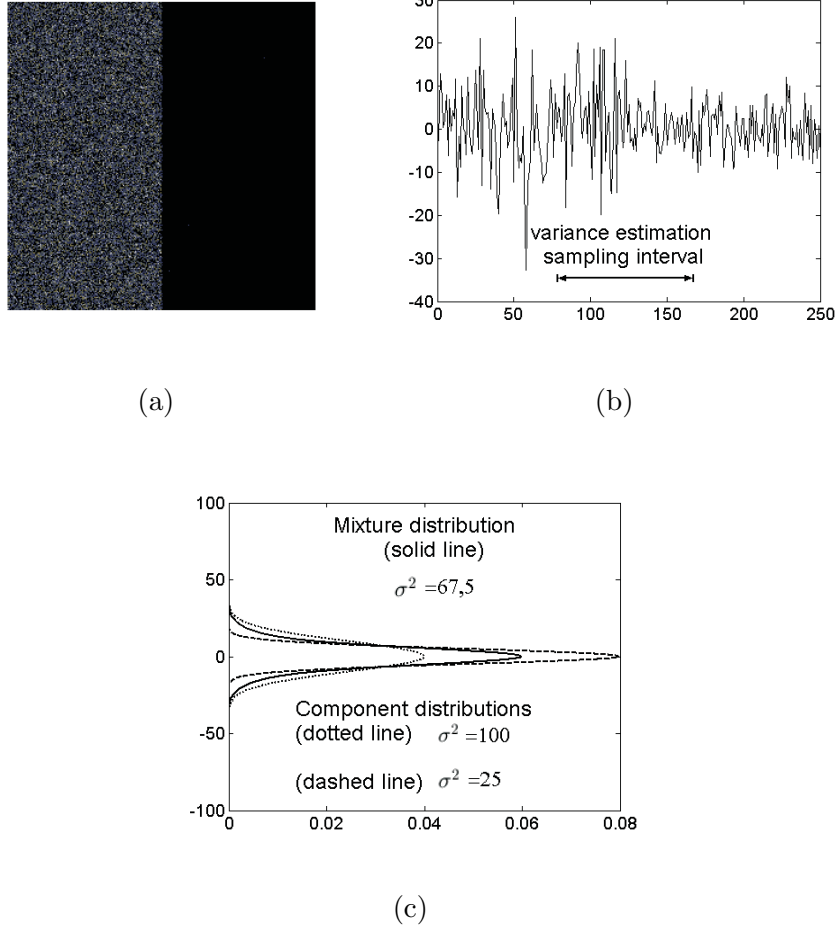


Fig. 6. Variance estimation using classical ML strategy for the edge image region: (a) original image; (d) one-dimensional original image fragment; (e) correspondent individual components and resulting pdfs with estimated variances.

Therefore, the difference with the model definition in the coordinate domain is that a particular sample from a region  $\mathcal{R}_i$  covered by a particular model  $\theta_i \in \{\Theta_1, \Theta_2, \dots, \Theta_v\}$  that is assumed to be a zero-mean Gaussian pdf  $p_X(x_j|\theta_i) = \mathcal{N}(0, \sigma_i^2[j])$ . Thus, non-stationary data behavior in regions is assumed according to the parallel source splitting paradigm [33] where globally non-Gaussian data might be considered as locally Gaussian with non-stationary variance.

#### 4 ALGORITHM IMPLEMENTATION AND BENCHMARKING

According to the benchmarking results summarized in [34], 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.  $X[k] \sim \mathcal{N}(\bar{x}[k], \sigma_X^2[k])$ , denoising will consist in the application of the Wiener filtering based on the ML local variance estimate, using the region partition map (Figure 5) as a 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 (correspondingly, 29.41 dB and 29.30 dB).

To prove the efficiency of the proposed model in practice we developed two versions of the AWGN removal algorithm for the critically sampled and 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 the side information (geometrical prior (GP) information) about regions  $\mathcal{R}_i$ , and inverse transform block (IT). The critically sampled wavelet transform is implemented using the orthogonal Beylkin filter pair [35] (5 decomposition levels were selected as in [17]). 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 based on the previously denoised data from the previous decomposition level. 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  $5 \times 5$  to  $11 \times 11$  depending on the decomposition level.

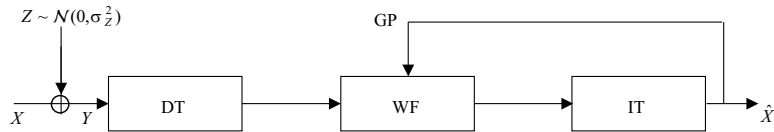


Fig. 7. Structure of the proposed denoising algorithm in the transform domain.

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). It is important to note that, contrarily to the coordinate domain set-up, in the transform domain case validity of local Gaussianity that corresponds to the proposed model, is justified by the parallel source splitting and, therefore, the role of segmentation consists only in region separation.

For optimization, we performed a number of tests to establish the optimal number of quantization bins. It was determined that from the point of view of the denoised image quality 4 bins was the best choice. Using this simpler segmentation, we obtained denoising results similar to those attained using the Cornell University segmentation software (performance loss in terms of the denoised image quality was at most 0.06 dB), while reducing segmentation time by a factor of 2.5. Therefore, all following experiments were obtained using this quantization-based segmentation.



Fig. 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 scheme presented in Figure 1.b, the transform was accomplished using 9/7 CDF [36] biorthogonal filter pair and 4 levels of decomposition. In addition, we took into account the fact that coefficient variances in high frequency subbands  $\Sigma_X^2$  are distributed according to a marginal distribution that can be very closely approximated by the Rayleigh law (Figure 9):

$$p_{\Sigma_X^2}(\sigma_X^2) = \frac{\sigma_X^2}{s^2} \exp\left(-\frac{(\sigma_X^2)^2}{2s^2}\right), \quad (19)$$

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

$$\sigma_{MAP}^2 = \frac{s^2(|M| - 1)}{2} \left[ -1 + \sqrt{1 + \frac{4}{s^2(|M| - 1)} \sum_{i=1}^{|M|} (x[i])^2} \right], \quad (20)$$

where  $|M|$  is a cardinality of the sampling window. The scale parameter  $s$

is estimated in each subband using the ML estimation in a  $3 \times 3$  sampling window:

$$s^2 = \frac{1}{2N} \sum_{i=1}^N \sigma_{3 \times 3}^2 [i]. \quad (21)$$

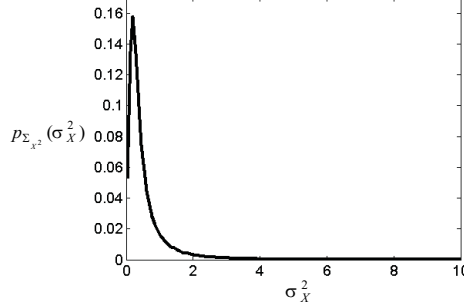


Fig. 9. Histogram of the variance of the high frequency subband (first decomposition level, diagonal orientation) of test image Lena in the overcomplete transform domain.

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 subbands from the fourth decomposition level). A complete sampling space dimensionality  $|M|=15 \times 15$  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 a set of twelve 8 bit  $512 \times 512$  test images for 100, 225, 400, 625 noise variances of the AWGN. Due to the fact that only two standard test images Lena and Barbara are used for experimental validation of the most of existing Bayesian denoising algorithms, for the fair comparison purpose, only results for these images are presented in this paper and compared with the best Bayesian denoising techniques. Due to the fact that none of the candidates is simultaneously 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 among known Bayesian techniques, but for the case of the overcomplete domain the method proposed in [22] 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 spatial directions (steerable pyramid transform [23]), and to the assumption that data in the linear transform domain are still correlated. A more complex 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 [22] provides poorer quality (in terms of the PSNR) of the denoised image than the technique proposed in this paper.

Table 1

Comparison of average PSNR [dB] for several denoising methods and both test images.

Denoising method	Noise standard deviation				Transform type
	10	15	20	25	<b>Wavelets</b>
<b>Noisy image</b>	28.13	24.63	22.10	20.17	Db8
Liu & Moulin [10]	33.61	31.30	—	—	Db8
Mihcak <i>et al.</i> [17]	33.52	31.32	29.81	28.64	Db8
Chang <i>et al.</i> [11]	—	30.51	29.07	28.01	Symmlet 8
Romberg <i>et al.</i> [37]	32.89	30.58	29.09	28.01	Db8
Xiao <i>et al.</i> [38] ( <i>HMM</i> )	33.49	31.30	29.80	28.68	Db8
Xiao <i>et al.</i> [38] ( <i>AHMF</i> )	33.58	31.41	29.92	28.80	Db8
Fan & Xia [39]	33.55	31.35	29.90	28.75	Db8
<b>Proposed method</b>	<b>33.68</b>	<b>31.51</b>	<b>29.92</b>	<b>28.84</b>	Beylkin
					<b>Overcomplete</b>
Chang <i>et al.</i> [11]	—	31.82	30.31	29.12	Symmlet 8
Li & Orchard [21]	34.12	32.04	30.55	29.40	CDF 10/18
Fan & Xia [39]	34.30	32.20	30.70	29.55	Symmlet 8
Portilla <i>et al.</i> [22]	34.38	32.39	31.01	29.95	Steerable
<b>Proposed method</b>	<b>34.41</b>	<b>32.37</b>	<b>30.98</b>	<b>29.86</b>	CDF 9/7

In order to complete experimental validation of the developed algorithms, denoising results obtained using these techniques versus those one presented in [17] and [22] are given in Figures 10 and 11 for visual quality comparison.

## 5 DENOISING BASED ON EDGE PROCESS MODEL

Based on the experimental results of Section 4, one can conclude that to enhance the denoising performance it is necessary to take into account local data relationships 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



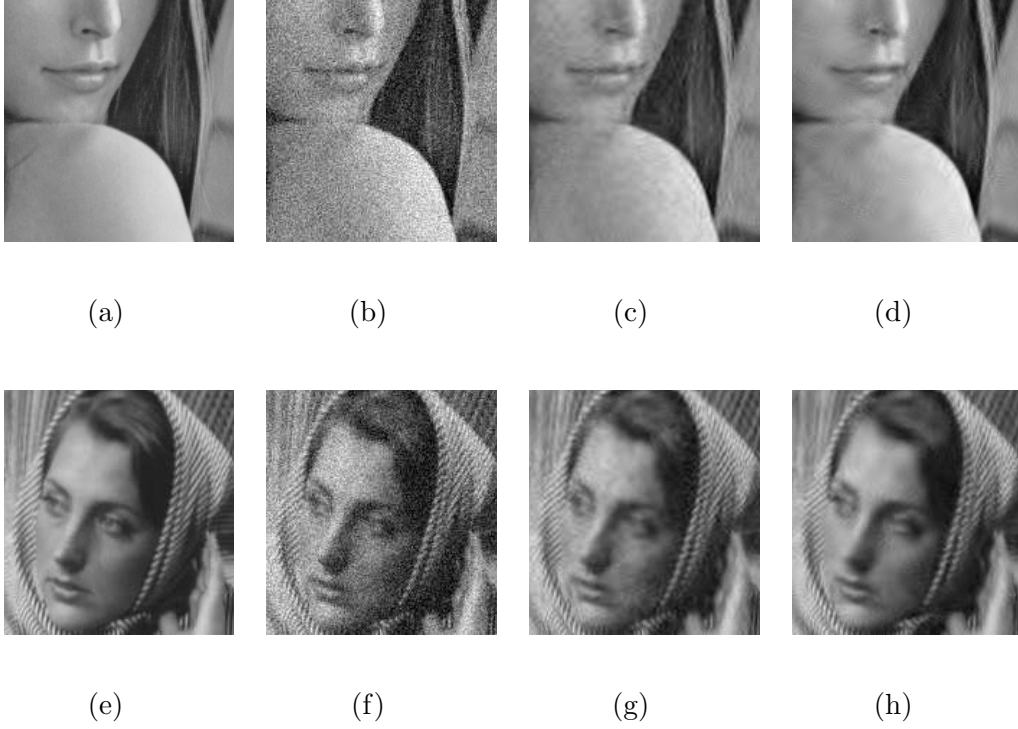


Fig. 10. Experimental results: (a) and (e) fragments of original test images; (b) and (f) the same fragments corrupted by zero-mean AWGN with  $\sigma_Z^2 = 400$ ; (c) and (g) DWT domain denoising results; (d) and (h) DOT domain denoising results.



Fig. 11. Denoising results of test image fragments (Figure 10,a,e) corrupted by zero-mean AWGN with  $\sigma_Z^2 = 400$  (from Figure 10,b,f): (a) and (b) denoised by Michak *et al.* [17]; (c) and (d) denoised by Portilla *et al.* [22].

because no linear transform is able to completely decorrelate the edges of real images. This phenomenon is illustrated in Figure 10 where a simple example of step edge (Figure 12, a) is transformed using non-decimated wavelet transformation (Figure 12,b). Therefore, if one finds a way to completely “remove” the edges from the subband data, this will allow an increase in performance by providing additional decorrelation (Figure 12,c).

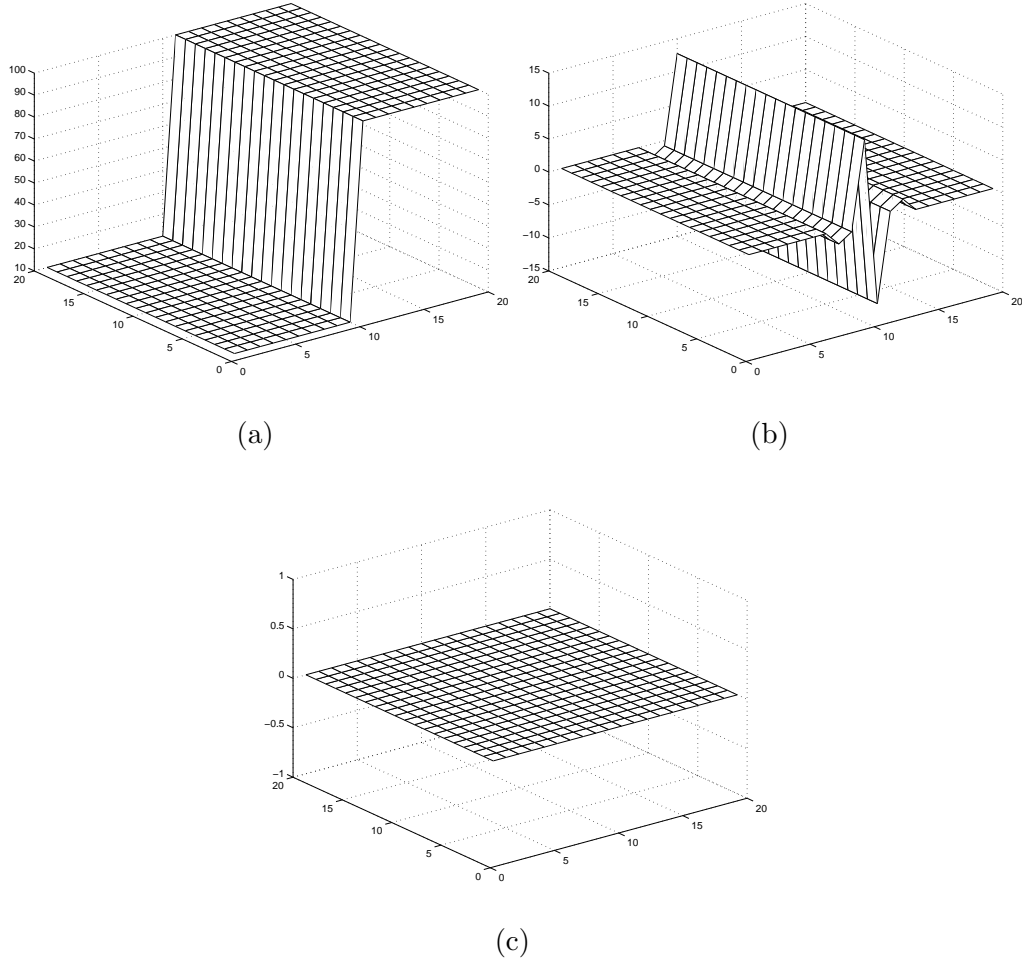


Fig. 12. Edge parameter estimation and edge subtraction from the high frequency subband data: (a) original data representing a test step edge in the coordinate domain; (b) non-decimated transform of the edge data; (c) subtracted edge data.

As a practical example, let us consider the fragment of Lena test image (Figure 2,a) in the overcomplete transform domain shown as a 3-D plot in Figure 13,a with the correspondent histogram presented in Figure 13,b. Application of the estimation and subtraction of the edge information detected using the proposed quantization based segmentation applied directly to the high frequency subband, from the original data leads to the following three main results (Figure 13,b-d):

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

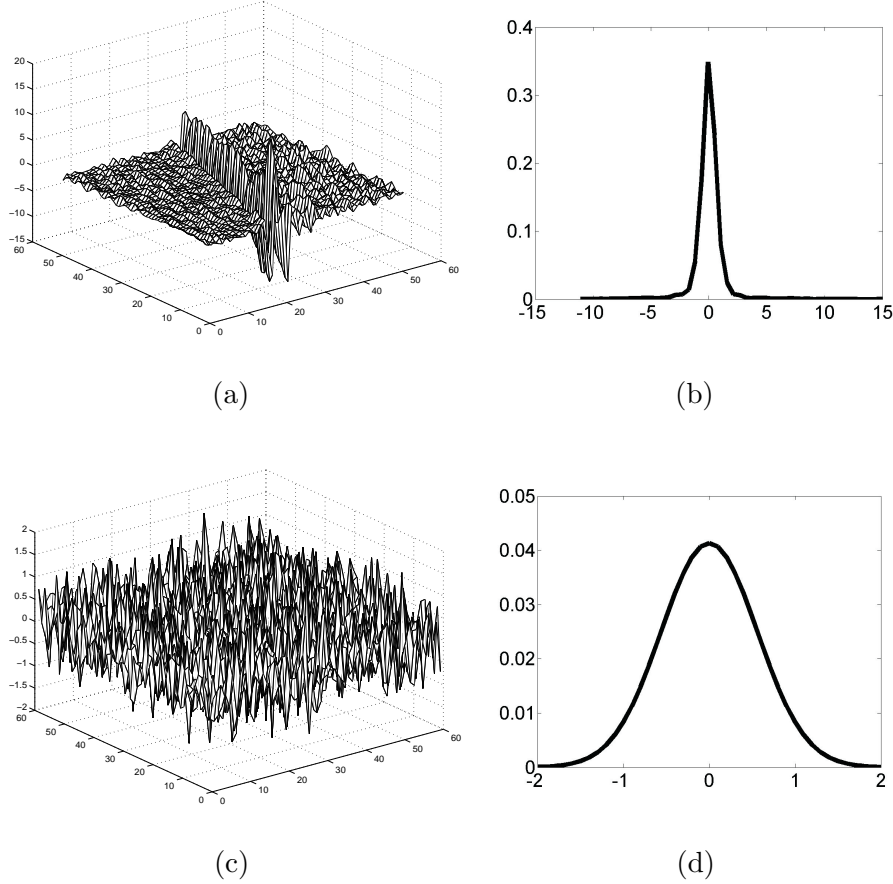


Fig. 13. Edge subtraction from the high frequency subband data: (a) original data with the variance  $\sigma^2=1.81$ ; (b) correspondent histogram; (c) subtracted edge data with the variance  $\sigma^2=0.31$ ; (d) correspondent histogram.

### 5.1 EP model definition for real images

Our goal is to introduce the *edge process model* (EP) and to compare it with the EQ model. The EQ model belongs to the class of intraband stochastic image models and assumes that the wavelet coefficients are Gaussian (in the original paper of Lopresto *et al.* the Generalized Gaussian [16]) distributed, with zero mean and variances that depend on the coefficient location within each subband. It is also assumed that the variance is slowly varying.

We assume that each subband of the multiresolution critically sampled transform has its own support  $S_l$ ,  $l = 1, \dots, 3W$ , where  $W$  is the number of dyadic decomposition levels such that  $S_i \cap S_j = \emptyset, i \neq j$  and  $\cup_l S_l = S$ . For non-decimated wavelet transform without downsampling used in our modeling, each subband has the same support as above but the dimensionality of each  $S_l$  is the same as the original image. According to the partition approach applied to the EQ model, we assume that only one region  $\mathcal{R}_l$  is given within the

subband  $S_l$  and all coefficients in this subband belong to the same region  $\mathcal{R}_l$ :

$$\mathcal{R}_l = \{\mathbf{x}_l : X_l[i] \sim \mathcal{N}(0, \sigma_{X_l}^2[i])\}, \quad (22)$$

i.e. all coefficients are considered to be independently distributed with zero-mean Gaussian distribution and different local variances  $\sigma_{X_l}^2[i]$ . Equivalently, this means that only one stochastic model out of  $\theta_l \in \{\Phi_1, \Phi_2, \dots, \Phi_L\}$  is applied to the whole support  $S_l$  and  $p_{X_l}(\mathbf{x}_l | \theta_l)$  follows an i.i.d. Gaussian pdf. In the following, we will only consider the image model for one subband. Contrarily to the EQ model, the proposed edge process model assumes two distinctive sets of coefficients in wavelet domain for each subband, i.e. those belonging to the flat regions and those belonging to the edge and texture regions. Moreover, it is assumed that a transition corresponding to an edge or to a fragment of texture consists of several distinct mean values that propagate along the transition (Figure 12). In the following we will refer to the transition simply as the edge. These mean values could be considered as the reconstruction levels of a scalar uniform threshold quantizer (UTQ) designed for the given subband  $S_l$  that is characterized by the global Generalized Gaussian ( $\mathcal{GGD}$ ) pdf  $X_l \sim f_{X_l}(x_l)$  and  $f_{X_l}(x_l) = \mathcal{GGD}(\mu_l, \gamma_l, \lambda_l)$  with mean  $\mu_l = 0$ , shape parameter  $\gamma_l$  and scale  $\lambda_l$ :

$$\bar{x}_j^l = \frac{\int_{d_j-1}^{d_j} x_l f_{X_l}(x_l) dx_l}{\int_{d_j-1}^{d_j} f_{X_l}(x_l) dx_l}, \quad (23)$$

where  $d_j - 1 \leq \bar{x}_j^l \leq d_j$  and  $\{d_j\}$  are decision levels of the UTQ. It is assumed that the UTQ has decision levels uniformly-spaced and the reconstruction levels are selected to minimize a mean square error (MSE). The dead-zone of UTQ  $[-T; T]$  is chosen to be equal to  $2\Delta$ , where  $\Delta$  is a quantization step-size between  $\{d_j\}$ .

The variation of the coefficients with the same mean is supposed to be low along the edge. According to the above introduced amplitude-based partition approach, the edge process model is defined as:

$$\mathcal{R}_1 = \{\mathbf{x} : X[i] \sim \mathcal{N}(0, \sigma_X^2[i])\}, \quad (24)$$

$$\mathcal{R}_2 = \{\mathbf{x} : X_j[i] \sim \mathcal{N}(\bar{x}_j[i], \sigma_{X_j}^2[i])\}, \quad (25)$$

where  $\mathcal{R}_1 \cup \mathcal{R}_2 = S$  and  $S$  represents a particular subband. The equation (25) assumes the proper separation of the regions with the distinctive statistics according to the case **B** (Figure 14). If the width of the bin  $\Delta$  is chosen to be relatively small in comparison with the flatness of the pdf, one can use the uniform approximation of the region  $\mathcal{R}_2$  statistics (case **C** in Figure 14) with

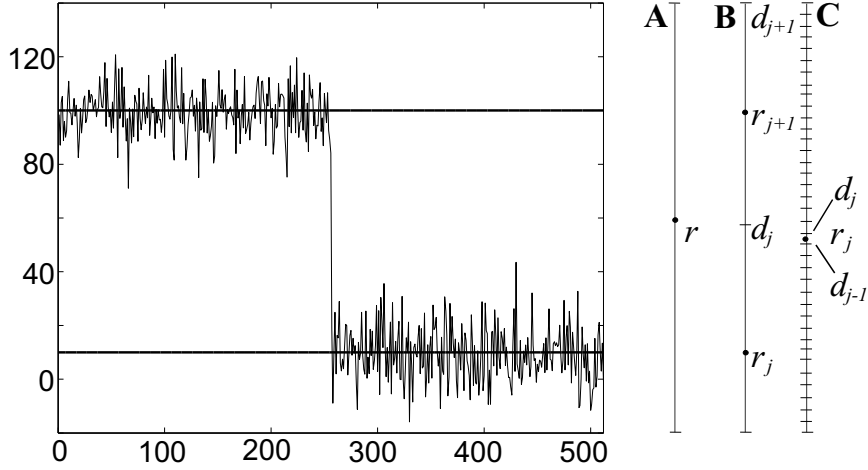


Fig. 14. Synthetic 1D example of non stationary mean data and amplitude-based classification.

mean  $\bar{x}_j[i]$  and interval  $(-\Delta/2, +\Delta/2)$ :

$$\mathcal{R}_2 = \{\mathbf{x} : X_j[i] \sim \mathcal{U}(\bar{x}_j[i], -\Delta/2, +\Delta/2)\}. \quad (26)$$

However, similarly to one-dimensional case we are not looking for the best possible quantized version approximation of the data but rather one aims at receiving the proper region separation. Therefore, the high-rate mode of quantization is out of interest in our formulation.

The region  $\mathcal{R}_1$  represents all flat regions within a subband assumed to be zero-mean Gaussian random variables with the local variance  $\sigma_X^2[i]$ . The region  $\mathcal{R}_2$  corresponds to the texture and edge regions. Each distinctive geometrical structure corresponding to the edge or texture transition within  $\mathcal{R}_2$  is decomposed into a set of local mean constellations. Moreover, a particular mean value  $\bar{x}_j[i]$ ,  $j = 1, \dots, J$  propagates along the edge creating the so-called *edge process*. Therefore, the coefficients on the edge are considered to have one of the possible mean values from the set  $\{\bar{x}_j[i]\}$ , contrarily to the EQ model which does not differentiate flat and edge regions and assumes zero-mean for all coefficients. Moreover, we can also assume that the variation of the coefficients with respect to the mean values (and this is especially true for the overcomplete transform) is very small. Therefore, the EP model assumes that the image consists of flat regions that are considered to be i.i.d. stationary Gaussian with almost “deterministic” edge occlusions which have a clearly defined geometrical structure depending on the mutual orientation of the edge and the subband basis function. Moreover, normally transitions along the edge have longer stationary length than the transitions within the texture (that explains the existence of higher correlations along the edges); this provides higher redundancy of the support for more accurate model parameter estimation. Due to this fact, the stationarity condition is more strict for the

edges than for the textures. Finally, all this leads to the conclusion that the real variance of the subbands is very low and is mostly determined by the flat regions and by the accuracy of edge shape approximation, contrarily to the EQ model or even more to the spike model [40] where the huge spikes of image coefficients with large variance can occur due to the edge that is supposed to model the wavelet coefficients' sparsity. No relationship or special geometrical spatial structure is assumed among the spikes in spike model contrary to the EP model where the "spikes" belonging to the same edge are treated jointly along the direction of edge propagation.

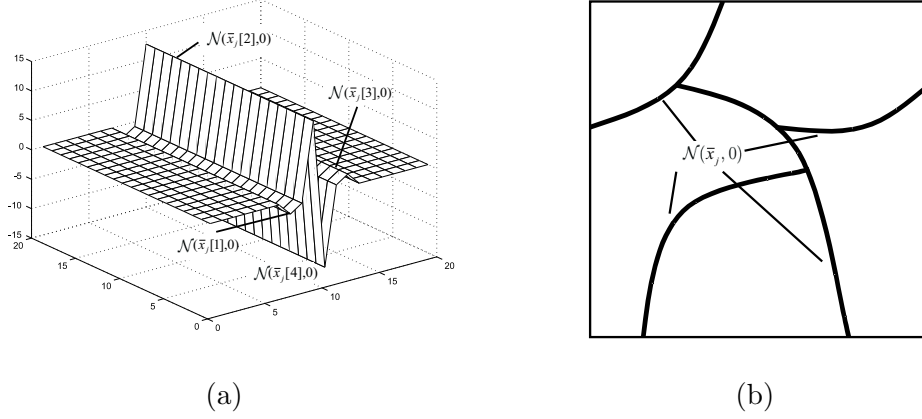


Fig. 15. Graphical interpretation of the edge process model: (a) edge representation in the highpass subband; (b) correspondence to the model subband partition map.

## 5.2 EP model verification on real images

To verify the EP model efficiency we incorporated it into the proposed non-decimated biorthogonal wavelet transform denoising algorithm (Figure 7). Since the EQ model is used in the denoising method of Mihcak *et al.* [17], we have chosen this method for the sake of comparison.

Both the EQ and EP models belong to the Gaussian family of distributions with only difference in the estimation of model parameters. In this case, the resulting estimate will be in the form of Wiener filter and the corresponding variance of the estimator can be found as:

$$\sigma_{MAP}^2 = \frac{1}{N} \text{tr} \left[ E[(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^T] \right] = \frac{1}{N} \sum_{i=1}^N \frac{\sigma_X^2[i] \sigma_Z^2}{\sigma_X^2[i] + \sigma_Z^2}, \quad (27)$$

where  $\sigma_X^2[i]$  is the local image variance of the EQ or EP models. Obviously, the lower image variance, the lower variance of the estimator.

Moreover, to avoid any subjective implementation issues (extraction of edges)

of a particular model parameter estimation we have used an *empirical upper bound* [41] as a criterion for the comparison. The empirical upper bound assumes that the stochastic model parameter estimation is performed based on the clean original image. In the case of the EQ model it refers to the estimation of the local image variances for each subband. We have investigated two modifications of the EQ model, either using a fixed window for all coefficients, or using a bootstrap version of the EQ denoiser described by Mihcak *et al.* [17]. In the case of the EP model, we performed three sets of tests aiming at investigating transform influence on the performance of denoising and at the establishing the power of prior knowledge of model parameters. In all cases, the parameters of the model (positions of the edge coefficients, local means magnitudes and local variances of flat regions) were estimated from the original image high frequency subbands using the proposed quantisation-based separation technique. Experimental results are presented in Table 2 for two images Lena and Barbara and compared to those obtained for the EQ model.

The denoised test images in both DWT and DOT domains for the situation when the AWGN variance is equal to 400 and side information about positions and means of the edge coefficients is available while the local variances are estimated from the noisy data are presented in Figure 16.

In case of the EP model, we assume three possible situations of different model priors available at the estimator (denoiser), i.e. when only positions of the EP coefficients are known and the mean values and the variances are estimated from noisy images directly; when both positions and the means values are known and the variances are estimated; and when both positions and the means and the variances are available. The image denoising was performed in two transform domains, i.e. DWT (*Db8*) and DOT (*9/7*). In case of the DWT domain, the EP model was applied to the first decomposition levels while the Wiener filter was applied to the rest ones. In case of the DOT, the EP model was applied to all decomposition levels. The results of the EP model performance in the DWT domain indicate that the EP model outperforms the EQ model in terms of empirical upper bound for all types of prior side information. The increase of amount of side information at the denoiser increases the EP model performance advantage from about 1 dB up to about 2 dB over the EQ model. The same results are obtained for the DOT where the performance gap is increased up to 1.5-6.0 dB. This means that not only model selection is important, but also that the transform domain could have a significant impact on the denoising problem. The obtained results also justify the performance of the EP model parameter estimation and power of prior knowledge in different domains. In particular, the knowledge of the EP means plays an important role for the critically sampled DWT domain where the amount of data is reduced in comparison with the DOT domain. Therefore, one can observe practically the same denoiser performance in the DOT for the situation when only positions are available and when both the positions and

the means computed based on the mean MSE centroid estimate from the clean subbands are available. The ML-estimation of means demonstrates that the means estimated from the noisy subbands and those estimated from the clean subband pdf practically coincide. This indicates that it suffices to estimate the means from the noisy image without loss of the performance. The situation is different for the critically sampled DWT where the reduction of the sampling space reduces the accuracy of mean estimate and thus the availability of the clean mean constellations enhances the denoiser performance. Finally, the additional prior information about local variances increases the performance gap between the EQ and EP models even more drastically. In conclusion, the usage of the EP model as opposed to the EQ model is justified according to this reference application.



Fig. 16. Denoising results of the fragments of Lena and Barbara test images corrupted by zero-mean AWGN with  $\sigma_Z^2 = 400$  when side information information about edge coefficient positions and means is available: (a) and (b) DWT domain set-up; (c) and (d) DOT domain set-up.

Based on the presented benchmarking results 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.

Despite very promising theoretical results, one should admit that reliable estimation of the edge information (especially, the positions of transition sample) is an open and challenging problem in denoising application and will be considered in our future research.

## 6 Conclusions 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



Table 2

Comparison of empirical bounds for the EP and EQ models in a denoising application, based on resulting PSNR [dB].

Image model	Noise standard deviation				Wavelet type
	10	15	20	25	
	<b>L E N A</b>				
Noisy image	28.13	24.63	22.10	20.17	
EQ (fixed window)	34.44	32.21	30.50	29.18	<i>Db8</i>
EQ ( [17], bootstrap)	34.58	32.59	31.17	30.06	<i>Db8</i>
EP (positions)	<b>35.46</b>	<b>33.42</b>	<b>31.91</b>	<b>30.73</b>	<i>Db8</i>
EP (positions and means)	<b>35.93</b>	<b>33.93</b>	<b>32.36</b>	<b>31.21</b>	<i>Db8</i>
EP (positions, means, variances)	<b>36.24</b>	<b>34.37</b>	<b>32.88</b>	<b>31.77</b>	<i>Db8</i>
EP (positions)	<b>37.92</b>	<b>36.34</b>	<b>35.08</b>	<b>33.97</b>	<i>9/7, overcomplete</i>
EP (positions and means)	<b>37.92</b>	<b>36.36</b>	<b>35.07</b>	<b>34.03</b>	<i>9/7, overcomplete</i>
EP (positions, means, variances)	<b>38.37</b>	<b>37.03</b>	<b>36.06</b>	<b>35.16</b>	<i>9/7, overcomplete</i>
	<b>B A R B A R A</b>				
Noisy image	28.14	24.63	22.11	20.18	
EQ (fixed window)	32.97	30.45	28.73	27.38	<i>Db8</i>
EQ ( [17], bootstrap)	32.84	30.46	28.86	27.65	<i>Db8</i>
EP (positions)	<b>33.74</b>	<b>31.80</b>	<b>30.36</b>	<b>29.31</b>	<i>Db8</i>
EP (positions and means)	<b>34.22</b>	<b>32.32</b>	<b>30.97</b>	<b>29.80</b>	<i>Db8</i>
EP (positions, means, variances)	<b>34.42</b>	<b>32.61</b>	<b>31.34</b>	<b>30.38</b>	<i>Db8</i>
EP (positions)	<b>36.66</b>	<b>35.26</b>	<b>34.19</b>	<b>33.30</b>	<i>9/7, overcomplete</i>
EP (positions and means)	<b>36.68</b>	<b>35.27</b>	<b>34.19</b>	<b>33.31</b>	<i>9/7, overcomplete</i>
EP (positions, means, variances)	<b>37.03</b>	<b>35.81</b>	<b>34.95</b>	<b>34.25</b>	<i>9/7, overcomplete</i>

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 previously denoised lowpass subband data. We have developed two versions of the denoising algorithms that exploit this model in critically sampled and overcomplete transform domains. Benchmarking with known denoising techniques within the Bayesian estimation framework has demonstrated that our approach is competitive with the best existing denoising techniques in both transform domain.

Aiming at 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 case of the EP model we treat data in the flat regions 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 a set of experiments assuming that the edge region spatial statistics were available from the original image (the so-called performance empirical upper bound was estimated). In this case the significant gain in the PSNR was obtained for all tested AWGN variances. As it 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 [42] and watermarking [43] where attacks and watermark are power limited due to perceptual constraints on image fidelity.

## 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", by SNF Professorship grant No PP002-68653/1, 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. The authors also acknowledge the valuable comments of anonymous reviewers.

## References

- [1] M. Lindenbaum, M. Fischer, and A. M. Bruckstein, "On gabor contribution to image enhancement," *Pattern Recognition* **27**, pp. 1–8, 1994.

- [2] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Patt. Anal. Mach. Intell.* (12), pp. 629–639, 1990.
- [3] L. Yaroslavsky, *Digital Picture Processing - An Introduction*, Springer Verlag, 1985.
- [4] L. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, 1996.
- [5] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing* **19** (1-3), pp. 553–572, 2003.
- [6] L. Rudin and S. Osher, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, 1992.
- [7] A. Buades, B. Coll, and J. M. Morel, "On image denoising methods," <http://www.cmap.polytechnique.fr/~peyre/cours/>.
- [8] A. Haddad and Y. Meyer, "Variational methods in image processing," <http://www.inrialpes.fr/is2/people/pgoncalv/WAMA2004/lectures/Haddad.ps>.
- [9] M. Nikolova, "Regularization functions and estimators," in *IEEE Int. Conf. On Image Proc*, pp. 457–460, (Lausanne, Switzerland), September 1996.
- [10] J. Liu and P. Moulin, "Image denoising based on scale-space mixture modeling of wavelet coefficients," in *IEEE Int. Conf. On Image Proc*, (Kobe, Japan), October 1999.
- [11] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with content modeling for image denoising," in *Proc. of 5th IEEE International Conference on Image Processing ICIP98*, (Chicago, USA), October 1998.
- [12] A. Synyavskyy, S. Voloshynovskiy, and I. Prudius, "Wavelet-based map image denoising using provably better class of stochastic i.i.d. image models," *Facta Universitatis* **14**, 2001.
- [13] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis*, (Pittsburg, USA), October 1998.
- [14] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Wavelet shrinkage: Asymptopia? (with discussion)," *Journal of the Royal Statistical Society Series B* **57**, pp. 301–369, 1995.
- [15] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors," in *IEEE Trans. on Inf. Theory*, **45**, pp. 909–919, Apr. 1999.
- [16] S. LoPresto, K. Ramchandran, and M. Orhard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Data Compression Conference 97*, pp. 221–230, (Snowbird, Utah, USA), 1997.

- [17] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters* **6**, pp. 300–303, December 1999.
- [18] S. Vaseghi, *Advanced Signal Processing and Noise Reduction*, John Wiley & Sons, Inc., 1998.
- [19] C. Chrysafis, *Wavelet Image Compression: Rate Distortion Optimizations and Complexity Reductions*. PhD thesis, University of Southern California, Los Angeles, USA, February 2000.
- [20] R. R. Coifman and D. L. Donoho, "Translation invariant denoising," in *Wavelets and Statistics, Springer Lecture Notes in Statistics 103*, pp. 125–150.
- [21] X. Li and M. T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *Proc. IEEE Int. Conf. Image Proc.*, Sept. 2000.
- [22] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Adaptive wiener denoising using a gaussian scale mixture model in the wavelet domain," in *8th IEEE Int. Conf on Image Processing*, (Thessaloniki, Greece), October 2001.
- [23] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. on Information Theory* **38**, pp. 587–607, 1992.
- [24] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis* **10**, pp. 234–253, May 2001.
- [25] J. Starck, E. Candes, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.* **11**, pp. 670–684, June 2002.
- [26] E. Candes and D. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise  $c^2$  singularities," *Comm. Pure Appl. Math.* **57**(2), pp. 219–266, 2004.
- [27] M. Do and M. Vetterli, "Framing pyramids," *IEEE Trans. on Signal Proc.* **51**, pp. 2329–2342, September 2004.
- [28] B. Matei, *Methodes multiresolutions non-linéaires. Applications au traitement d'image*. PhD thesis, Université Pierre et Marie Curie - Paris VI, Paris, France, November 2002.
- [29] B. Matei, "Denoising using nonlinear multiscale transform (english. abridged french version)," *C. R., Math., Acad. Sci. Paris 338*, 2004.
- [30] E. L. Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. on Image Proc.*, p. to appear.
- [31] R. Wilson, "Multiresolution gaussian mixture models: Theory and applications," *Research Report RR404, Department of Computer Science, University of Warwick, U K*, December 1999.
- [32] C. U. S. Software, "<http://lci.die.unifi.it/publications/swmc99-04.ps.gz>,"

- [33] A. Hjørungnes, J. Lervik, and T. Ramstad, "Entropy coding of composite sources modeled by infinite gaussian mixture distributions," in *IEEE Digital Signal Processing Workshop*, pp. 235–238, 20–24 January 1996.
- [34] V. Strela, "Denoising via wiener filtering in wavelet domain," in *3rd Europ. Conf. on Mathematics*, (Barcelona, Spain), July 2000.
- [35] M. Wickerhauser, *Adaptive Wavelet Transform: From Theory to Software*, A.K. Press, 1994.
- [36] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Proc.* **1**(2), pp. 205–220, 1992.
- [37] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," in *SPIE Technical Conference on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, July 1999.
- [38] S. Xiao, I. Kozintsev, and K. Ramchandran, "Stochastic wavelet-based image modeling using factor graphs and its application to denoising," in *Proc. SPIE: Visual Commun. and Image Proc.*, SPIE, 2000.
- [39] G. Fan and X.-G. Xia, "Wavelet-based statistical image processing using hidden Markov tree model," in *Proc. 34th Ann. Conf. Inform. Science and Systems*, 2000.
- [40] C. Weidmann and M. Vetterli, "Rate-distortion analysis of spike processes," in *Data Compression Conference*, (Snowbird, USA), March 1999.
- [41] K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.
- [42] J. Vila, O. Koval, and S. Voloshynovskiy, "Facial image compression using overcomplete transforms," in *Proceedings of SPIE: Electronic Imaging 2004, Image and Video Communications and Processing VI*, (San Jose, CA, USA), January 2004.
- [43] 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/SPIEs annual Symposium, Electronic Imaging 2003: Security and Watermarking of Multimedia Content*, SPIE, 2003.