The Edge Process Model and Its Application to Information-Hiding Capacity Analysis

Sviatoslav Voloshynovskiy, Oleksiy Koval, M. Kivanc Mihcak, and Thierry Pun

***Abstract*—In this paper, the problem of capacity analysis of data-hiding techniques in a game information-theoretic framework is considered. Capacity is determined by the stochastic model of the host image, by the distortion constraints, and by the side information about the watermarking channel state available at the encoder and at the decoder. The importance of the proper modeling of image statistics is emphasized, and for this purpose, a novel stochastic nonstationary image model is proposed that is based on geometrical priors, the so-called edge process model. Being mathematically simple and tractable, the edge process model outperforms the estimation-quantization (EQ) and spike process models in reference applications such as denoising. Finally, this model allows us to obtain a realistic estimate of maximal embedding rates, and in particular, it is shown that the expected capacity limit of real images is significantly lower than previously reported.**

***Index Terms*—Capacity, edge process model, estimation-quantization (EQ) model, information theory, spike process model, stochastic image model, watermarking.**

I. Introduction

An important problem in digital data hiding is the investigation of fundamental capacity limits, i.e., somehow an analog to Shannon's limit in digital communications. The recent work proposed by Moulin advocates a game-theoretic approach for the evaluation of data-hiding capacity [18]. In this approach, the data-hiding capacity is considered to be the solution of a max–min two-player zero-sum game between the data-hider attempting at maximizing reliable information transmission and the attacker aiming at decreasing it. No specific form of encoder/decoder and blockwise memoryless attack channel is assumed, but it is rather supposed that both data-hider and attacker are doing the best to achieve their goals. Therefore, one is looking for the maximum rate of reliable communications, over the best possible data-hiding strategy, and the worst attack that satisfies the specified constraints.

An emerging practical problem is the application of the game-theoretic paradigm to the analysis of the data-hiding capacity of real images. The solution to this problem should formally justify a number of existing practical algorithms and allow to fairly evaluate their performance and potential capabilities. An important aspect of this problem is to develop appropriate models for attack channels, for distortion metrics, and for image statistics. The analysis of these three items has great impact on the solution of the max–min problem. Therefore, it is justified by Moulin and O'Sullivan that a minimum mean-square error (MMSE) estimator of the host signal and a *Gaussian test channel* from rate distortion theory are the worst-case memoryless attacks for a given constrained attack distortion D_2. A squared-error distortion measure $d(x,y) = (x-y)^2$ is selected for the analysis due to its wide usage in communication theory and the fact that it often yields nice closed-form results. A Gaussian model of the host image is selected as a source model, since it provides the *upper bound* on capacity for non-Gaussian sources with bounded variance as well. It is also assumed that the maximum admissible distortion for the data-hider is D_1 while for the attacker it is constrained by D_2 (for further details, we refer the reader to [5], [18], and [20]).

In a data-hiding communication setup in the case when a fixed attacking channel is assumed, it is possible to design a codebook allowing perfect host interference cancellation [6]. This is not the case when the same problem is considered using game theory apparatus. An important assumption here is that the stochastic model of the source image is shared among all involved into the games parties. In the original analysis of Moulin and Mihcak [20], it is proposed to use an *estimation-quantization* (EQ) [11] and a *spike process model* [33] to evaluate the data-hiding capacity of real images due to their superior performance in reference applications. Therefore, the final watermark energy allocation and the attack distortion distribution are performed according to the selected models. Although this approach is intuitively justified, it could potentially lead to the overestimation of the actual capacity since the obtained estimate of the rate of reliable communications is highly sensitive to the source model selection. This means that the higher is the variance of the source, the larger will be the capacity estimate. Therefore, although the selection of the EQ and spike process models have been well justified for the game-theoretic analysis of information hiding, new more powerful source models can appear requiring to reassess the obtained fundamental capacity limits for those models.

The goal of this paper is to introduce a new class of nonstationary stochastic image models based on geometric priors that show superior performance in some reference applications such as denoising over the EQ and spike process models. The proposed model is applied to the game-theoretic setup and new capacity results are obtained. Since the proposed model has

Manuscript received October 26, 2004; revised June 23, 2005. This paper was partially supported by Swiss National Science Foundation (SNF) grants No 21-064837.01 and PP00268653, by the Swiss National Centers of Competence in Research (NCCR) IM2 and by the European Commission through the IST Programmes under contracts IST-2002-507932-ECRYPT and FP6-507609-SIMILAR. The information in this document reflects only the author's views, is provided as is, and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zixiang Xiong.

S. Voloshynovskiy, O. Koval, and T. Pun are with University of Geneva-CUI, Geneva 4, Switzerland.

M. K. Mihcak is with Microsoft Research, Redmond, WA 98052 USA (e-mail: svolos@cui.unige.ch).

Digital Object Identifier 10.1109/TSP.2006.871965

considerably lower local variances, the obtained practical embedding capacity is also smaller than those obtained for the EQ/spike process models. Therefore, we demonstrate that, although the information-theoretic game approach is an excellent framework for providing the absolute limits of information-hiding capacity, the practical limits are highly sensitive to the model selection and to the transform domain used. Essentially, we show that the capacity under the EP model is much lower than the one given by the EQ or spike process models.

Section II introduces the *edge process* (EP) model and provides the comparison of this model with the EQ and spike process models for the reference applications. Section III briefly reviews the main results of Moulin *et al.* game-theoretic approach and presents the actual capacity for the EP model for a number of test images. A new attack based on the EP model is presented in Section IV. Section V summarizes some open issues of the game-theoretic approach and concludes the paper.

Notation: We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, and corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively. The superscript N is used to denote length-N vectors $\mathbf{x} = x^N = \{x[1], x[2], \ldots, x[N]\}$ with ith element $x[i]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. The mathematical expectation of a random variable $X \sim p_X(x)$ is denoted by $E_{p_X}[X]$ or simply by $E[X]$ and $\mathrm{Var}[X]$ denotes the variance of X. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of a set.

II. Edge Process Model

The model selection is a very important but at the same time very ambiguous issue. It involves a lot of subjective experience dealing with the estimation, detection and rate-distortion problems.

The selection of the EQ or spike process models is justified by their excellent performance in some reference applications such as image compression and denoising and their good fit to the parallel Gaussian channel model [11], [15], [16]. In a more general case, the advantages of one model over another are considered based on the satisfaction of a list of requirements which determine its suitability for the practical applications: a) model simplicity (preferably Gaussian-type models due to easy integration and differentiation); b) model ability to lead to a closed-form analytical solution; c) model and result tractability and existence of performance bounds (preferably Gaussian-type models due to the corresponding upper and lower bounds in channel capacity and rate-distortion theory); and d) model robustness in the sense of the model applicability to a wide class of real images.

A. EP Model: Definition and Experimental Validation

1) Stochastic Image Modeling—Main Trends and Assumptions: During the last few years, significant efforts have been devoted to the development of accurate stochastic image models. This has had impact on all image processing applications such as compression [11], [26], restoration [2], denoising [19], [22], [24], data-hiding [29], and image communications through noisy channels [8], [14], [17]. The most advanced stochastic models are applied to a transform-based image representation. In particular, the image is often presented in the discrete cosine transform (DCT) or discrete wavelet transform (DWT) domains. Image modeling in the transform domain has a number of advantages in comparison with the coordinate domain. Such transforms aim at achieving approximate decorrelation and energy compaction, thus resulting in sparse data as well as nice fit to the properties of the human visual system (HVS) [27]. The DWT has a number of advantages over the DCT, which resulted in the necessity to reconsider the DCT based JPEG compression standard and led to the development of the new DWT-based JPEG2000 standard. Good energy compaction properties of the wavelet transforms and the remarkable properties of wavelet domain representations, such as sparsity, locality, and multiresolution, have made wavelet-domain techniques successful and popular. The best recent image restoration, denoising, and image communication systems also exploit the DWT.

In the image compression and estimation literature, various stochastic models have been proposed to characterize the dependencies among wavelet coefficients. These models can be classified into three categories: 1) those exploiting interscale dependencies, 2) those exploiting intrascale dependencies, and 3) those exploiting both dependencies. Since Liu and Moulin [10] have shown that the intraband dependencies are stronger than interband dependencies, we will concentrate only on intraband models.

The most simple and widely used class of intraband stochastic image models is a family of independent and identically distributed (i.i.d.) generalized Gaussian distributions (GGD) [12]. The GGD model captures the global behavior of wavelet coefficients. A particular case of this model is the Laplacian probability density function (pdf), which is obtained when the shape parameter of the GGD is equal to 1. A number of practical image coders and denoisers are designed based on the Laplacian model [19], [34]. However, an even more significant gain can be achieved when the coefficients are considered to be locally Gaussian rather than globally Laplacian. The corresponding procedure of local image coefficients splitting or classification based on their statistical properties is known as a *source splitting* [7]. Therefore, it is important to establish the mathematical relationship between local and global stochastic models. This link can be found based on the *infinite Gaussian mixture model*. According to this model, the global Laplacian pdf $p_X(x)$ is obtained as a weighted mixture of zero-mean conditionally Gaussian pdfs (conditioned on local variance σ_x^2) and exponential prior on σ_x^2 that capture the local image statistics

$$p_X(x) = \int_0^\infty p_{X|\Sigma_x^2}\left(x|\sigma_x^2\right) p_{\Sigma_x^2}\left(\sigma_x^2\right) d\sigma_x^2, \tag{1}$$

where $p_{X|\Sigma_X^2}(x|\sigma_x^2) = (1/\sqrt{2\pi\sigma_x^2})e^{-(x^2/2\sigma_x^2)}$ and $p_{\Sigma_x^2}(\sigma_x^2) = \lambda_1 e^{-\lambda_1\sigma_x^2}$ and λ_1 is the scale parameter of Laplacian distribution and

$$p_X(x) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \lambda_1 e^{-\lambda_1\sigma_x^2} d\sigma_x^2 = \sqrt{\frac{\lambda_1}{2}} e^{-\sqrt{2\lambda_1}|x|}. \tag{2}$$

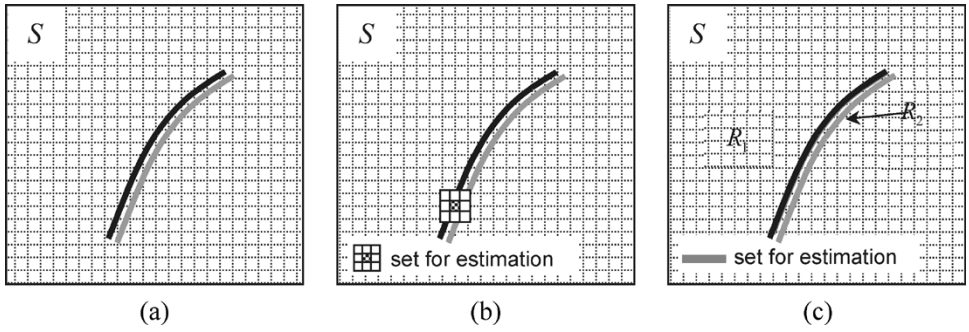


Fig. 1. Explanation of the EQ and EP models: (a) the edge structure in the transform domain (critically sampled or overcomplete transform); (b) ML-estimation of the local image variance for the edge coefficient in the local set using a stationary assumption; and (c) ML-estimation of the local image variance for the edge coefficient using a stationary set of coefficients only along the edge (set R_2).

This simple relationship provides a fundamental link between the global and local statistics of image coefficients. Therefore, the same data can be considered to be locally zero-mean Gaussian with the variance distributed according to the exponential pdf and, at the same time, having Laplacian global statistics. The Gaussian mixture model is also the basis for a *parallel channel decomposition* of stochastic image sources that makes it possible to use the simple relationships that exists for Gaussian statistics.

The number of Gaussian channels is limited in practice by K instead of an infinite number. Therefore, the source $\mathbf{X}$ is split into K classes according to their variances, such that the samples with variance in the intervals $[0;\sigma^2_{x_1}), [\sigma^2_{x_1};\sigma^2_{x_2}), \ldots, [\sigma^2_{x_{K-1}};+\infty)$ belong to the first, second,..., Kth classes, accordingly. From another perspective, the data from these classes can be also considered as different length vectors generated by Gaussian pdf's with different variances $\{\sigma^2_{x_i}\}$ governed by an exponential law. The joint consideration of all samples results in the Laplacian pdf.

It is important to note that one of the state-of-the-art image compression algorithms, the EQ codec [11], is based on a variant of this model. In fact, omitting the practical details of side information communications between the encoder and the decoder, Hjorungnes *et al.* [7] were the first who theoretically demonstrate that the rate gain between Laplacian and Gaussian Mixture models can be as much as 0.312 bits/sample for a high-rate regime. This is achieved by the proper design of entropy coders for each subclass of coefficients. In addition, the analog of the EQ model was used by Mihcak *et al.* [15], [16] in the context of image denoising where state-of-the-art results were demonstrated.

2) EQ and Spike Process Models—Open Issues: The goal of this section is to analyze the problem of model parameter estimation and to point out the main drawbacks of the EQ/spike process models that neglect the nonstationary mean behavior of the wavelet coefficients. These drawbacks are related to the usage of a *maximum-likelihood* (ML) estimation of model parameters, i.e., the local image variances. To illustrate these drawbacks, let us consider Fig. 1, where some edge structure is presented in the transform domain [Fig. 1(a)]. The edge structure consists of two distinctive parallel propagating sets of coefficients with different mean values of opposite sign polarity.

In the scope of the EQ/spike process models, the ML estimation is used to estimate the local variance of the wavelet coefficients in some local neighborhood $\Omega(i)$. We assume an $L \times L$ square window $\Omega(i)$ centered at location i. The estimation of the local variance according to the ML estimate is

$$\hat{\sigma}^2_x[i] = \frac{1}{|\Omega|} \sum_{\omega \in \Omega(i)} |x[\omega]|^2 \tag{3}$$

where $|\Omega|$ is the cardinality of $\Omega(i)$ for all i. Although this ML variance estimator is widely used in the image processing community, its usage is justified only for approximately locally stationary data, which is not always the case for real images. The condition of stationarity is especially violated in the vicinity of edges and textures. As a result, a lot of outliers from the wavelet coefficients with different local means are in the square window in the vicinity of an edge and one obtains extremely high estimates for the local variance. The detailed analysis of this phenomenon is presented in our previous papers [30], [32] on the example of one-dimensional (1-D) edge structure. The ML estimate corresponding to (3) is shown in Fig. 1(b) for the edge coefficient in the transform domain. As in the 1-D case [30], [32], one can clearly observe the aforementioned issue of the ML estimate due to the nonstationarity of the wavelet coefficients. The local window contains both coefficients belonging to the edge and coefficients belonging to the flat region. Contrarily, Fig. 1(c) demonstrates a situation where the ML-estimate of the image variance for the edge coefficient is taken using only those coefficients along the stationary edge, corresponding to the stationarity condition for the ML-estimate. This practically corresponds to the wavelet subband separation into two large regions R_1 and R_2. The region R_1 represents the flat regions in images and can be assumed to have zero mean in the wavelet domain while the region R_2 corresponds to some edge structure with distinctive mean values along the direction of edge propagation. Therefore, it is obvious that neglecting the nonzero mean or applying the ML-estimate to nonstationary data, one obtains a highly overestimated variance, especially in the regions of the textures and edges where the condition of stationarity is violated.

3) EP Model Definition for Real Images: In this section, we introduce a new stochastic image model that allows to separate various wavelet subband coefficients into distinct regions and to treat them separately.

We consider an image $\mathbf{x}$ with a support S as a realization of a random field $\mathbf{X}$ with distinct stochastic behavior in different re-

gions. Let $R_1, R_2, \ldots, R_M$ be the set S partitioning, where for all i R_i are disjoint connected sets, i.e., $R_i \cap R_j = \emptyset$, $i \neq j$ and $\cup_i R_i = S$. Let $\mathbf{x}_l$ denote the subset of image pixels supported by the regions R_l. In our model, we assume that each region R_l is fully covered by the model $\theta_l \in \{\Phi_1, \Phi_2, \ldots, \Phi_L\}$ and that no two neighboring R_l contain the same model. In particular, we assume that the pixels in the image subregion $\mathbf{x}_l$ are distributed with a joint pdf $p_{\mathbf{X}}(\mathbf{x}_l|\theta_l)$.

Our goal is to introduce the EP model 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.*, a generalized Gaussian [11]) 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 $S_l, l = 1, \ldots, 3M$, where M 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 the nondecimated 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 approximately local stationary assumption of the data under the EQ model, we assume that only one region R_l is given within the subband S_l and that all coefficients in this subband belong to the same region R_l, as follows:

$$R_l = \left\{\mathbf{x}_l : X_l[i] \sim \mathcal{N}\left(0, \sigma^2_{x_l}[i]\right)\right\} \tag{4}$$

i.e., all coefficients are considered to be approximately locally i.i.d. with Gaussian distribution, but with different local variances $\sigma^2_{x_l}[i]$. Equivalently, this means that only one stochastic model out of $\{\Phi_j\}$ is applied to the whole support S_l and $p_{\mathbf{X}_l}(\mathbf{x}_l|\theta_l)$ 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 EP 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. In the following we will refer to the transition simply as the edge.

Due to the above assumption, one can distinguish two main approaches to efficiently separate the regions or to simply detect the edge. The first approach exploits the *spatial separation* of the regions which requires to know an inter region boundary. Once the boundary is specified, the ML-estimate is applied to both regions independently. The knowledge of the separation boundary requires a relatively a small amount of side information to be known in order for the ML estimate to avoid inaccuracies in the variance estimation. This approach is obviously very simple in the one-dimensional case. However, some technical difficulties might arise in the two-dimensional (2-D) case of real images when the exact separation boundary for all regions should be specified.

The second approach exploits the *amplitude-based separation* of the regions assuming that the regions statistics can be reliably distinguished and that there is a considerable difference between their mean values. One can consider the region separation problem as a corresponding problem of pulse amplitude modulation (PAM) digital communications [25] represented by the selected mean values. Moreover, the amplitude classification problem is equivalent to a multiple hypothesis testing problem or to a change point detection problem [1], [4] and appropriate techniques can be used for this purpose as well. Finally, the region-based segmentation based on K-means or projection onto convex sets (POCS) methods that are well-known in computer vision can be of benefit [3].

However, in our approach, we will follow the well-known duality of region segmentation and compression problems [9]. We consider the mean values as the reconstruction levels of a scalar uniform threshold quantizer (UTQ) designed for the given subband S_l that is characterized by the global GGD $X_l \sim p_{X_l}(x_l)$, and we assume $p_{X_l}(x_l) = \mathcal{GGD}(\mu_l, \gamma_l, \lambda_l)$ to be the GGD pdf with mean $\mu_l = 0$, shape parameter γ_l and scale λ_l, as follows:

$$\bar{x}^l_j = \frac{\int_{d_j-1}^{d_j} x_l p_{X_l}(x_l) dx_l}{\int_{d_j-1}^{d_j} p_{X_l}(x_l) dx_l} \tag{5}$$

where $d_j - 1 \leq \bar{x}^l_j \leq d_j$ and $\{d_j\}$ are the UTQ decision levels. It is assumed that the UTQ has uniformly spaced decision levels and that reconstruction levels are selected to minimize the MSE. The dead zone of UTQ $[-T, T]$ is chosen to be 2Δ, where Δ 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 amplitude-based partition approach, the EP model is defined as

$$R_1 = \left\{\mathbf{x} : X[i] \sim \mathcal{N}\left(0, \sigma^2_x[i]\right)\right\} \tag{6}$$

$$R_2 = \left\{\mathbf{x} : X_j[i] \sim \mathcal{N}\left(\bar{x}_j[i], \sigma^2_{x_j}[i]\right)\right\} \tag{7}$$

where the subscript j is used to indicate the data behavior along the jth local edge, $R_1 \cup R_2 = S$ and S represents a particular subband. Equation (7) assumes a proper separation of the regions with distinctive statistics. If the width of the bin Δ is chosen to be relatively small compared to the flatness of the pdf, one can use an uniform approximation of the region R_2 statistics

$$R_2 = \left\{\mathbf{x} : X_j[i] \sim \mathcal{U}\left(\bar{x}_j[i], -\frac{\Delta}{2}, +\frac{\Delta}{2}\right)\right\}. \tag{8}$$

However, we are not looking here for the best possible quantized approximation of the data but rather to obtain the proper region separation. Therefore, the high-rate mode of quantization is out of interest in our formulation.

The region R_1 represents all flat regions within a subband assumed to be zero-mean Gaussian random variables with the local variance $\sigma^2_x[i]$. The region R_2 corresponds to the texture and edge regions. Each distinctive geometrical structure corresponding to the edge or texture transition within R_2 is decomposed into a set of local mean constellations. Moreover, a particular mean value $\bar{x}_j[i]$, $j = 1, \ldots, J$, where J is the number of mean levels, 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

(a) (b)

Fig. 2. Explanation of the EP model: (a) Original Lena image and (b) five-level orthogonal wavelet transform $(Db8)$.

all coefficients. We can also assume that the variation of the coefficients with respect to the mean values (this is especially true for the overcomplete transform) is very small. Thus, the EP model assumes that the image consists of random Gaussian processes with zero-mean and some small local variance for the flat regions with almost "deterministic" edge occlusions. The parameters and orientations of edge occlusions depend on the mutual orientation of the edge and of the subband. Moreover, transitions along the edge usually have longer stationary length than the transitions within the texture (which 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 edge shape approximation accuracy. This phenomenon is contrarily to the one observed when the data are modeled by the EQ or spike process models where 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 particular geometrical spatial structure is assumed among the spikes in the spike process model. This is not the case when the EP model where the "spikes" belonging to the same edge are treated jointly along the direction of edge propagation.

4) EP Model Experimental Validation on Real Images: To demonstrate the main features of the EP model on real images we performed a number of experiments using a set of standard test images. We show here results obtained using the image Lena of size 512×512 [Fig. 2(a)]. Fig. 2(b) illustrates the results of the selected image decomposition into a five-level wavelet transform pyramid with Daubechies orthogonal wavelet Db8 filter. A similar four-level decomposition was performed in the case of discrete overcomplete transform (DOT) based on the 9/7 filter pair [13].

The absolute values of the first horizontal subbands are shown in Fig. 3(a) and (c) for the DWT and DOT, respectively. The DWT first level subband has a two times smaller linear size in comparison with the DOT subband. The estimation of the mean along the edge according to the EP model, followed by its subtraction, results in the more uniform random field shown in Fig. 3(b) and (d). It is important to note that the amplitude of the coefficients is considerably reduced. The "edge" is visually less detectable and the subbands are more decorrelated.

As an example of edge description according to the above transforms, we have selected a fragment of the first horizontal subband of the DWT on the shoulder of Lena image [Fig. 2(b)] that corresponds to a well distinguished propagating edge [Fig. 4(a)] and the corresponding fragment for the DOT [Fig. 4(c)]. One can also observe some variations along the edge due to the critically sampled character of the DWT. The variation along the edge is smoother for the DOT due to the absence of downsampling [Fig. 4(c)]. Small variations of the transformed coefficients are observed on both sides of the edge in the case of the DWT and DOT. The corresponding fragments after the EP mean estimation and subtraction are shown in Fig. 4(b) and (d) for the DWT and DOT, respectively.

Since both EQ and spike process models refer to the local variance as the model parameter, we visualize the subband local variances in Fig. 5(a) for the EQ model and in Fig. 5(b) for the EP model. A seven-mean constellation for each sign was used in the EP model for the R_2 region. Obviously, this simple constellation scheme causes some approximation error that gives rise to an increase of the local variance in the vicinity of the edge. More powerful analytical edge shape approximation techniques can be used for this purpose and the work is underway [28]. The ML variance estimate was used in both cases (3). One can also observe a significant decrease of the local image variance with respect to the peak variance values. The decrease of the variance has a crucial impact on the performance of denoising and compression algorithms as well as on the capacity estimation problem.

To demonstrate the model ability to carry out and to capture only the significant image components with a certain level of sparsity, we performed a set of experiments. First, the image was decomposed using the DWT$(Db8)$ into a five-level pyramid and into a four-level DOT (9/7) pyramid. Second, the EP model was applied and the edges were replaced by their mean estimates according to the EP model. In both cases, all the information about the flat part of the image was completely discarded. The quality of the obtained reconstructed images in terms of PSNR were 36.95 dB for the DWT domain case and 41.72 dB for the DOT case.

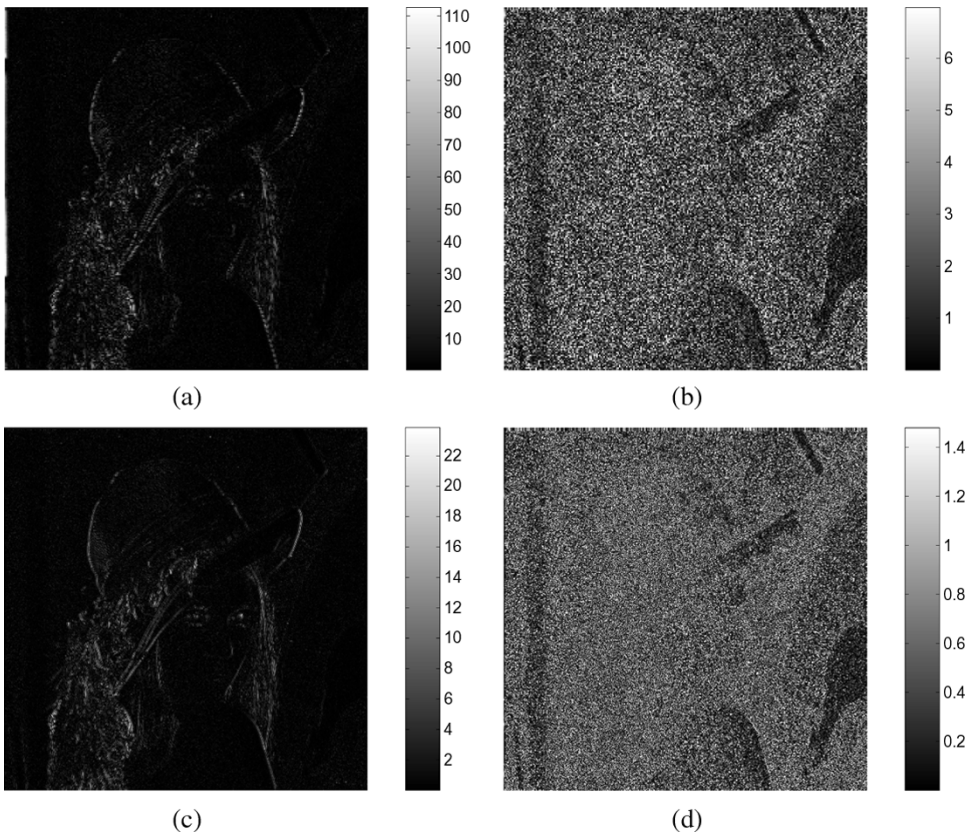


Fig. 3. Explanation of the EP model based on Lena image (absolute values of the results are shown): (a) First horizontal subband of the Lena image in the DWT ($Db8$) and (c) DOT (9/7) decompositions; (b) and (d) the same subbands after EP model mean subtraction along the edges for the DWT and DOT, respectively.

We now briefly summarize the main features of the EP image model. First, the EP model offers an additional data "decorrelation" even for fixed transform basis functions, which could be a very useful feature for many applications such as denoising, compression and watermarking. It should also be noted that the complexity of this "decorrelation" transform, besides model overhead, still remains almost the same as the complexity of the corresponding wavelet or overcomplete transforms. Second, the resulting distribution of the subband coefficients is close to Gaussian assuming a proper selection of decision boundaries. This further considerably simplifies the analysis and guaranties the existence of closed-form solutions for many applications, contrarily to non-Gaussian image models. Since the data is Gaussian and decorrelated, this also brings us an additional benefit the independent character of the coefficients. This allows the modeling of joint subband pdf as a product of independent pdf's of each coefficient. This also supports the idea to use the parallel Gaussian channels for the analysis of the data-hiding capacity. Third, the subtraction of the local mean along the edges makes data more "stationary" and considerably reduces the value of variance estimated based on the ML-estimator thereby reducing the uncertainty. This has an important impact on the performance of denoising and compression algorithms. Further, this provides a completely different justification for the data-hiding algorithm performance as opposed to the EQ model (smaller host interference for spread spectrum-based information-hiding techniques and different capacity estimation limits).

B. EP Model: Justification

To our knowledge the EQ model produces state-of-the-art results in image denoising and compression. Therefore, to have a fair justification we will compare the proposed EP model with the EQ model in three sets of tests: image generation from model parameters, operational entropies and reference applications.

1) Image Regeneration: The goal of this section is to demonstrate the power of the proposed model with respect to the generation of images from their statistical descriptions. The experiment performed here are similar to the tests performed in [23] while investigating the performance of the EQ model.

The image Lena is used as a test target example. The image is decomposed into the wavelet pyramid using 9/7 biorthogonal pair. The subbands of the decomposed image are used to estimate the parameters of the four models under test. We keep the same low-pass subband for all simulations.

In the first case, the Gaussian version of the EQ model is investigated for the sake of fair comparison. The local variances are estimated in all high-frequency subbands of the DWT from the target image. At the same time, a pseudo-random excitation field is generated in the coordinate domain from a unit-variance Gaussian distribution and then transformed to the DWT domain. The DWT coefficients of the pseudorandom excitation field are multiplied by the corresponding estimated standard deviations (except for the low-pass subband). The resulting image is transformed back to the coordinate domain using inverse DWT [Fig. 6(a)].

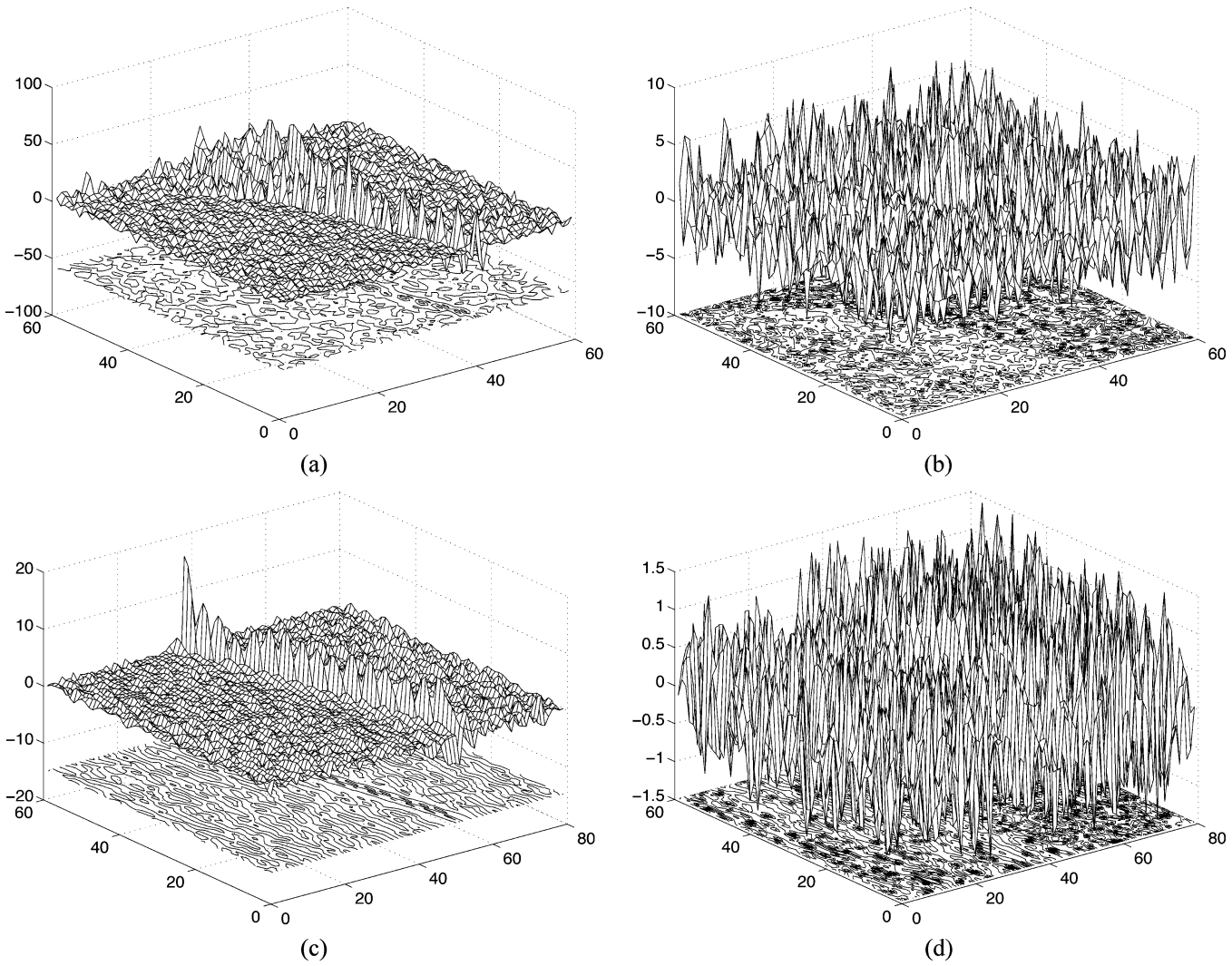


Fig. 4. Example of real edge on Lena's shoulder: (a) and (c) original Lena's shoulder in the first horizontal subbands of the DWT and DOT domains according to Fig. 3(a) and (c); (b) and (d) the same fragments after EP mean subtraction.

The second model discussed in [23] represents a hybrid semi-deterministic/semi-stochastic version of the EQ. In this model, the sign of the DWT coefficients extracted from the DWT coefficients of the target image (deterministic part of the model) is used in addition to the local variances (stochastic part of the model). Since some extra side information (or constraints on the image generation that relaxes a prior ambiguity) is used, one can expect that the resulting image will be closer to the target image according to its statistics. The resulting image generated according to this procedure is shown in Fig. 6(b) and supports this conclusion.

Two similar sets of experiments have been performed according to the EP model. The only difference with the previous tests consists in the usage of additional priors given in the form of the EP model parameters, i.e., the means of the DWT coefficients propagating along the edges. We assume that these means also belong to the deterministic part of the image model. Therefore, in addition to the previous local variance priors, the DWT coefficients of the random excitation field are replaced by the EP means in corresponding subbands keeping the same positions and orientations. Therefore, the means of the EP model will completely replace the DWT random excitation field coefficients disregarding their actual initial values. The resulting image is shown in Fig. 6(c). In addition, the usage of the target image subbands sign is demonstrated in Fig. 6(d). An obvious enhancement of both objective and perceptual image quality is observed in the case of the EP model.

2) Operational Entropies: The main objective of this subsection is to demonstrate the "ompressibility" of the EP model. In other words, we would like to show the efficiency of the proposed model in the image compression application. Since the entropy represents an average length of the code for lossless data representation, we have selected this measure for different model comparison.

The setup of the demonstration is the following. Having a data base of 30 512×512 grayscale test images, we estimated the entropy of the high frequency wavelet transform coefficients after the application of a four level transform. The entropies are calculated based on three different assumptions about the stochastic properties of wavelet coefficients assuming that the model parameters are perfectly available at both encoder and decoder [31]. In the first case, we assume that the statistics in each

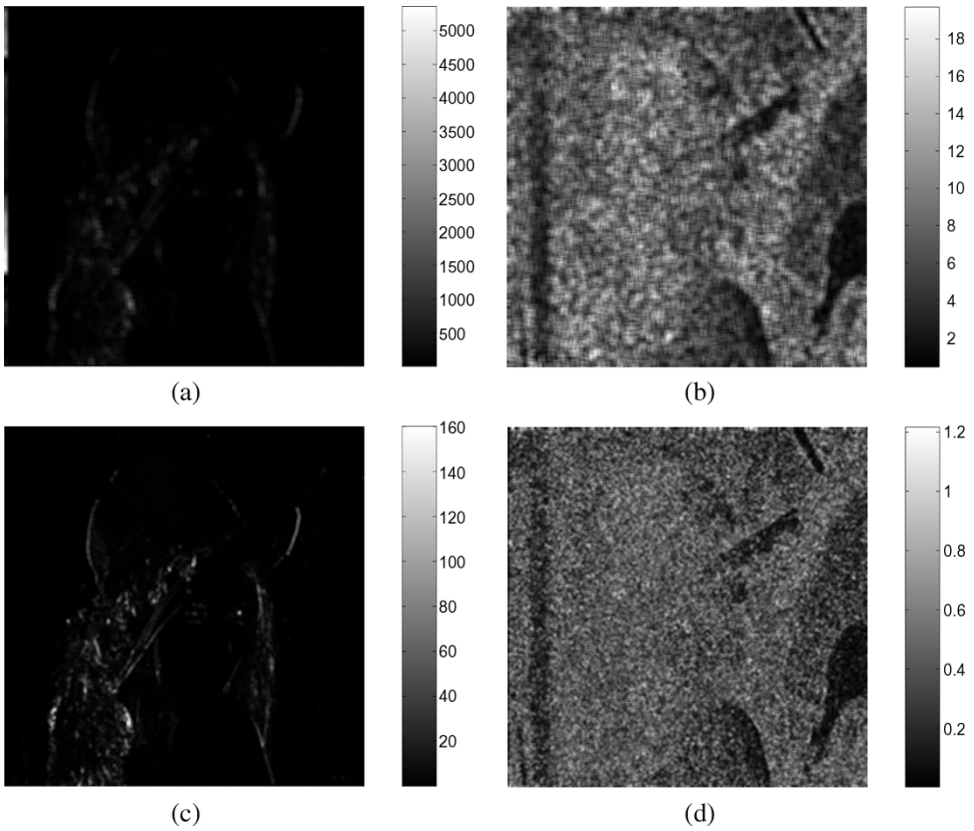


Fig. 5. Variance estimation for the subbands shown in Fig. 3: (a) and (c) ML variance estimation in a 5 × 5 window used in the EQ/spike process model for the DWT and DOT, respectively; (b) and (d) ML variance estimation for the EP model computed in a 5 × 5 window and the same subbands.

wavelet subband can be modeled using a global i.i.d. Laplacian distribution. In this case, one can obtain

$$H_L = \log_2\left(\frac{2e}{\nu}\right) \tag{9}$$

where ν is the parameter of the Laplacian pdf.

Taking into account the cardinality of each subband, we finally calculate the average number of bits per transformation sample that should be used to represent the data in the high-frequency subbands without losses.

In the second case, we apply the Laplacian source splitting property to connect the previous global model with the EQ model. Having an infinite Gaussian mixture representation of the Laplacian pdf (1), the entropy of the EQ, calculated in each subband, is determined by

$$H_{\mathrm{EQ}} = \frac{1}{|S_l|}\sum_{i=1}^{|S_l|}\frac{1}{2}\log_2\left(2\pi e\sigma_x^2[i]\right) \tag{10}$$

where $|S_l|$ is the cardinality of the subband $l, l \in \{1, 2, \ldots, 12\}$ and $\sigma_x^2[i]$ is a variance of the ith zero-mean component of the infinite Gaussian mixture estimated using the ML strategy in a [9 × 9] noncausal window.

Finally, to estimate the entropy of the coefficients in the high frequency wavelet subbands, we assume a locally Gaussian behavior of the data in the subbands and the availability of the mean values along the edge propagation direction. In this case we apply (10), where the variance estimation is performed taking advantage from the availability of the EP information.

The results of those simulations are presented in Fig. 7. They demonstrate that for the selected test images the average amount of bits needed for the lossless representation of the high-frequency subband coefficients is 5.03, 4.37, and 3.54 for the Laplacian, EQ, and EP model, respectively, that is clearly lower for the EP model.

3) EP Model Performance in Reference Applications: An important issue of the model selection is its performance in some reference applications. Therefore, the validity of the EP model was additionally investigated in two reference applications, namely in denoising [32] and in compression [28]. The reason to choose these two reference applications is due to the existence of benchmarking results for the EQ model.

In the denoising, the observed noisy image $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{x} + \mathbf{z} \tag{11}$$

where $\mathbf{z}$ is i.i.d. AWGN $Z[i] \sim \mathcal{N}(0, \sigma_z^2)$. We use the MAP estimator to get the estimate of $\mathbf{x}$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} -\left[\ln p_{Y|X}(y|x) + \ln p_X(x)\right] \tag{12}$$

where $\ln p_{Y|X}(y|x) \sim -(1/\sigma_z^2)\|\mathbf{y}-\mathbf{x}\|^2$ corresponds to the log-likelihood function for the above case of AWGN and $p_X(x)$ represents the priori distribution that corresponds to the EQ and EP models. Since the EQ model is used in the denoising method of Mihcak *et al.* [15], we have chosen this method for the sake of comparison.

Both EQ and EP models belong to the Gaussian family of models with the only difference being in the estimation of the

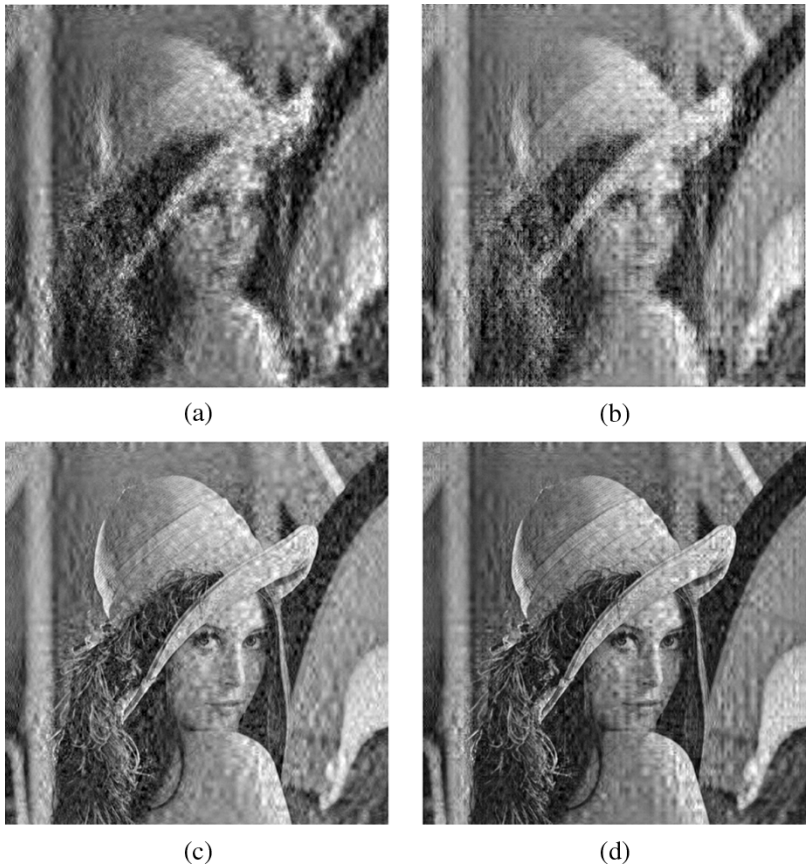


Fig. 6. Example of real image generation according to the different stochastic image models: (a) Image generated from the EQ model variance (PSNR = 19.18 dB); (b) image generated from the EQ model variance and sign of the original subband coefficients (PSNR = 20.56 dB); (c) image generated from the EP model (PSNR = 24.66 dB); and (d) image generated from the EP model and sign of the original subband coefficients (PSNR = 25.49 dB).

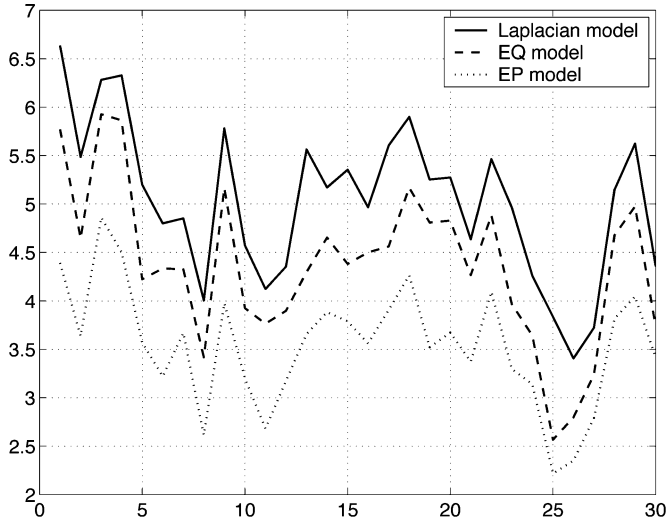


Fig. 7. Operational entropies (in bits per pixel) under (a) global i.i.d. Laplacian model, (b) the EQ model, and (c) the EP model.

model parameters. In this case, the resulting estimate will be in the form of a Wiener filter and the corresponding variance of the estimator can be found as

$$\sigma^2_{\mathrm{MAP}} = \frac{1}{N}\mathrm{tr}\left[E\left[(\mathbf{X}-\hat{\mathbf{X}})^T(\mathbf{X}-\hat{\mathbf{X}})\right]\right] = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_x^2[i]\sigma_z^2}{\sigma_x^2[i]+\sigma_z^2} \tag{13}$$

where $\mathrm{tr}[\cdot]$ states for the matrix trace operation and $\sigma_x^2[i]$ is the local image variance of the EQ or EP models. Obviously, the lower the image variance, the lower the variance of the estimator will be. Thus, taking into account the considerable decrease of local variances in the case of the EP model in comparison to the EQ (Fig. 5), the advantages of the EP models are obvious. More results demonstrating the EP model performance in this reference application can be found in [32].

The lossy compression problem can be formulated in a way similar to denoising as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{C}_X}[D+\lambda R] \tag{14}$$

where $D = (1/N)\mathrm{tr}[E[(\mathbf{X}-\hat{\mathbf{X}})^T(\mathbf{X}-\hat{\mathbf{X}})]]$ is the distortion between the original data and its reconstructed version that corresponds to the above negative log-likelihood function. R is the rate allocated to represent the original image in the codebook $\mathcal{C}_X$ with $|\mathcal{C}_X| = 2^{NR}$, and $R = H(\hat{X}^N) = h(X^N) - \log_2\Delta$. $h(X) = -\int p_X(x)\log_2 p_X(x)$ represents the differential entropy. λ is the Langrange multiplier.

The critical issue is the selection of an accurate stochastic image model for the original image $p_X(x)$. Since, both the EQ and the EP models are Gaussian and treat coefficients independently, $h(X^N) = \sum_{i=1}^{N} h(X_i)$ with the differential entropy $h(X_i) = (1/2)\log_2(2\pi e\sigma_{xi}^2)$. Obviously, the smaller the variance of the individual coefficients, the greater the gain in performance that can be achieved as was the case with the denoising

application. The results presented in Fig. 7 clearly support this conclusion.

Therefore, taking into account all previous results concerning image variances of the EQ and EP models, one can conclude that the EP model can theoretically attain superior performance. The extensive simulation results performed in [28] and [32] experimentally support this conclusion.

III. EP Model Application to Digital Data-hiding

A. Data-Hiding Game-Theoretic Approach

In this section, we review the main results of Moulin and Mihcak [20] for the sake of convenience.

We assume that a message m, uniformly distributed over the message set $\mathcal{M}$ with cardinality $|\mathcal{M}|$, is encoded based on a secret key into some watermark $\mathbf{w} = w^N = \{w[1], w[2], \ldots, w[N]\}$ and embedded into a host data (image) $\mathbf{x} = x^N = \{x[1], x[2], \ldots, x[N]\}$. We denote by $\mathbf{x}$ a 2-D sequence representing the luminance of the original image. The ith element of $\mathbf{x}$ is denoted by $x[i]$ where $i = M_1 n_1 + n_2$, $1 \leq i \leq N$ and $\mathbf{x} \in \mathbb{R}^N$ and $N = M_1 \times M_2$ is the size of the host image. The embedding rule can be expressed as a mapping, as follows:

$$w[i] = f_i(m, x^i) \tag{15}$$

$$y'[i] = x[i] + w[i] \tag{16}$$

where m is a particular realization of the random message M, y'_i, $1 \leq i \leq N$ is the stego data, and x^i represents $x^N = \{x[1], x[2], \ldots, x[N]\}$ (the so-called noncausal side information). The admissible distortion for watermark embedding is D_1

$$E\left[d_1^N(\mathbf{X}, \mathbf{Y'})\right] \leq D_1 \tag{17}$$

where $d_1^N(\mathbf{a}, \mathbf{b}) = (1/N)\sum_{i=1}^N d_1(a[i], b[i])$ denotes N-vector distortion between the vectors $\mathbf{a}$ and $\mathbf{b}$, and $d_1(a[i], b[i])$ denotes the element-wise distortion between the ith elements $a[i]$ and $b[i]$.

The attacker, aiming at imparing reliable communications, modifies the stego data $\mathbf{y'}$ and produces the attacked data $\mathbf{y}$. The admissible attacker distortion according to Moulin definition [18] is D_2 that is defined in the same way as (17) between the vectors $\mathbf{x}$ and $\mathbf{y}$, as follows:

$$E\left[d_2^N(\mathbf{X}, \mathbf{Y})\right] \leq D_2. \tag{18}$$

The decoder produces the estimate of $\hat{m}$ using

$$\hat{m} = g(y^N) \tag{19}$$

where $g(\cdot)$ denotes the decoding rule and $\mathbf{y} = y^N = \{y[1], y[2], \ldots, y[N]\}$ is the distorted stego data. The decoding error occurs when $\hat{m} \neq m$. Finally, Moulin and O'Sullivan considered the above Gel'fand–Pinsker problem in the following game setup [21]:

$$C = \max_{f(\cdot)} \min_{p(y'|y)} \left[I(U;Y) - I(U;X)\right] \tag{20}$$

with the maximization over all embedding strategies $f(\cdot)$ and minimization over all attacking strategies $p(y'|y)$ subject to the embedding and attacking distortion constraints (17) and (18), respectively.

B. Capacity for the Parallel Gaussian Channels and Spike Process Model

Moulin and Mihcak [20] have furthermore used the EQ-model and the so-called *spike process* model introduced by Weidmann and Vetterli [33]. Under the spike process model there are two types of channels: those with large variance, i.e., strong channels $\sigma_x^2 \gg D_1, D_2$, and those with low variance, i.e., weak channels $\sigma_x^2 \ll D_1, D_2$. The image components are assumed to be very sparse and independent and decomposed into K channels using some multirate transform. Assuming that $\sigma_x^2 \gg D_1, D_2$, for $0 \leq k \leq K^* - 1$ and $\sigma_x^2 \ll D_1, D_2$ for $K^* - 1 < k \leq K - 1$, the capacity for the spike process model is then determined as [20]

$$C = \frac{1}{2} r^* \log_2 \left(1 + \frac{D_1}{D_2 - D_1}\right) \tag{21}$$

where $r^* = \sum_{i=0}^{K^*-1} r_i \in \{0, 1\}$ is the fraction of strong signal components.

The MMSE filter cannot considerably help killing the watermark in the strong channels, because it does not produce an accurate estimate of the host signal. The variance of the MMSE/MAP estimators in the Gaussian case for both the host image and the watermark is determined as $\sigma^2_{\mathrm{MMSE}}[k] = \sigma_x^2[k] d_1[k] / \sigma_x^2[k] + d_1[k]$. According to the spike process model assumption for the strong channels $\sigma_x^2[k] \gg d_1[k]$. Therefore, $\sigma^2_{\mathrm{MMSE}}[k] \rightarrow 1$ and the MMSE estimator term $(\sigma_x^2[k]/\sigma_x^2[k] + d_1[k]) \rightarrow 1$, resulting in $\hat{x}[k] \rightarrow y'[k]$ as a one-to-one mapping, i.e., no reliable estimate is produced in this case. The only attack that is introduced in this case is a *test channel* from the rate-distortion theory.

C. Data-Hiding Capacity for the EP Model

It should be pointed out that Moulin and Mihcak have used both the EQ and spike process models in the original paper. Our objective is to validate the data-hiding capacity for the EP model since it produces more realistic results in the cited reference applications.

It is also important to underline that Moulin and Mihcak assume that the model parameters are perfectly known to both involved parties as side information. Therefore, for the sake of fair comparison, we also assume that the parameters of the EP model are available to the data-hider and to the attacker.

Assuming the parallel Gaussian channel energy allocation for the spike (EQ) model, we estimate the data-hiding capacity in the case of the EP model. We use the technique proposed by Moulin and Mihcak [20] for this purpose. The capacity is estimated for the attacker distortion $D_2 = 2D_1$ and for $D_2 = 5D_1$ for several synthetic images shown in Fig. 8 (Table I) and test grayscale images Lena, Barbara, and Baboon of size 512×512 (Table II). The results for the spike process model are not surprising. The images with large amount of transitions like

Fig. 8. Synthetic test images: (a) Step edge; (b) diagonal step edge; (c) cross; and (d) circle.

TABLE I
COMPARISON OF TOTAL DATA-HIDING CAPACITIES (IN BITS) FOR THE SYNTHETIC TEST IMAGES OF SIZE OF 512 × 512, FOR JUST NOTICEABLE DISTORTION D_1 FOR THE SPIKE AND EP MODELS

Image	D_1	$D2 = 2D_1$		$D2 = 5D_1$	
		(spike)	(EP)	(spike)	(EP)
Step edge	25	3232	344	1030	74
Diagonal step edge	25	9739	570	2976	114
Cross	25	4742	511	1427	96
Circle	25	3509	230	752	43

TABLE II
COMPARISON OF TOTAL DATA-HIDING CAPACITIES (IN BITS) FOR REAL TEST IMAGES OF SIZE OF 512× 512, FOR JUST NOTICEABLE DISTORTION D_1 FOR THE SPIKE AND EP MODELS

Image	D_1	$D2 = 2D_1$		$D2 = 5D_1$	
		(spike)	(EP)	(spike)	(EP)
Lena	10	31855	7857	6951	1026
Barbara	10	54632	8480	13045	1237
Baboon	25	80952	2568	17562	436

Baboon and Barbara are characterized by high local variances $\{\sigma_x^2[k]\}$ and a large embedding distortion D_1. Thus, the capacity estimate is higher than for the rather smooth image Lena. The EP model produces much lower estimates of data-hiding capacity due to the more accurate assumptions about image statistics in the transition regions. This surprising phenomenon is justified by the capabilities of the EP model to correctly account for the locally smooth structures on the Baboon and Barbara textured regions for local variance computation.

Since the variance of the EP model is significantly lower in these regions, which are supposed to be the main carriers of the watermark according to the max-min approach and EQ/spike process model, the difference in the data-hiding capacity between the spike and the EP model is considerable. One can argue that the capacity for the EQ/spike process model is higher and thus it should be considered as an upper bound. The difference between the EQ model capacities in the wavelet and DCT domains was explained by the fact that wavelets produce better independence between the channels and better fit to the Gaussian distribution [20]. However, comparing the EP and EQ models, as it was shown is Section V, one can clearly observe that these conditions are even better met with the EP model than with the EQ model. Therefore, the existence of larger capacities for the EQ model does not mean that these results should be considered as practically reachable data-hiding upper bounds.

IV. ATTACK BASED ON THE EP MODEL

The statistics of real images under the EP model can also be used to develop a more involved attacking strategy. The main idea of this attack is based on the nice approximating feature of the EP model.

We have shown that, by preserving only the information about means along the edges, one can obtain a high quality image. Therefore, instead of "killing" the rate of spike coefficients with large variance according to the Gaussian test channel and the assumption about zero mean, we propose to set to zero all flat regions and to replace all edges by the corresponding mean estimates, thus completely killing the watermark in these regions. This attack is an asymptotic case of the Gaussian test channel for zero rate. In this case, the attack distortion will be proportional to the variance of the edge process, that is relatively small in the case of the EP model contrary to the zero-mean EQ model.

If we assume some edge wavelet coefficient to be a constant value θ along the EP defined edge structure, and that a max-min energy allocation for the i.i.d. watermark $\mathbf{W}$ with energy σ_w^2 is performed, we obtain the model

$$y[i] = x[i] + w[i] = \theta + w[i]. \tag{22}$$

Assuming the watermark to be i.i.d. Gaussian $W[i] \sim \mathcal{N}(0, \sigma_w^2)$, we can apply the ML-estimate for θ

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{|\Omega'|} \sum_{j \in \Omega'} y[j] \tag{23}$$

where Ω' is a support of the edge and $|\Omega'|$ is the length of the edge. The ML mean estimate is the unbiased estimate $E[\hat{\theta}_{\mathrm{ML}}] = \theta$ with the variance of the estimate $\mathrm{Var}[\hat{\theta}_{\mathrm{ML}}] = E[|\hat{\theta}_{\mathrm{ML}} - \theta|^2] = (1/|\Omega'|)\sigma_w^2$. Therefore, the longer the edge $|\Omega'|$, the more accurate mean estimate one can receive. In the case of an i.i.d. Gaussian assumption about region R_2 (see (16)) $X[i] \sim \mathcal{N}(\theta, \sigma_x^2)$, the variance of the mean estimator of θ will be $\mathrm{Var}[\hat{\theta}_{\mathrm{ML}}] = \sigma_x^2 + \sigma_w^2/|\Omega'|$. In the case of a simple replacement of the i.i.d. edge signal $X[i] \sim \mathcal{N}(\theta, \sigma_x^2)$ by the sample mean estimated from the watermarked image, the resulting variance of the replacement error will consist of $(1/|\Omega'|)\mathrm{tr}[E[(|\hat{\theta}_{\mathrm{ML}} - \mathbf{X}|)^T(|\hat{\theta}_{\mathrm{ML}} - \mathbf{X}|)] = (\sigma_x^2) + (\sigma_w^2 + \sigma_x^2|\Omega'|)$ that represents the above mentioned asymptotic case of zero rate. Although this attack completely kills the watermark in the flat regions and replaces the edge structures by the EP model mean, the introduced distortion will be smaller than those for the EQ model and Gaussian test channel. This came from using the redundancy along the stationary edge to get a more accurate estimate of the mean and lower local variance contrarily to the EQ/spike process models where all coefficients are treated separately.

V. Conclusion

We have considered the problem of data-hiding capacity for real images by applying the results obtained by Moulin and Mihcak for parallel Gaussian channels. A new stochastic image model has been introduced for the wavelet domain, the so-called EP model, that more accurately treats the data in the regions of edges and textures. We emphasize the crucial role of the model selection on determining capacity. Although the EP model is more accurate, as has been demonstrated in a reference application, and fits better the conditions of proper image decomposition into the parallel Gaussian model, the obtained results considerably deviate from those obtained for the EQ/spike process models argued to yield "upper bounds" on actual capacity. We have in fact shown that the capacity is likely to be much lower than previously thought. We also demonstrate the important role of model mismatch in the extended data-hiding games. Finally, a new attack based on the proposed EP model is presented in the paper.

Acknowledgment

The authors would like to thank F. Pérez-González (University of Vigo, Spain) and P. Moulin (University of Illinois at Urbana-Champaign) for many helpful and interesting discussions.

References

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] M. Belge and E. Miller, "Wavelet domain image restoration using edge preserving prior models," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, Chicago, IL, Oct. 1998, pp. 103–107.

[3] A. Bovik, *Handbook of Image and Video Processing*. New York: Academic, 2000.

[4] J. Chen and A. K. Gupta, "Likelihood procedure for testing change point hypothesis for multivariate Gaussian model," *Random Operators Stochastic Equations*, vol. 3, no. 3, pp. 235–244, 1995.

[5] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[6] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] A. Hjorungnes, J. Lervik, and T. Ramstad, "Entropy coding of composite sources modeled by infinite Gaussian mixture distributions," in *Proc. IEEE Digital Signal Processing Workshop*, Jan., 20–24 1996, pp. 235–238.

[8] R. L. Josshi, H. Jafarkhani, J. H. Kasner, T. R. Fischer, N. Farvardin, M. W. Marcellin, and R. H. Bamberger, "Comparison of different methods of classification in subband coding of images," *IEEE Trans. Inf. Theory*, no. 6, pp. 1473–1486, Nov. 1997.

[9] J. Li and R. M. Gray, *Image Segmentation and Compression Using Hidden Markov Models*. Boston, MA: Kluwer/Plenum (Springer), 2000.

[10] J. Liu and P. Moulin, "Analysis of interscale and intrascale dependencies between image wavelet coefficients," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 1, Vancouver, BC, Canada, Oct. 2000, pp. 669–672.

[11] S. LoPresto, K. Ramchandran, and M. Orhard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proc. Data Compression Conf.*, Snowbird, UT, 1997, pp. 221–230.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[13] S. G. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1997.

[14] H. Man, F. Kossentini, and M. J. Smith, "A family of efficient and channel error resilient wavelet/subband image coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 95–108, Feb. 1999.

[15] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 6, Phoenix, AZ, Mar. 1999, pp. 3253–3256.

[16] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.

[17] M. J. Ruf and J. W. Modesti, "Operational rate-distortion performance for joint source and channel coding of images," *EEE Trans. Inf. Theory*, vol. 8, no. 3, pp. 305–320, Mar. 1999.

[18] P. Moulin, "The role of information theory in watermarking and its application to image watermarking," *Signal Process. (*Special Issue on Information Theoretic Issues in Digital Watermarking*)*, vol. 81, no. 6, pp. 1121–1139, 2001.

[19] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.

[20] P. Moulin and M. K. Mihcak, "A framework for evaluating the data-hiding capacity of image sources," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1029–1042, Sep. 2002.

[21] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[22] B. Natarajan, "Filtering random noise from deterministic signals via data compression," *Proc. IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2595–2605, Nov. 1995.

[23] A. Ortega and K. Ramchandran, "Rate-distortion techniques in image and video compression," *IEEE Trans. Inf. Theory*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[24] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive wiener denoising using a Gaussian scale mixture model in the wavelet domain," presented at the 8th Int. Conf. Image Processing (ICIP), Thessaloniki, Greece, Oct. 2001.

[25] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[26] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[27] B. Vidacovic, *Statistical Modeling by Wavelets*. New York: Wiley, 1999.

[28] J. Vila, O. Koval, and S. Voloshynovskiy, "Facial image compression using overcomplete transforms," in *Proc. SPIE: Electronic Imaging 2004 Image Video Communications Processing VI*, vol. 5308, San Jose, CA, Jan. 2004, pp. 1060–1072.

[29] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. 3rd Int. Workshop Information Hiding*, vol. 1768, Dresden, Germany, Sep. 29–Oct. 1 1999, pp. 212–236.

[30] S. Voloshynovskiy, O. Koval, F. Deguillaume, and T. Pun, "Data hiding capacity-security analysis for real images based on stochastic nonstationary geometrical models," presented at the IS&T/SPIE's Annu. Symp. Electronic Imaging 2003: Image Video Communications Processing V, Santa Clara, CA, Jan., 20–24 2003.

[31] ——, "Visual communications with side information via distributed printing channels: Extended multimedia and security perspectives," presented at the SPIE Photonics West, Electronic Imaging 2004, Multimedia Processing Applications, San Jose, CA, Jan. 18–22, 2004.

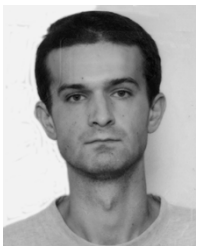
[32] S. Voloshynovskiy, O. Koval, and T. Pun, "Wavelet-based image denoising using nonstationary stochastic geometrical image priors," presented at the IS&T/SPIE's Annu. Symp. Electronic Imaging 2003: Image Video Communications Processing V, Santa Clara, CA, Jan., 20–24 2003.

[33] C. Weidmann and M. Vetterli, "Rate-distortion analysis of spike processes," presented at the Data Compression Conf., Snowbird, UT, Mar. 1999.

[34] Y.Yoo , A. Ortega, and B. Yu, "Image subband coding using context-based classification and adaptive quantization," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1702–1715, Dec. 1999.

Sviatoslav Voloshynovskiy received the Radio Engineer degree from Lviv Polytechnic Institute, Lviv, Ukraine, in 1993 and the Ph.D. degree in electrical engineering from State University "Lvivska Polytechnika," Lviv, Ukraine, in 1996.

From 1998 to 1999, he was a visiting scholar with the University of Illinois at Urbana-Champaign. Since 1999, he has been with the University of Geneva, Switzerland, where he is currently an Associate Professor with the Department of Computer Science and head of the Stochastic Image Processing group. His current research interests are in information-theoretic aspects of digital data hiding, visual communications with side information, and stochastic image modeling for denoising, compression, and restoration. He has coauthored over 100 journal and conference papers in these areas and holds nine patents. He has served as a consultant to private industry in the above areas.



Oleksiy Koval received the Master's and Ph.D. degrees, both in electrical engineering, from the National University "Lvivska Politechnika," Lviv, Ukraine, in 1996 and 2002, respectively. He also received the Ph.D. degree in stochastic image modeling from the University of Geneva, Switzerland, in 2004.

From 1996 to 2001, he was a Researcher and Ph.D. student with the Department of Synthesis, Processing and Identification of Images, Institute of Physics and Mechanics, Lviv, Ukraine. Since 2002, he has been with Stochastic Image Processing Group, Computer Vision and Multimedia Laboratory, University of Geneva, where he is currently a Postdoctoral Fellow. Research interests cover stochastic image modeling for different image processing applications, digital watermarking, information theory, and communications with side information.

M. Kivanc Mihcak was born in Turkey. He received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1996 (valedictorian) and the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, in 1999 and 2002 respectively.

Between 1996 and 2002, he was with the Electrical and Computer Engineering Department at the University of Illinois, Urbana-Champaign, in the Image Formation and Processing Group, at the Beckman Institute. Currently, he is with the Crypto Group, Microsoft Research, Redmond, WA. His research interests include information hiding, watermarking, fingerprinting, hashing, multimedia security, as well as statistical signal processing, information theory, data compression, and wavelets.

Thierry Pun received the Ph.D. degree in image processing from the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland, in 1982.

In 1986, he joined the University of Geneva, Switzerland, where he is currently Full Professor in the Computer Science Department and head of the Computer Vision and Multimedia Laboratory. Since 1979, he has been active in various domains of image processing, image analysis, and computer vision. He has authored or coauthored over 200 journal and conference papers in these areas as well as seven patents and led or participated in a number of national and European research projects. His current research interests, related to the design of multimedia information systems and multimodal interaction, focus on data hiding, image and video watermarking, image and video content-based information retrieval systems, EEG signals analysis, and brain–computer interaction.