Facial Image Compression Based on Structured Codebooks in Overcomplete Domain

J. E. Vila-Forcén, S. Voloshynovskiy, O. Koval, and T. Pun

Stochastic Image Processing Group, CUI, University of Geneva, 24 rue du Général-Dufour, Geneva 1211, Switzerland

We advocate facial image compression technique in the scope of distributed source coding framework. The novelty of the proposed approach is twofold: image compression is considered from the position of source coding with side information and, contrarily to the existing scenarios where the side information is given explicitly; the side information is created based on a deterministic approximation of the local image features. We consider an image in the overcomplete transform domain as a realization of a random source with a structured codebook of symbols where each symbol represents a particular edge shape. Due to the partial availability of the side information at both encoder and decoder, we treat our problem as a modification of the Berger-Flynn-Gray problem and investigate a possible gain over the solutions when side information is either unavailable or available at the decoder. Finally, the paper presents a practical image compression algorithm for facial images based on our concept that demonstrates the superior performance in the very-low-bit-rate regime.

1. INTRODUCTION

The urgent demand of efficient image representation is recognized by the industry and research community. Its necessity is highly increased due to the novel requirements of many authentication documents such as passports, ID cards, and visas as well as recent extended functionalities of wireless communication devices. The document, ticket, or even entry pass personalization are often requested in many authentication or identification protocols. In most cases, classical compression techniques developed for generic applications are not suitable for these purposes.

Wavelet-based [1, 2] lossy image compression techniques [3–6] have proved to be the most efficient from the rate-distortion point of view for the rate range of 0.2–1 bits per pixel (bpp). The superior performance of this class of algorithms is justified by both decorrelation and energy compaction properties of the wavelet transform and by the efficient adaptive both interband (zero trees [5]) and intraband (estimation quantization (EQ) [7, 8]) models that describe the data in the wavelet subbands. Recent results in wavelet-based image compression show that some modest performance improvement (in terms of peak signal-to-noise ratio (PSNR) up to 0.3 dB) could be achieved either taking into account the nonorthogonality of the transform [9] or using more complex higher-order context models of wavelet coefficients [10].

During years, a standard benchmark database of images for wavelet-based compression algorithm evaluation was used. It includes several 512×512 grayscale test images (like Lena, Barbara, Goldhill) and the verification was performed for the rates 0.2–1 bpp. In some applications, which include person authentication data like photo images or fingerprint images, the operational conditions might be different. In this case, especially for strong compression (below 0.15 bpp), the resulting image quality of the state-of-the-art algorithms is not satisfactory enough (Figure 1). Therefore, for this kind of applications more advanced techniques are needed to satisfy the fidelity constrains.

In this paper, we address the problem of classical wavelet-based image compression enhancement by using side information within a framework of distributed coding of correlated sources. Recently, it was practically shown that it is possible to achieve a significant performance gain when the side information is available at the decoder, while the encoder has no access to the side information [11]. Using the side information from an auxiliary analog additive white Gaussian noise (AWGN) channel in the form of a noisy copy of the input image at the decoder, it was reported a PSNR enhancement in the range of 1–2 dB depending on the test image and the compression rate. It could be noted that the performance of this scheme strongly depends on the state of the auxiliary channel, which should be known in advance at the encoding stage. Moreover, it is assumed that the noisy copy
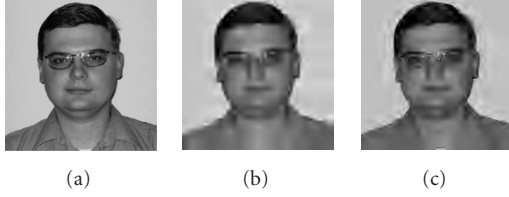
(a)   (b)   (c)

FIGURE 1: (a) 256×256 8-bit test image *Slava*. Results of compression with rate 0.071 bits per pixel (bpp) using (b) JPEG2000 standard software (PSNR is 25.09 dB) and (c) state-of-the-art EQ coder (PSNR is 26.36 dB).
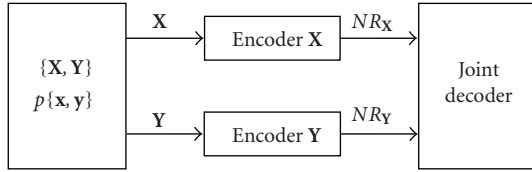

FIGURE 2: Slepian-Wolf coding.

of the original image should be directly available at the decoder. This situation is typical for the distributed coding in the remote sensing applications or can be simulated as in the case of analog and digital television simulcast [11]. In the case of single-source compression, the side information is not directly available at the decoder.

The main goal of this paper consists in the development of a concept of single-source compression within a distributed coding framework using *virtually* created side information. This concept is based on the accurate approximation of a source data using a structured codebook, which is shared by the encoder and decoder, and the communication of the residual approximation term within the classical wavelet-based compression paradigm.

The paper is organized as follows. In Section 2, fundamentals of source coding with side information are presented. In Section 3, an approach for single-source distributed lossy coding is introduced. A practical algorithm for a very-low-bit-rate compression of passport photo images is developed in Section 4. Section 5 contains the experimental results and Section 6 concludes the paper.

Notation 1. *Scalar random variables are denoted by capital letters X, bold capital letters $\mathbf{X}$ denote vector random variables, letters x and $\mathbf{x}$ are reserved to denote the realization of scalar and vector random variables, respectively. The superscript N is used to denote N-length vectors $x^N = \mathbf{x} = \{x_1, x_2, \ldots, x_N\}$, where the ith element is denoted as x_i. $X \sim p_X(x)$ or $X \sim p(x)$ indicates 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 $E[X]$. $H(X)$, $H(X, Y)$, $H(X \mid Y)$ denote the entropy of the random variable X, the joint entropy of the random variables X and Y, and the conditional entropy of the random variable X given Y, respectively. By $I(X; Y)$ and $I(X; Y \mid Z)$, we denote the mutual information*
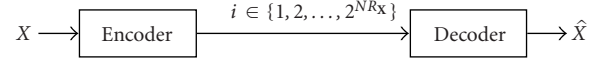

FIGURE 3: Lossy source coding system without side information.

between the random variables X and Y, and the conditional mutual information between the random variables X and Y given the random variable Z, respectively. $R_{\mathbf{X}}$ denotes the rate of communications for the random variable $\mathbf{X}$. Calligraphic font $\mathcal{X}$ is used to indicate sets $X \in \mathcal{X}$, and $|\mathcal{X}|$ indicates the cardinality of a set. $\mathcal{R}^+$ is used to represent the set of positive real numbers.

2. DISTRIBUTED CODING OF CORRELATED SOURCES

2.1. Slepian-Wolf encoding

Assume that it is necessary to encode two discrete-alphabet pair wisely independent and identically distributed (i.i.d.) random variables $\mathbf{X}$ and $\mathbf{Y}$ with joint distribution $p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^{N} p_{X_k Y_k}(x_k, y_k)$. A Slepian-Wolf [12, 13] code allows performing lossless encoding of $\mathbf{X}$ and $\mathbf{Y}$ individually using two separate encoders, and the decoding is performed jointly as presented in Figure 2. Using a random binning argument, it was shown that the efficiency of such a code is the same as in the case when joint encoding is used. It means that the encoder bit rates pair $(R_{\mathbf{X}}, R_{\mathbf{Y}})$ is achievable when the following relationships hold:

$$\begin{aligned} R_{\mathbf{X}} &\geq H(\mathbf{X} \mid \mathbf{Y}), \\ R_{\mathbf{Y}} &\geq H(\mathbf{Y} \mid \mathbf{X}), \\ R_{\mathbf{X}} + R_{\mathbf{Y}} &\geq H(\mathbf{X}, \mathbf{Y}). \end{aligned} \tag{1}$$

2.2. Lossy compression with side information

In the lossy compression setup, it is necessary to achieve the minimal possible distortions for a given target coding rate. Depending on the availability of side information, several possible scenarios exist [14].

No side information is available

Imagine that it is needed to represent an i.i.d. source sequence $\mathbf{X} \sim p_X(x)$, $\mathbf{X} \in \mathcal{X}^N$ using the encoding mapping $f_E : \mathcal{X}^N \rightarrow \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\}$ and the decoding mapping $f_D : \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\} \rightarrow \mathcal{X}^N$ with the minimum average bit rate R bits per element. The fidelity of representation is evaluated using the average distortion $D = (1/N) \sum_{k=1}^{N} E[d(x_k, \hat{x}_k)]$, where the distortion measure $d(x, \hat{x})$ is determined in general as a mapping $\mathcal{X}^N \times \widehat{\mathcal{X}}^N \rightarrow \mathcal{R}^+$. Due to Shannon [12, 15], it is well known that the optimal performance of such a compression system (Figure 3) (the minimal achievable rate for certain distortion level) is determined by the rate-distortion function,

$$R_{\mathbf{X}}(D) = \min_{p(\hat{x}|x): \sum_{\hat{x}, x} p(\hat{x}|x) d(\hat{x}, x) \leq D} I(X; \hat{X}). \tag{2}$$
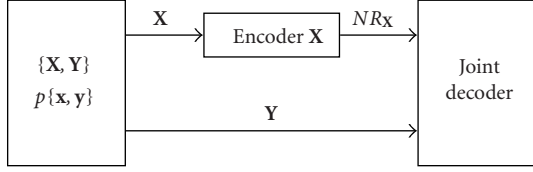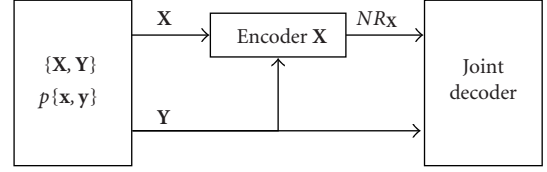
FIGURE 4: Wyner-Ziv coding.

FIGURE 5: Berger-Flynn-Gray coding.

Side information is available only at the encoder

In this case, the performance limits coincide with the previous case and the rate-distortion function could be determined using (2) [16].

Side information is available only at the decoder (Wyner-Ziv coding)

Fundamental performance limits of source coding systems with side information available only at the decoder (Figure 4) were established by Wyner and Ziv [12, 17]. The Wyner-Ziv problem could be formulated in the following way: given the side information only at the decoder, what will be the minimum rate $R_{\mathbf{X}}$ necessary to reconstruct the source $\mathbf{X}$ with average distortion less than or equal to a given distortion value D? By other words, assume that we have a sequence of independent drawings of pairs $\{X_k, Y_k\}$ of dependent random variables, $\{\mathbf{X}, \mathbf{Y}\} \sim p(\mathbf{x}, \mathbf{y})$, $(\mathbf{X}, \mathbf{Y}) \in \mathcal{X}^N \times \mathcal{Y}^N$. Our goal is to construct an $R_{\mathbf{X}}$-bits-per-element encoder $f_E : \mathcal{X}^N \to \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\}$ and joint decoder $f_D : \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\} \times \mathcal{Y}^N \to \widehat{\mathcal{X}}^N$ such that the average distortion satisfies the fidelity constraint:

$$E[d(\mathbf{X}, f_D(\mathbf{Y}, f_E(\mathbf{X})))] = \sum_{x,\hat{x}} p(x, y) p(\hat{x} \mid x, y) \le D. \quad (3)$$

Using the asymptotic properties of random codes, it was shown [17] that the set of achievable rate-distortion pairs of such a coding system will be bounded by the Wyner-Ziv rate-distortion function:

$$R_{\mathbf{X}}(D)_{X|Y}^{WZ} = \min_{p(u|x)p(\hat{x}|x,y)} [I(U; X) - I(U; Y)], \quad (4)$$

where the minimization is performed over all $p(u \mid x) p(\hat{x} \mid x, y)$ and all decoder functions f_D satisfying the fidelity constraint (3). U is an auxiliary random variable such that $|\mathcal{U}| \le |\mathcal{X}| + 1$ and $Y \to X \to U$ forms a Markov chain. Hence, (4) could be rewritten as follows:

$$R_{\mathbf{X}}(D)_{X|Y}^{WZ} = \min_{p(u|x)p(\hat{x}|x,y)} I(U; X \mid Y), \quad (5)$$

where the minimization is performed over all $p(u \mid x) p(\hat{x} \mid x, y)$ subject to the fidelity constraint (3).

It is worth to note that for the case of zero distortions, the Wyner-Ziv problem corresponds to the Slepian-Wolf problem, that is, $R_{\mathbf{X}}(0)_{X|Y}^{WZ} = H(X \mid Y)$.

Lossy compression of correlated sources (Berger-Flynn-Gray coding)

This problem was investigated by Berger [18] and Flynn and Gray [19], and the general scheme is presented in Figure 5.

As in the previous case, Berger-Flynn-Gray coding refers to the sequence of pairs $\{\mathbf{X}, \mathbf{Y}\} \sim p(\mathbf{x}, \mathbf{y})$, $(\mathbf{X}, \mathbf{Y}) \in \mathcal{X}^N \times \mathcal{Y}^N$, where now $\mathbf{Y}$ is available at both encoder and decoder, while in the Wyner-Ziv problem it was available only at the decoder. It is necessary to construct an $R_{\mathbf{X}}$-bits-per-element joint coder $f_E : \mathcal{X}^N \times \mathcal{Y}^N \to \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\}$ and a joint decoder $f_D : \{1, 2, \ldots, 2^{NR_{\mathbf{X}}}\} \times \mathcal{Y}^N \to \widehat{\mathcal{X}}^N$ such that the average distortion satisfies $E[d(\mathbf{X}, f_D(\mathbf{Y}, f_E(\mathbf{X}, \mathbf{Y})))] \le D$. In this case, the performance limits are determined by the conditional rate-distortion function,

$$R_{\mathbf{X}}(D)_{X|Y}^{BFG} = \min_{p(\hat{x}|x,y)} I(\hat{X}; X \mid Y), \quad (6)$$

where the minimization is performed over all $p(\hat{x} \mid x, y)$ subject to the fidelity constraint (3). The Berger-Flynn-Gray rate in (6) is, in general, smaller than the Wyner-Ziv rate (5) since the availability of the correlated source $\mathbf{Y}$ at both encoder and decoder makes possible to reduce the ambiguity about $\mathbf{X}$.

Comparing the rate-distortion performance of different coding scenarios with the side information, it should be noted that, in general, the following inequalities hold [20]:

$$R_{\mathbf{X}}(D) \ge R_{\mathbf{X}}(D)_{X|Y}^{WZ} \ge R_{\mathbf{X}}(D)_{X|Y}^{BFG}. \quad (7)$$

The last inequality becomes equality, that is, $R_{\mathbf{X}}(D)_{X|Y}^{WZ} = R_{\mathbf{X}}(D)_{X|Y}^{BFG}$, only for the case of Gaussian distribution of the source $\mathbf{X}$ and mean square error (MSE) distortion measure. For any other pdf, performance loss exists in the Wyner-Ziv coding. It was shown in [20] that this loss is upper bounded by 0.5 bit,

$$R_{\mathbf{X}}(D)_{X|Y}^{WZ} - R_{\mathbf{X}}(D)_{X|Y}^{BFG} \ge 0.5. \quad (8)$$

Therefore, due to the fact that natural images have highly non-Gaussian statistics [8, 21, 22], compression of this data using the Wyner-Ziv strategy will always lead to the performance loss. The main goal of subsequent sections consists in the extension of the classical distributed coding setup to the case of a single-source coding scenario in the very-low-bit-rate regime.
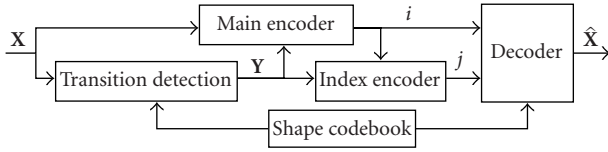
FIGURE 6: Block diagram of single-source distributed coding with side information.
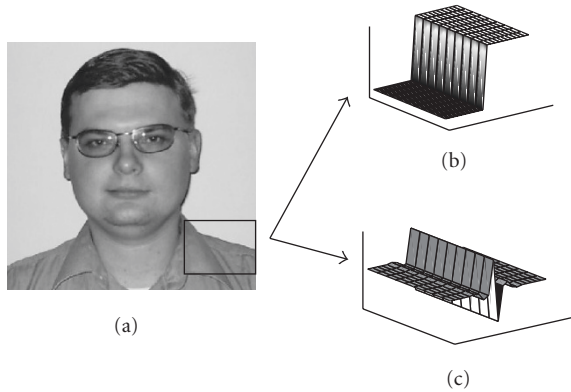

FIGURE 8: Shape cosets from the shape codebook **Y**.

FIGURE 7: (a) Test image *Slava* and its fragment (marked by square): two-region modeling of the fragment, (b) in the coordinate domain, and (c) in the nondecimated wavelet transform domain.

one represents edges and textures. It is supposed that the samples from the latter subset propagate along the transition direction (Figure 7(c)). Accurate tracking of the region separation boundary in the coordinate domain setup or transition profile propagation in the transform domain setup allowed to achieve image denoising results that are among the state-of-the-art for the case of AWGN [24].

3. PRACTICAL APPROACH: DISTRIBUTED SOURCE CODING OF A SINGLE SOURCE

3.1. Coding block diagram

The block diagram of a practical single-source distributed coding system with side information is presented in Figure 6. The system consists of two main functional parts. The first part includes the main encoder that is working as a classical quantization-based lossy coder with varying rates. The second part includes the block of transition detection that approximates the image edges and creates some auxiliary image **Y**, as a close approximation to **X**. The index encoder communicates the parameters of approximation model to the decoder. The shape codebook is shared by both transition detection block and decoder.

The intuition behind our approach is based on the assumption that natural images in the coordinate domain can be represented as a union of several stationary regions of different intensity levels or in the nondecimated wavelet transform domain [23] using edge process (EP) model. This assumption and the EP model have been used in our previous work in image denoising where promising results have been reported [24].

Under the EP model, an image in the coordinate domain (Figure 7(a)) is composed of a number of nonoverlapping smooth regions (Figure 7(b)). Accordingly, in the critically sampled or nondecimated wavelet transform domain, it is represented as a union of two types of subsets: the first one contains all samples from flat image areas, while the second
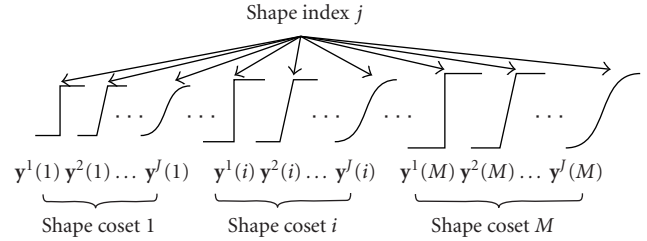
3.2. Codebook construction

Contrarily to the image denoising setup, in the case of lossy wavelet-based image compression we are interested in considering not the behavior of edge profile along the direction of edge propagation, but the different edge profiles. Due to the high variability of edge shapes in real images and the corresponding complexity of the approximation problem, we will exploit a structured codebook for shape representation. It means that several types of shapes will be used to construct a codebook where each codeword represents one edge of some magnitude. A schematic example of such a codebook is given in Figure 8, where several different edge profiles are exploited for image approximation. This structured codebook has a coset-based structure, where each coset contains the selected triple of edge profiles of a certain amplitude.

More formally, the structured codebook $\mathbf{Y} = \{\mathbf{y}(i)\}$, where $i = 1, 2, \ldots, M$, and a coset (9) can be represented as in Figure 9:

$$\mathbf{y}(i) = \begin{cases} y_1^1(i) & y_2^1(i) & \cdots & y_N^1(i) \\ y_1^2(i) & y_2^2(i) & \cdots & y_N^2(i) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^J(i) & y_2^J(i) & \cdots & y_N^J(i) \end{cases}. \tag{9}$$

Here, $\mathbf{y}^j(i)$ represents the shape j from the shape coset i. All shape cosets i consist of the same shape profiles, that is, $j \in \{1, 2, \ldots, J\}$, and $i \in \{1, 2, \ldots, M\}$ for the example presented in Figure 8.

Important points about the codebook are as follows: (a) it is image independent, (b) the considered shapes are unidimensional, (c) the codewords shape could be expressed analytically, for instance, using apparatus of splines, and (d) the codebook dimensionality is determined by the type of transform used and the compression regime. Therefore, a concept of successive construction refinement [25] of the codebook
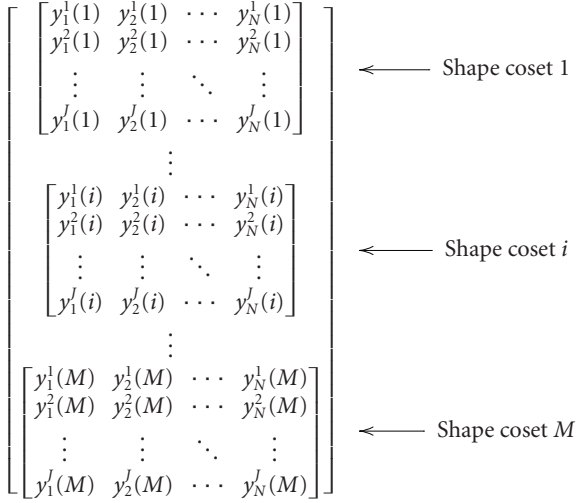
$$\begin{bmatrix} \begin{bmatrix} y_1^1(1) & y_2^1(1) & \cdots & y_N^1(1) \\ y_1^2(1) & y_2^2(1) & \cdots & y_N^2(1) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^J(1) & y_2^J(1) & \cdots & y_N^J(1) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} y_1^1(i) & y_2^1(i) & \cdots & y_N^1(i) \\ y_1^2(i) & y_2^2(i) & \cdots & y_N^2(i) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^J(i) & y_2^J(i) & \cdots & y_N^J(i) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} y_1^1(M) & y_2^1(M) & \cdots & y_N^1(M) \\ y_1^2(M) & y_2^2(M) & \cdots & y_N^2(M) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^J(M) & y_2^J(M) & \cdots & y_N^J(M) \end{bmatrix} \end{bmatrix}$$

⟵ Shape coset 1

⟵ Shape coset i

⟵ Shape coset M

FIGURE 9: Structured codebook: shape coset index i (or magnitude) is communicated explicitly by the main encoder as transition location and magnitude of quantized coefficients, and shape index j ($j \in \{1, 2, \ldots, J\}$) is encoded by index encoder.

might be used. The intuition behind this approach could be explained using the coarse-fine quantization framework presented in Figure 10.

It means that for the case of high compression ratios, when there is not much rate to code the shape index, a single shape profile will be used (like a coarse quantizer). In other regimes (at medium or at high rates), it is possible to improve the fidelity of approximation adding more edge shapes to the codebook. In this case, we could assume that the high-rate quantization assumption becomes valid.

The task of real edge approximation according to the shape codebook can be formulated, for instance, like a classical ℓ^2 norm approximation problem,

$$\widetilde{\mathbf{y}}^j(i) = \operatorname*{argmin}_{\{\mathbf{y}^j(i)\}, 1 \le i \le M, 1 \le j \le J} \left\| \mathbf{x} - \mathbf{y}^j(i) \right\|^2, \quad (10)$$

where the minimization is performed over the whole codebook in each image point.

3.3. Practical implementation: high-, medium-, and low-bit-rate regimes

It is clear that in the presented setup, the computational complexity of image approximation in each point will be significant, and can be unacceptable in some realtime application scenarios. To simplify the situation, searching space dimensionality might be significantly reduced using techniques that simplify the edge localization. Canny edge detector [26] can be used for this purpose.

The edge of a real image could be considered as a noisy or distorted version of the corresponding codeword $\{y_1^j(i), y_2^j(i), \ldots, y_N^j(i)\}$ (edge shape) with respect to the codebook $\mathbf{Y}$, that is, some correlation between an original
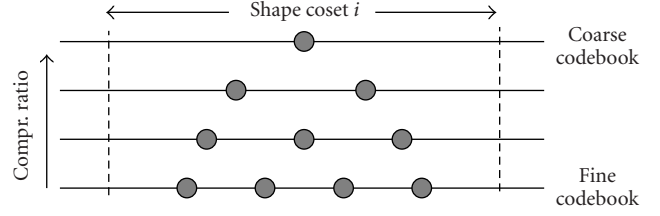

FIGURE 10: Successive refinement codebook construction.

edge and a codeword can be assumed. Therefore, the structure of the codebook is similar to the structure of a channel coset code [27], meaning that the distance between codewords of equal magnitude (Figure 8) in the transform domain should be large enough to perform correct shape approximation.

The coding strategy can be performed in a distributed manner. In general, the main encoder performs the quantization of the edge and communicates the corresponding indices of reconstruction levels to the decoder. This information is sufficient to determine the shape coset index i at the decoder for different compression regimes, including even very-low-bit-rate regime (besides the case when quantization to zero is performed). The index j of edge shape within a coset is communicated by the index encoder to the decoder. Having the coset index and the shape index, the decoder looks in the coset bin i for $\mathbf{y}^j(i)$ and generates the reproduction sequence $\hat{\mathbf{x}} = f_D(\hat{\mathbf{x}}'(i), \widetilde{\mathbf{y}}^j(i))$, where $\hat{\mathbf{x}}'(i)$ is the data reproduced at the decoder based only on the index i.

In the case of high rates, the main encoder performs a high-rate (high-accuracy) approximation of the image edges. It means that the index encoder does not produce any output, that is, both edge magnitude and edge shape could be reconstructed directly from the information contained in the main decoder bit stream. Therefore, the role of side information represented by the fine codebook consists in the compensation of quantization noise influence.

For middle rates, the edge magnitude prediction is still possible using the main encoder bitstream. However, the edge shape approximation accuracy for this regime is not high enough to estimate the edge shape and its index should be communicated to the decoder by the index encoder. One can note that in such a way we end up with vector-like edge quantization using the off-line designed edge codebook. The role of the side information remains similar to the previous case and targets the compensation of quantization error.

At low rates, a single codeword (optimal in the mean square error sense) should be chosen to represent all shapes within the given image (coarse codebook in Figure 10). In more general case, one can choose a single shape codeword that is the same for all images. This is a valid assumption for the compression of image databases with the same type of images. Contrarily to the above case of middle rates, the decoder operates with a single edge codeword that can be applied to all cases where the edge coefficients are partially preserved in the corresponding subbands. Moreover, the edge

reconstruction is possible even when the edge coefficients in some subbands are completely discarded by the deadzone quantization.

The practical aspects of implementation of the presented single source coding system with side information are out of the scope of the paper. In the following section, we will present an application of the proposed framework to the very-low-bit-rate compression of passport photo images.

4. DISTRIBUTED CODING OF IMAGES WITH SYMMETRIC SIDE INFORMATION: COMPRESSION OF PASSPORT PHOTOS AT VERY LOW BIT RATES

In this section, the case of single source distributed coding system with side information is discussed for the case of very-low-bit-rate (less than 0.1 bpp) compression of passport photo images. The importance of this task is justified by the urgent necessity to store personal information on the capacity restricted media authentication documents that include passports, visas, ID cards, driver licenses, and credit cards using digital watermarks, barcodes, or magnetic strips. In this paper, we assume that the images of interest are 8-bit gray scale images of 256×256 size. As it was shown in Figure 1, existing compression tools are unable to provide the satisfactory quality solution to this task.

The scheme presented in Figure 6 is used as a basic setup for this application. As it was discussed earlier, for the case of very-low-bit-rate regime, only one shape profile (simple step edge) is exploited. Therefore, index encoder is not used in this particular case since only one index is possible as its output and, therefore, it is known a priory by the decoder. Certainly, better performance can be expected if one approximates transitions using complete image codeword (Figure 6). The price to pay for that is additional $\log_2 J$ bits of side information per shape, where J is the number of edge shapes within each coset.

In the next subsections, we discuss in details the particularities of encoding and decoding at the very-low-bit rates.

4.1. Transition detection

Encoding

Due to the fact that high-contrast edges consume a significant amount of the allocated bit budget for the complete image storage, it would be beneficial from the reconstructed image quality perspective to reduce the ambiguity about these image features.

On the first step, the position of the principal edges (the edges with the highest contrast) is detected using the Canny edge detector. Due to the fact that the detection result is not always precise (some position deviation is possible), actual transition location is detected using the zero-crossing concept.

Zero-crossing concept is based on the fact that in the nondecimated wavelet transform domain (algorithm *a trois* [23]
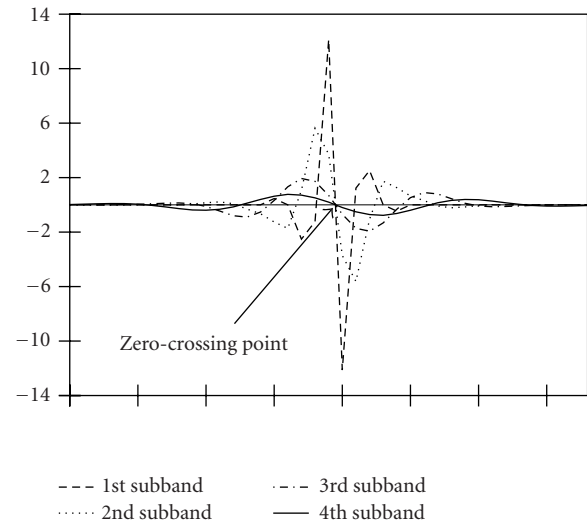

FIGURE 11: Zero-crossing concept: 4-level decomposition of the step edge in the nondecimated domain.

is used for its implementation) all the representations of an ideal step edge from different decomposition levels in the same spatial orientation cross the horizontal axis in the same point referred to as the *zero-crossing* point (Figure 11). This point coincides with a spatial position of the transition in the coordinate domain. Besides, the magnitudes of principle peaks (maximum and minimum values of the data in the vicinity of transition in the nondecimated domain) of the components are related pairwise from high to low frequencies with certain fixed ratios which are known in advance. Therefore, when the position of the *zero-crossing* point is given and, at least, one of the component peak magnitudes is known from original step edge, it is possible to predict and to reconstruct the missing data components with no error.

Consequently, if it is known at the decoder that an ideal step edge with a given amplitude is presented in a given image location, it is possible to assign zero rate to the predictable coefficients at the encoder, allowing higher quality reconstruction of unpredictable information.

Decoding

For this mode, it is assumed that the low-resolution version of the original data obtained using main encoder bitstream is already available. Detection of the coarse positions of main edges is performed on the interpolated image analogically to the encoder case. To adjust detection results, a new concept of *zero-crossing* detection is used.

In the targeted very-low-bit-rate compression scenario, the data are severely degraded by quantization. To make zero-crossing detection more reliable in this case, more levels of the nondecimated wavelet transform can be used. The additional reliability is coming from the fact that the data at the very low-frequency subbands almost do not suffer from quantization. The gain in this case is limited by the information that is still presented at these low frequencies: only the
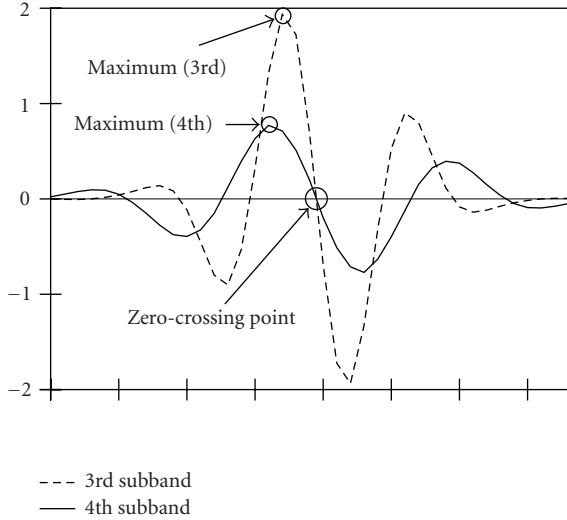
FIGURE 12: Zero-crossing concept: decoding stage.

--- 3rd subband
— 4th subband

edges propagating in all the subbands could be detected in such a way.

In order to reconstruct high-frequency subbands, both edge position and edge magnitude are predicted using low-frequency subbands. In Figure 12, the position of the zero-crossing point is estimated based on the data from the 3rd and 4th subbands. Having their maximum magnitude values, the reconstruction of high frequency subbands can be performed accurately based on the fixed magnitude relationships (Figure 11).

4.2. Main encoder

To justify the main encoder structure, we would like to point out that the main gain achieved recently in wavelet-based lossy transform image coding is due to the accuracy of the underlying stochastic image model.

One of the most efficient and accurate stochastic image models that represent images in the wavelet transform domain is based on the parallel splitting of the Laplacian source firstly introduced by Hjorungnes et al. [28]. The main underlying assumption here is that global i.i.d. zero-mean Laplacian data can be represented, without loss according to the Kullback-Leibler divergence, using an infinite mixture of Gaussian pdfs with zero-mean and exponentially distributed variances,

$$\frac{\lambda}{2}e^{(-\lambda|x|)} = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}e^{(x^2/2\sigma^2)}\lambda e^{(-\lambda\sigma^2)}d\sigma^2, \qquad (11)$$

where λ is the parameter of the Laplacian distribution. The Laplacian distribution is often used to model the global statistics of the high-frequency wavelet coefficients [8, 21, 22].

Hjorungnes et al. [28] were the first who demonstrated that, if the side information (the local variances) are available

at both encoder and decoder, the gain in the rate-distortion sense of coding the Gaussian mixture instead of the global Laplacian source is given by

$$R_L(D) - R_{MG}(D) \approx 0.312 \text{ bit/sample}, \qquad (12)$$

where $R_L(D)$ and $R_{MG}(D)$ denote the rate-distortion functions for the global i.i.d. Laplacian source and the Gaussian mixture, respectively.

The practical problem of the side information communication to the decoder was elegantly solved in [7, 8]. The developed EQ coder is based on the assumption of the slow varying nature of the local variances of the high-frequency subband image samples. As a consequence, this variance can be accurately estimated (predicted) given its quantized causal neighborhood.

According to the EQ coding strategy, the local variances of the samples in the high-frequency wavelet subbands are estimated based on the causal neighborhood using maximum likelihood strategy. When it is available, the data from the parent subband are also included to enhance the estimation accuracy.

At the end of the estimation step, the coefficients are quantized using a uniform threshold quantizer selected accordingly to the results of the rate-distortion optimization. In particular, the Lagrange functional should be minimized, that on the sample level is given by

$$y_i = r_i + \lambda d_i, \qquad (13)$$

where r_i is the rate corresponding to the entropy of the quantizer output applied to the ith sample, d_i is the corresponding distortion, and λ is the Lagrange multiplier. The encoding of the quantized data is performed using the bin probabilities of the quantizers, where the samples fall, by an arithmetic coder.

While at the high-rate regime the approximation of the local variance field by its quantized version is valid, in the case of low rates it fails. The reason for that is the quantization to zero most of the data samples that makes local variance estimation extremely inaccurate.

The simple solution proposed in [7, 8] consists in the placement of all the coefficients that fall into the quantizer deadzone in the so-called *unpredictable class*, and the rest in the so-called *predictable class*. The samples of the first one are considered to be distributed globally as an i.i.d. generalized Gaussian distribution, while the infinite Gaussian mixture model is used to capture the statistics of the samples in the second one. This separation is performed using a simple rate-dependent thresholding operation. The parameters of the unpredictable class are exploited in the rate-distortion optimization and are sent to the decoder as side information.

The experimental results presented in [7, 8] allow to conclude about the state-of-the-art performance of this technique in the image compression application.

TABLE 1: Benchmarking of the developed compression method versus existing lossy encoding techniques.

Bytes (bpp)	*Slava* (PSNR, dB)				*Julien* (PSNR, dB)				*Jose* (PSNR, dB)			
	ROI JPEG 2000	ROI SPIHT	EQ	DSSC	ROI JPEG 2000	ROI SPIHT	EQ	DSSC	ROI JPEG 2000	ROI SPIHT	EQ	DSSC
300 (0.037)	21.04	25.27	10.33	25.93	19.87	22.82	9.76	23.35	20.76	26.11	10.39	27.29
400 (0.049)	22.77	26.08	18.56	26.81	21.47	23.43	17.94	23.89	23.21	27.30	19.86	28.27
500 (0.061)	24.92	26.85	22.36	27.41	23.07	23.81	21.86	24.15	25.50	28.20	25.09	28.81
600 (0.073)	25.78	27.41	25.96	27.85	23.78	24.17	22.81	24.28	26.39	28.74	27.09	29.10
700 (0.085)	26.66	27.96	27.12	28.09	24.53	24.44	23.61	24.50	28.08	29.31	28.37	29.35
800 (0.099)	27.39	28.56	27.71	28.16	25.04	24.68	24.24	24.56	28.72	29.89	29.31	29.46

(a)   (b)

FIGURE 13: Test image *Slava*: (a) region of interest and (b) background four-quadrant splitting.

Motivated by the EQ coder performance, we designed our main encoder using the same principles with several modifications as follows:

(i) at the very-low-bit-rate regime, most of the information at the first and the second wavelet decomposition levels is quantized to zero. We assume that all the data about strong edges could be reconstructed with some precision using the side information and do not allocate any rate to these subbands;

(ii) high-frequency subbands of the third decomposition level are compressed using a region of interest strategy (Figure 13(a)), where the region of interest is indicated using three extra bytes. The image regions outside of the region of interest will be reconstructed using low-frequency information, and four extra bytes for the mean brightness of the background of the photo image in four quadrants (Figure 13(b));

(iii) a 3×3 causal window is applied for local variance estimation;

(iv) no parent dependencies are taken into account on the stochastic image model, and only samples from the given subband are used [29].

The actual bitstream from the encoder is constituted by the data from the EQ encoder, three bytes determining the position of the rectangular region-of-interest, and four bytes characterizing the background brightness.

4.3. Index encoder

As it was mentioned in the previous subsection, only one edge profile (the step edge) is used at the very-low-rate regime. Thus, index encoder does not produce any output.

4.4. Decoder

The decoder performs the reconstruction of the compressed data using the main encoder output and the available side information. The bitstream of the main encoder is decompressed by the EQ decoder. The fourth wavelet transform decomposition level is decompressed using classical algorithm version, and the third level is reconstructed using region of interest EQ decoding.

Having two lowpass levels of decomposition, the low-resolution reconstruction (with two high-frequency decomposition levels equal to zero) of the original photo using wavelet transform is obtained. Final reconstruction of high-quality data is performed based on the interpolated image, and the transition detection block information in the nondecimated wavelet transform domain.

5. EXPERIMENTAL RESULTS

In this section, we present the experimental results of very-low-bit-rate passport photo compression based on the proposed framework of distributed single source coding with symmetrical side information (DSSC). A set of 11 images were used in our experiments. The results for three of them are presented in Table 1, Figures 14 and 15 versus those provided by the standard EQ algorithm as well as JPEG2000 with region of interest coding (ROI-JPEG2000) [30] and set partitioning in hierarchical trees algorithm with region of interest coding (ROI-SPIHT) [31].
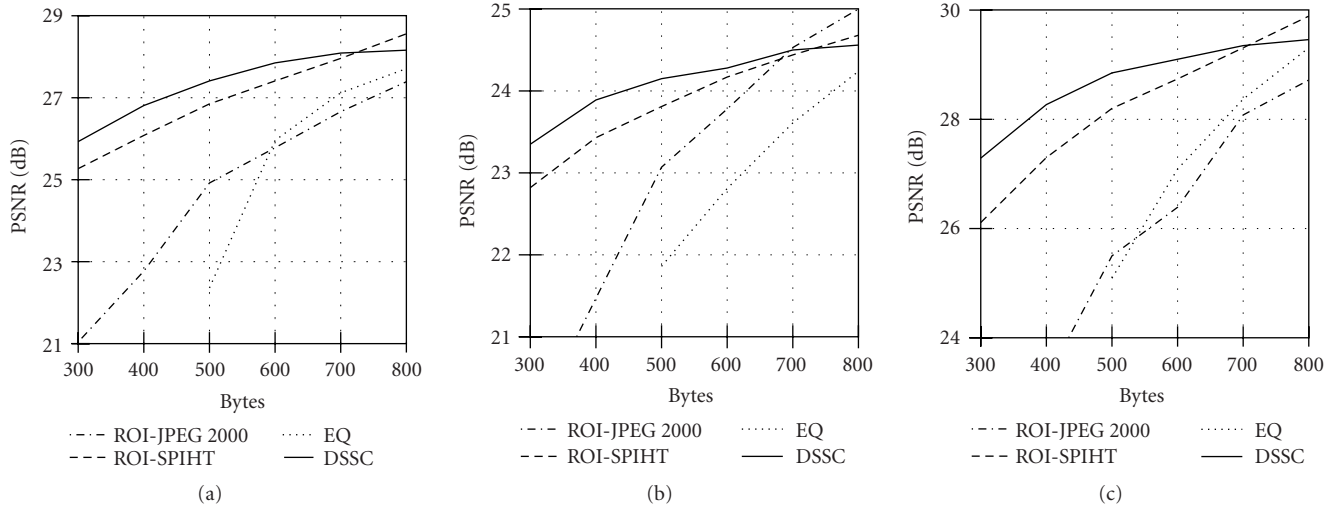
FIGURE 14: Benchmarking of the developed compression method versus existing lossy encoding techniques: (a) *Slava*, (b) *Julien*, and (c) *Jose* test images.
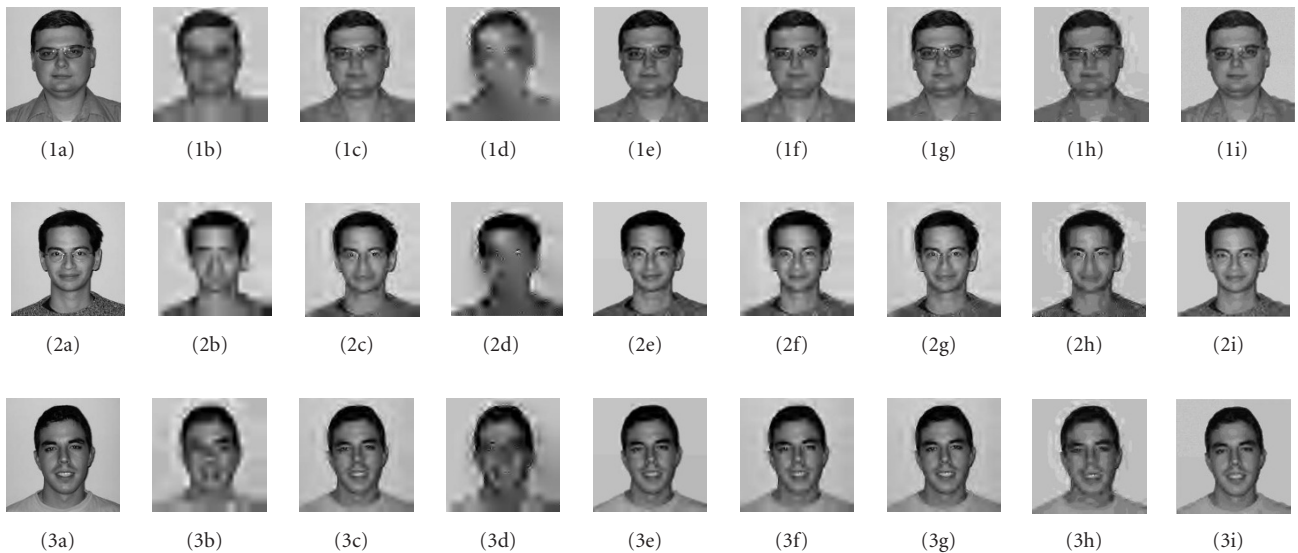

FIGURE 15: Experimental results. The first column: the original test images; the second column: ROI-JPEG2000 compression results for the rate 400 bytes; the third column: ROI-SPIHT compression results for the rate 400 bytes; the fourth column: EQ compression results for the rate 400 bytes; the fifth column: DSSC compression results for the rate 400 bytes; the sixth column: ROI-JPEG2000 compression results for the rate 700 bytes; the seventh column: ROI-SPIHT compression results for the rate 700 bytes; the eighth column: EQ compression results for the rate 700 bytes; and the ninth column: DSSC compression results for the rate 700 bytes.

The performance is evaluated in terms of the peak signal-to-noise ratio PSNR $= 10 \log_{10}(255^2/\|\mathbf{x} - \hat{\mathbf{x}}\|^2)$.

The obtained results allow to conclude about the proposed method advantages over the selected competitors for compression rates below 0.09 bpp in terms of both visual quality and PSNR. Performance loss at higher rate in our case in comparison with ROI-SPIHT and ROI-JPEG2000 is explained by the necessity of algorithm performance optimization for this rate regime that includes a modification of the unpredictable class definition.

6. CONCLUSIONS

In this paper, the problem of distributed source coding of a single source with side information was considered. It was shown that the compression system optimal performance for non-Gaussian sources can be achieved using the Berger-Flynn-Gray coding setup. A practical very-low-bit-rate compression algorithm based on this setup was proposed for coding of passport photo images. Experimental validation of this algorithm performed on a set of passport photos allows to

conclude its superiority over a number of existing encoding techniques at rates below 0.09 bpp in terms of both visual quality and PSNR. The realized performance loss of the developed algorithm at rates higher than 0.09 bpp is justified by the necessity of its parameters optimization for this rate range. This extension is a subject of our ongoing research.

DISCLAIMER

The information in this document reflects only the authors' 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.

ACKNOWLEDGMENTS

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa, USA, 1992.

[2] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[3] C. Chrysafis and A. Ortega, "Efficient context-based entropy coding lossy wavelet image compression," in *Proceedings of Data Compression Conference (DCC '97)*, pp. 241–250, Snowbird, Utah, USA, March 1997.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[5] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[6] Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 677–693, 1997.

[7] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Wavelet image coding via rate-distortion optimized adaptive classification," in *Proceedings of NJIT Symposium on Wavelet, Subband and Block Transforms in Communications*, Newark, NJ, USA, 1997.

[8] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proceedings of Data Compression Conference (DCC '97)*, pp. 221–230, Snowbird, Utah, USA, March 1997.

[9] A. Deever and S. S. Hemami, "What's your sign? efficient sign coding for embedded wavelet image coding," in *Proceedings of Data Compression Conference (DCC '00)*, pp. 273–282, Snowbird, Utah, USA, March 2000.

[10] X. Wu, "Compression of wavelet transform coefficients," in *The Transform and Data Compression Handbook*, K. R. Rao and P. C. Yip, Eds., chapter 8, pp. 347–378, CRC Press LLC, Boca Raton, Fla, USA, 2001.

[11] S. S. Pradhan and K. Ramchandran, "Enhancing analog image transmission systems using digital side information: a new wavelet-based image coding paradigm," in *Proceedings of Data Compression Conference (DCC '01)*, pp. 63–72, Snowbird, Utah, USA, March 2001.

[12] T. M. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, NY, USA, 1991.

[13] D. Slepian and J. K. Wolf, "Noiseless encoding of correlated information sourcea," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[14] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two sided state information," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1629–1638, 2002.

[15] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, International Convention Record*, vol. 7 (Part 4), pp. 142–163, 1959.

[16] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1971.

[17] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[18] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed., Springer, New York, NY, USA, 1977.

[19] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 773–787, 1987.

[20] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, no. 6, Part 2, pp. 2073–2084, 1996.

[21] 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*, vol. 6, no. 12, pp. 300–303, 1999.

[22] Y. Yoo, A. Ortega, and B. Yu, "Image subband coding using context based classification and adaptive quantization," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1702–1715, 1999.

[23] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, NY, USA, 1997.

[24] S. Voloshynovskiy, O. Koval, and T. Pun, "Wavelet-based image denoising using non-stationary stochastic geometrical image priors," in *Proceedings of IS&T/SPIE's 15th Annual Symposium, Electronic Imaging: Image and Video Communications and Processing 2003*, vol. 5022 of *Proceedings of SPIE*, pp. 675–687, Santa Clara, Calif, USA, January 2003.

[25] I. Kozintsev and K. Ramchandran, "Multiresolution joint source-channel coding using embedded constellations for power-constrained time-varying channels," in *Proceedings of*

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96), vol. 4, pp. 2343–2346, Atlanta, Ga, USA, May 1996.

[26] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[27] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.

[28] A. Hjorungnes, J. M. Lervik, and T. A. Ramstad, "Entropy coding of composite sources modeled by infinite Gaussian mixture distributions," in *Proceedings of IEEE Digital Signal Processing Workshop*, pp. 235–238, Loen, Norway, September 1996.

[29] K. R. Rao and P. C. Yip, Eds., *The Transform and Data Compression Handbook*, CRC Press, Boca Raton, Fla, USA, 2000.

[30] C. Christopoulos, J. Askelof, and M. Larsson, "Efficient methods for encoding regions of interest in the upcoming JPEG 2000 still image coding standard," *IEEE Signal Processing Letters*, vol. 7, no. 9, pp. 247–249, 2000.

[31] E. Atsumi and N. Farvardin, "Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees," in *Proceedings of IEEE International Conference on Image Processing. (ICIP '98)*, vol. 1, pp. 87–91, Chicago, Ill, USA, October 1998.

J. E. Vila-Forcén received the Telecommunications Engineer degree from the Carlos III University of Madrid, Spain, in 2001. In 2001–2002, he joined the Signal Theory and Communications Department of the Carlos III University, working in the development of the MPEG4 standard. Since 2002, he has been an Assistant Professor and a Ph.D. student at the Stochastic Image Processing Group, Computer Vision and Multimedia Lab, Department of Computer Science of the University of Geneva, Geneva, Switzerland. His current research interests are the information-theoretic aspects of digital data hiding, communications with side information, and stochastic image modeling for compression.

S. Voloshynovskiy received the Radio Engineer degree from Lviv Polytechnic Institute in 1993, and the Ph.D. degree in electrical engineering from State University *Lvivska Politechnika*, Lviv, Ukraine, in 1996. In 1998–1999, he has been with University of Illinois at Urbana-Champaign, USA, as a Visiting Scholar. Since 1999, he has been with 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 as well as nine patents. He has served as a consultant to private industry in the above areas.

O. Koval received his M.S. Degree in electrical engineering from the National University *Lvivska Politechnika*, Lviv, Ukraine, in 1996. In 1996–2001, he was with the Department of Synthesis, Processing, and Identification of Images, Institute of Physics and Mechanics (Lviv, Ukraine) as a researcher and Ph.D. student. He received his Ph.D. degree in electrical engineering from the National University *Lvivska Politechnika*, in 2002. Since 2002, he has been with Stochastic Image Processing Group, Computer Vision and Multimedia Lab, University of Geneva, from which he received his Ph.D. degree in stochastic image modeling in 2004, where he is currently a Postdoctoral Fellow. His research interests cover stochastic image modeling for different image processing applications, digital watermarking, information theory, and communications with side information.

T. Pun received his Ph.D. degree in image processing in 1982, at the Swiss Federal Institute of Technology in Lausanne (EPFL). He joined the University of Geneva, Switzerland, in 1986, where he is currently a Full Professor at the Computer Science Department and Head of the Computer Vision and Multimedia Lab. 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 to 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.