Lossy to Lossless Object-Based Coding of 3-D MRI Data

Gloria Menegaz, *Member, IEEE*, and Jean-Philippe Thiran, *Member, IEEE*

***Abstract*—We propose a fully three-dimensional (3-D) object-based coding system exploiting the diagnostic relevance of the different regions of the volumetric data for rate allocation. The data are first decorrelated via a 3-D discrete wavelet transform. The implementation via the lifting steps scheme allows to map integer-to-integer values, enabling lossless coding, and facilitates the definition of the object-based inverse transform. The coding process assigns disjoint segments of the bitstream to the different objects, which can be independently accessed and reconstructed at any up-to-lossless quality. Two fully 3-D coding strategies are considered: embedded zerotree coding (EZW-3D) and multidimensional layered zero coding (MLZC), both generalized for region of interest (ROI)-based processing. In order to avoid artifacts along region boundaries, some extra coefficients must be encoded for each object. This gives rise to an overheading of the bitstream with respect to the case where the volume is encoded as a whole. The amount of such extra information depends on both the filter length and the decomposition depth. The system is characterized on a set of head magnetic resonance images. Results show that MLZC and EZW-3D have competitive performances. In particular, the best MLZC mode outperforms the others state-of-the-art techniques on one of the datasets for which results are available in the literature.**

***Index Terms*—Multiresolution, objects, 3-D coding.**

I. Introduction

MEDICAL data are increasingly represented in digital form. Imaging techniques like magnetic resonance (MR), computerized tomography (CT) and positron emission tomography (PET) generate three-dimensional (3-D) data distributions. The representation in digital form enables the medical field to benefit from the know-how in signal and image processing, opening the way to new applications like computer-aided diagnosis, telemedicine, and, in general, new tools for improving health care. The focus here is on compression and coding. The limitations in transmission bandwidth and storage space on one side, and the growing size of medical image datasets on the other, push toward the design of *ad-hoc* tools for their efficient manipulation. The increasing demand has triggered a vast investigation on volumetric data compression, and a number of solution have been proposed so far exploiting the dependencies among data samples in the 3-D space. Among these, some are mainly concerned with video sequences [1]–[8], while others are focused on medical data [9]–[15]. The new trend in the field of data compression is founded on a redefinition of the role and the meaning of relevance in the information to be encoded. The focus is increasingly put on *semantics*. The different objects that are present in a scene are assigned a different *priority* in the encoding process, based on their importance in the framework of the considered application. The *a priori* knowledge about image content makes such approaches particularly suitable for medical images. In the same perspective, object-based algorithms are suitable for being combined with modeling techniques. The idea behind the so-called *model-based* approach to coding is to replace the real information with some synthetic representation of it in all the regions where the lossless constraint can be relaxed, assuming that the *information to be preserved* is the visual appearance. In this way, coding efficiency is improved by reducing the information to transmit. Some examples can be found in the field of video compression for multimedia (like video-telephony [5], [16] and surveillance [17]), medical imaging [18], as well as in emerging applications like stereoscopic imaging, used to obtain a 3-D perception of a scene [19].

Medical images usually consists of a region representing the part of the body under investigation (i.e., the heart in a CT or MRI chest scan, the brain in a head scan) on an often noisy background with no diagnostic interest. It seems thus very natural to process such data in a object-based framework: assign high priority to the semantically relevant object, to be represented with up-to-lossless quality, and lower priority to the background.

The agreement of the image processing community on object-based approaches is proved by the fact that the incoming standard for still image compression JPEG2000 [20] features region of interest (ROI)-based functionalities [21]–[23]. Nevertheless, 3-D data are out of the scope of the baseline JPEG2000. Even though some authors have addressed the task of object-based coding for medical images (see, for example, [24]–[26]), such an approach still deserves some investigation.

In this paper, we propose a fully 3-D wavelet-based coding system allowing random access to any object at the desired bit-rate. The distinguishing feature of the proposed system is the absence of artifacts along object borders, for any decoding rate. This is obtained by selecting, in each subband, the set of wavelet coefficients which are necessary for reconstructing the object (by inverse wavelet transform) *as if* the whole set of subband samples were available (e.g., the inverse transform were performed on the entire volume). A separable 3-D DWT is performed by the lifting steps scheme [27]. A nonlinear version mapping integers to integers is obtained by introducing

Manuscript received May 1, 2001; revised May 16, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Faouzi Kossentini.

G. Menegaz is with the Audio-Visual Communications Laboratory, School of Computer and Communication Sciences, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland (e-mail: Gloria.Menegaz@epfl.ch).

J.-P. Thiran is with the Signal Processing Institute, School of Engineering, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland.

Publisher Item Identifier 10.1109/TIP.2002.802525.

a rounding operation after each step [28], enabling lossless coding. Two coding strategies are considered: a 3-D version of the well-known embedded zerotree wavelet-based (EZW) algorithm [29] and the multidimensional layered zero coding (MLZC) technique [18]. These provide a fully embedded bitstream supporting a finely-graded up-to lossless range of bit-rates. Each object is encoded independently, to generate a segment of the global bitstream. In this paper, we do not address the issue of image segmentation, and we assume each object to be represented by a surface or region model (e.g., a binary mask) [30], [31], whose description is assumed to be available at both the encoder and decoder sides.

This paper is organized as follows. In Section II the integer wavelet transform via lifting steps is briefly revisited. Section III illustrates the object-based Inverse DWT (IDWT) and summarizes the procedure followed to select the relevant wavelet coefficients in the different subbands. The generalization of the EZW-3D and MLZC for object processing are presented in Section IV. Performances are analyzed in Sections V and VI derives conclusions.

II. Three-Dimensional Integer DWT via Lifting

The lifting steps scheme [27], [32] is particularly suitable for our purpose. First, it leads to an integer version of the discrete wavelet transform (DWT) in a very natural way [28]. This is of prime importance because it enables lossless coding. Then, the transform can be implemented *in-place*, minimizing the run-time memory requirements. This can have an important impact on the computational cost when large amounts of data (as volumes) must be handled. Finally, it asymptotically reduces the computational complexity by a factor four [33], [34]. Due to the rounding operations, the integer coefficients are different from the corresponding *true* wavelet coefficients. This compromises the approximation power of the wavelet basis and degrades compression performances in the medium-to-high quality range [35]. Since the number of rounding operations implied by each level of transformation, in each spatial dimension, increases with the number of lifting steps, filters corresponding to polyphase matrices which can be factorized in only two steps (i.e., *interpolating* filters) are most suitable. Furthermore, short filters minimize the amount of extra information to be encoded for each object to avoid artifacts along the boundaries. Accordingly, we adopt the 5×3 interpolating filter [36]. The separability of the transform allows an efficient implementation of the 3D-DWT by splitting it in three successive 1D-DWT, one for each spatial dimension.

III. Object-Based IDWT

Object-based processing concerns both transformation and coding. In the perspective of transformation it brings up a boundary problem. As discrete signals are nothing but sets of samples, it is straightforward to associate the idea of *object* to a *subset* of samples, usually sharing some common features. The problem of boundary conditions is greatly simplified when the DWT is implemented by the lifting steps scheme [27]. In this case, perfect reconstruction is ensured by construction, for any kind of signal extension at borders. Nevertheless, perfect reconstruction is not the only issue when dealing with a complete coding system. Our goal is to make object-based processing completely transparent with respect to the unconstrained general case where the signal is considered as a whole, in any working condition. Otherwise stated, we want the images decoded at a given quality to be *exactly* the same in the following conditions: 1) the signal has been encoded/decoded as a whole and 2) each object has been independently encoded and decoded at a given quality (e.g., quantization level). The perfect reconstruction condition is not enough to ensure the absence of artifacts—in terms of discontinuities at borders. Since quantized coefficients are approximations of the true values, any signal extension used to reconstruct two adjacent samples belonging to different objects (e.g., lying at the opposite sides of a boundary) would generate a discontinuity. To avoid this, the inverse transform must be performed *as if* the whole set of true coefficients were available. The use of the lifting scheme simplifies this task. The idea is to determine which samples are needed at the *input* of the synthesis chain to reconstruct a given sample at its *output*. The key of the proposed solution is to start at the finest resolution ($l = 1$) and select the set of wavelet coefficients which are needed (in each subband) to reconstruct the object in the signal domain (full resolution, $l = 0$). At this point, the problem has been solved for $l = 1$, or, equivalently, it has been projected to the next coarser level. Due to the recursivity of the IDWT, the approximation subband of level $l = 1$ becomes the reference (*critical*) set of samples that must be reconstructed without loss, and so on. By going through all the resolutions and successively iterating the procedure as described for $l = 1, \ldots, L$, the appropriate set of wavelet coefficients is selected. We call *generalized projection* of the object such set of coefficients.

Let GP be the corresponding operator, and let GP_η, be the set of samples obtained by applying GP in the direction $\eta = x, y, z$. The separability of the transform leads to the following composition rule:

$$GP_{zyx} = GP_z\{GP_y\{GP_x\{\cdot\}\}\}. \tag{1}$$

The set of wavelet coefficients to be encoded for each object are those belonging to its generalized projection. We refer to [18] and [37] for further details.

IV. Three-Dimensional Object-Based Coding

We restrict our analysis to the case of two disjoint regions. For simplicity, we will adopt the same terminology as in JPEG2000 and call ROI the object of interest and *background* the rest of the volume. The ROI is identified by a color code in a 3-D mask, that we assume to be available at both the encoder and decoder sides. The problem of shape representation and coding is not addressed in this work. However, it is worth mentioning that the proposed coding scheme was conceived for being integrated within a model-based coding system. The idea was to describe the object of interest by a predefined *ad-hoc* shape model. Such an approach was tailored on medical applications, due to the *a priori* knowledge about the image semantics. In a set of head MRI images for example, the object of interest is most probably the brain, and the rest of the image can be considered as the

background. If the "average" model for the generic patient were available to both the encoder and the decoder, only the deformation parameters needed to fit it to the current data would need to be transmitted. Different solutions can be envisaged in this respect. In [31] and [38] we proposed a parametric hybrid model for shape representation. Such a model is defined as a set of hybrid ellipsoids which are used to deform a reference shape both globally and locally. Such a parameterization allows to preserve the analytical representation during the fitting process, and has a number of properties making it suitable for the integration in a coding system. Among these are compactness, conciseness, availability of an inside–outside function and scale-invariance [30]. The final model can then be used either for deriving a mask (as it is the case in this work) or directly in analytical form via the corresponding inside/outside function.

Independent object coding has two major advantages. First, it is suitable for parallelization: different units can be devoted to the processing of the different objects simultaneously. Then, it is expected to improve coding efficiency when the objects correspond to statistically distinguishable sources. In what follows, the generalization of EZW-3D and MLZC coding systems for region-based processing are detailed.

A. Embedded Zerotree Wavelet Based Coding

The generalization of the classical EZW technique [29] for independent processing of 3-D objects is performed by applying the 3-D extension of the coding algorithm to the different objects, separately. The definition of the parent–children relationship is slightly modified with respect to the general case where the entire volume is encoded, to emphasize the semantics of the voxels as belonging to a particular region. Accordingly, the set of descendants of a wavelet coefficient $c(\mathbf{k}, l, j)$ at position $\mathbf{k}$ in subband (l, j) is identified by restricting the oct-tree to the domain of the generalized object projection $GP(l, j)$ in all the finer levels. More specifically, let T be a given oct-tree and let $T(l, j)$ identify the set of samples of the oct-tree in subband (l, j).

Definition 1: We define as *semantic oct-tree* the set of all subband samples ST

$$ST = \bigcup_{l,j} ST(l, j) \tag{2}$$

$$ST(l, j) = T(l, j) \cap GP(l, j). \tag{3}$$

Based on this, we derive a *semantically constrained* definition for a zerotree root.

Definition 2: A subband sample is a *zerotree root* if the all the coefficients which belong to the oct-tree originating in it are nonsignificant with respect to the current threshold.

Fig. 1 illustrates the semantically constrained oct-tree. Given a zerotree candidate point, as the significance of all the descendants lying outside the generalized projection is not relevant to the classification as zerotree root, we expect the number of successful candidate to increase with respect to the general case when *all* the descendants within T are required to be nonsignificant. This potentially augments coding efficiency. The inherent embedding resulting from the quantization strategy allows PSNR scalability for each object. Accordingly, each object can be reconstructed with increasing quality by progressively decoding the concerned portion of the bitstream.

B. Multidimensional Layered Zero Coding

The multidimensional layered zero coding (MLZC) is inspired by the *layered zero coding* (LZC) scheme proposed in [2] for multirate subband coding of video. It basically consists in successively applying a sequence of quantizers of decreasing bin-size to the subband structure, and encoding each corresponding significance map by context-adaptive arithmetic coding [39], [40]. LZC method takes advantage of the fact that the most probable symbol resulting from quantization is the *zero* symbol. High efficiency is achieved by splitting the entropy coding phase in two successive steps:

- *zero coding*: encodes a symbol representing the *significance* of the considered coefficients with respect to the current quantizer (i.e., $Q_n(x[i])$ being zero or nonzero);
- *magnitude refinement*: generates and encodes a symbol defining the value of each nonzero symbol.

Zero coding exploits some spatial or other dependencies among subband samples by providing such information to a *context adaptive* arithmetic coder [41]. The expected statistical relationships among coefficients are modeled by defining some *conditioning terms* which summarize the significance state of the samples belonging to a given neighborhood. The conditioning term $\chi_i[\mathbf{k}, l, j]$ for the symbol at position $\mathbf{k}$ in subband (l, j) is obtained by modeling both the *spatial* and the *interband* dependencies among wavelet coefficients via the terms $\chi^s[\mathbf{k}, l, j]$ and $\chi^f[\mathbf{k}, l, j]$

$$\chi[\mathbf{k}, l, j] = \chi^s[\mathbf{k}, l, j] + \chi^f[\mathbf{k}, l, j] \quad \forall l \neq L. \tag{4}$$

The spatial contribution $\chi^s[\mathbf{k}, l, j]$ is defined as a linear combination of the significance states of one or more samples in a neighborhood $N(\mathbf{k}, l, j)$

$$\chi^s[\mathbf{k}, l, j] = \sum_{p=0}^{P-1} w_p \sigma[\mathbf{k}', l, j] \quad \forall \mathbf{k}' \in N \tag{5}$$

where $p = p(\mathbf{k}')$. The weights $w_p = 2^p$ are such that each term of the summation contributes to the value of the pth bitplane of $\chi^s[\mathbf{k}, l, j]$, P is the bit depth of $\chi^s[\mathbf{k}, l, j]$, and σ is the distribution of the sequence of symbols $\sigma(\mathbf{k}, l, j)$ generated by quantizer Q_i. The fact that the weight depends on the spatial position $\mathbf{k}'$ within the neighborhood N allows to combine the contributions of different samples to the same bitplane of the conditioning term.

The interband term relates the significance state of the current sample to the one of its *parent* within the subband tree, according to

$$\chi^f[\mathbf{k}, l, j] = w_P \sigma[\mathbf{k}', l+1, j] \tag{6}$$

where $\mathbf{k}' = \lfloor \mathbf{k}/2 \rfloor$ is the position of the ancestor and $w_P = 2^P$ is the weight needed to define the MSB of the final context. The general rule does not apply to the coarsest subbands $l = L$, for which no parents can be identified. In this case, only the local-space contribution is considered. In the case of MLZC, 3-D local-scale conditioning terms are possible. Even though it

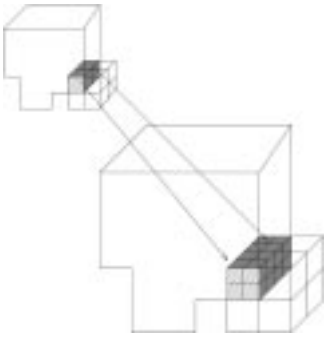

Fig. 1. Semantic oct-tree.

is reasonable to expect that 3-D contexts would increase coding efficiency due to the exploitation of the whole correlation among samples, some care must be devoted to the design of $\chi^s[\mathbf{k}]$ in order to keep the dimensionality of the conditioning space sufficiently small. The 3-D spatial support for the conditioning terms has been obtained by extending that of the best behaved two-dimensional (2-D) ones (in terms of compression performances) to the adjacent physical layers, as shown in Fig. 2. The set of local-scale bidimensional settings that have been considered is illustrated in Fig. 3. Very little modifications are needed to adapt the MLZC system to object-based processing. As for the EZW, each coefficient is encoded *if and only if* it belongs to the generalized projection of the considered object.

Equation (5) is generalized for this case by assuming that the significance state of any sample outside the generalized projection is zero

$$\chi^s[\mathbf{k}, l, j] = \sum_{p=0}^{P-1} w_p \tilde{\sigma}[\mathbf{k}', l, j] \tag{7}$$

$$\tilde{\sigma}[\mathbf{k}', l, j]) = \begin{cases} \sigma[\mathbf{k}', l, j], & \forall\, \mathbf{k}' \in GP(l, j) \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

V. Results and Discussion

In the framework of ROI-based coding, the weight assigned to a voxel depends on its semantics. This is assumed as the ground for the judicious allocation of available resources (e.g., bit-budget, bandwidth). The efficiency improvement is thus to be intended in the sense of *prioritization* of the information to be transmitted. The impact of the nonrelevant information must not be underestimated. This is particularly important when dealing with medical images, where the background often encloses the majority of the voxels. For a typical MRI dataset for instance, about the 90% of the voxels belong to the background. It is thus of prime importance to classify them *a priori* in order to assign higher priority to the ROI. Coding efficiency results from the tradeoff between the improvement due to the separation of sources with different statistics and the degradation due to the overhead implied by the border voxels.

The test set consists of an MRI head scan of 256 × 256 × 128 voxels. We assume that the object of interest is the brain, and consider the rest as background. The dataset is presented in Fig. 4. Particularly, Fig. 4(a) shows a representative image of the

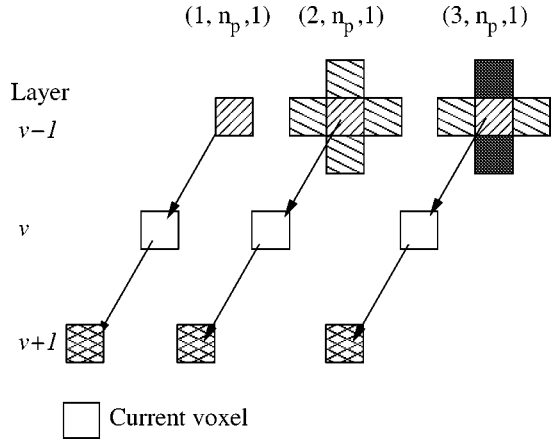

Fig. 2. Three-dimensional spatial support of the conditioning terms of a sample in subband image ν. The extensions to both the previous $(\nu - 1)$ and the next $(\nu + 1)$ subband images are represented. Squares with same pattern represent the positions of symbols whose significance states are combined in the definition of the corresponding $\chi^s[\mathbf{k}, l, j]$.

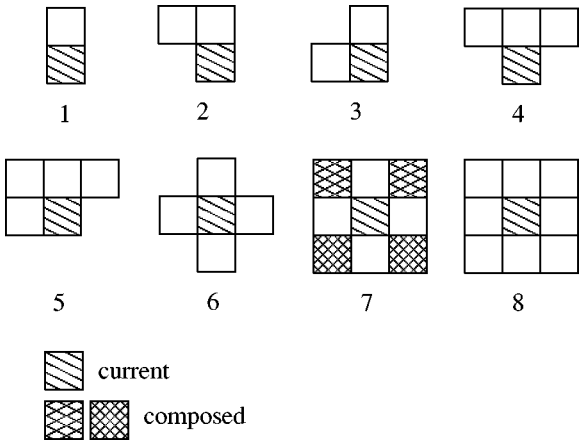

Fig. 3. Spatial supports for bidimensional conditioning terms. Squares with same pattern represent voxels whose significance states are combined in the definition of the corresponding $\chi_i^s[\mathbf{k}, l, j]$.

set, Fig. 4(b) represents the *mask*, or *atlas*, (which is used to select the object), Fig. 4(c) is the object of interest as segmented by the mask, and Fig. 4(d) is the background. In our work, the brain segmentation has been performed by a directional watershed, as described in [42]. Some coding results are also provided for the 8-bit head MR image volume obtained by the Mallinckrodt Institute of Radiology, Image Processing Laboratory, that we label as *MR-MRI*. It consists in a set of 58 images of size 256 × 256 representing the saggital view of a head which has been used as test set by other authors [10], [43], [44] for the evaluation of the compression performances of 3-D systems.

A. Compression Efficiency

The performances of the EZW-3D and MLZC coding systems has been analyzed by comparison with the 2-D counterparts—namely EZW-2D and MLZC-2D. The JPEG [45] and JPEG2000 [20] standards have also been considered. For 2-D encoding, the images have been processed independently. For all the wavelet-based methods—namely EZW, MLZC and JPEG2000—$L = 3$, 4 levels of decomposition and the 5 × 3 [36] filter have been chosen. All of the seven prediction

modalities of the lossless JPEG mode (JPEG-LS) were tested and the best one—corresponding to $k = 7$—was retained. Results show that the best behaved context for MLZC-2D is the (070) with inter-band conditioning, so it has been used to define the 3-D spatial conditioning terms. Accordingly, the $\chi^s[\mathbf{k}, l, j]$ have been constructed as illustrated in Fig. 2, with (070) being the spatial support in layer ν. Fig. 5 shows the lossless rates for the whole volume (WHOLE) when varying the shape of the support in the adjacent layers $(\nu - 1)$ and $(\nu + 1)$. The most performant are the (271) and (370) with inter-band conditioning. The first one has been retained for the evaluation of the object-based performances. It consists in the corner pixels of the first order neighborhood in the current layer (ν), a cross-shaped support in the previous layer $(\nu - 1)$, and the pixel in the same spatial position as the current one in the next $(\nu+1)$ layer (see Figs. 2 and 3). For the head MRI dataset, performances tend to improve when extending the generalized neighborhood used for conditional arithmetic coding, in both the 2-D and 3-D cases. The intuition for this is that when the size of the neighborhood used for conditioning increases, more information is available for entropy coding. Nevertheless, such a conclusion does not hold in general. The limit on the improvement in compression efficiency with the size of the support of χ is set by the volume size [18]. For relatively small volumes, the set of symbols is not sufficient for the probability tables of the entropy coder to adapt to the statistic of the source, raising a tradeoff between the expected improvement due to the increase of information and degradation due to an excessive growth of the conditioning space.

In order to compare the performances of MLZC and EZW-3D systems with other state-of-the-art techniques, the same set of experiments has been done on the MR-MRI dataset. In general, the trend is the same as for the MRI set, namely the best behaved contexts are (070) and (271) with interband conditioning for MLZC and MLZC-2D, respectively. Table I compares the average lossless rates of each of the considered 2-D algorithms to those provided by MLZC and EZW-3D, for both datasets. Among the 2-D algorithms, MLZC-2D with context (070) outperforms the others. JPEG2000 results in a lossless rate slightly lower than EZW-2D for MRI. All 2-D schemes provide a sensible improvement over JPEG-LS. For MRI, the lowest lossless rate corresponds to the EZW-3D scheme, which in this case slightly outperforms MLZC. Nevertheless, the MLZC method is faster and less computationally demanding than EZW-3D. The zerotree algorithm scans the whole tree of descendants of any potential zerotree-root for every quantization step. This makes it inefficient for 3-D data. For MLZC, the encoding time depends on the context, and increases with the size of the neighborhood. Efficiency can thus be improved by choosing spatial conditioning terms of small support.

For the MR-MRI set, some results are available in the literature. We refer here to those presented in [10]. The first one was obtained for $L = 3$ and using the integer version of the $(2 + 2, 2)$ filter (as defined in [36]) on the whole volume. The second was based on a two levels integer transform with the $(1 + 1, 1)$ filter on 16 slice coding units, and the compression efficiency data were averaged over the volume. The coding scheme—*3-D CB-EZW*—was a version of EZW-3D exploiting context modeling. The corresponding lossless rates are 2.285 and 2.195 bit/voxel, while the best MLZC mode results in 2.143 bit/voxel, slightly improving such a result. However,

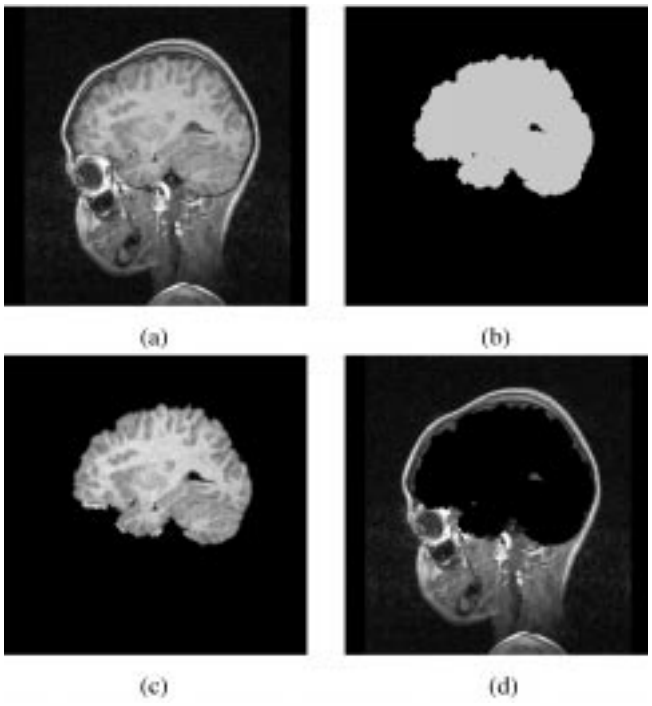

Fig. 4. Saggital view of the CT of a brain: (a) original image; (b) mask; (c) object of interest; and (d) background.

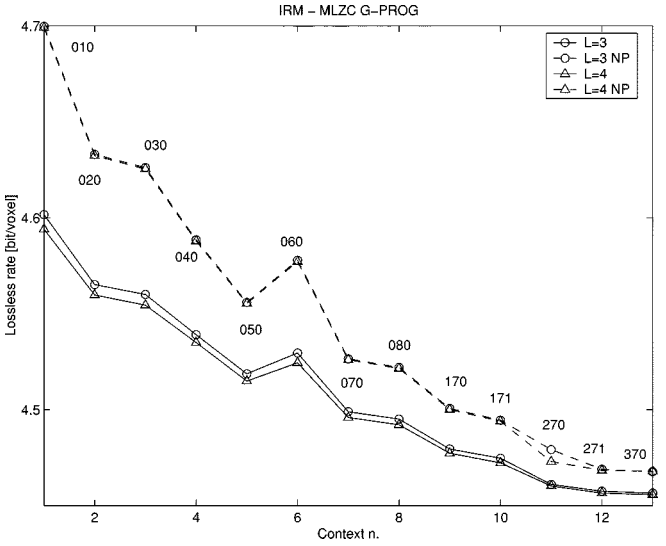

Fig. 5. Lossless rates for the whole volume (WHOLE) for $L = 3, 4$ and 5×3 filter. Continuous lines: global conditioning; dashed lines: spatial conditioning.

TABLE I
LOSSLESS RATES FOR MRI AND MR-MRI DATASETS. FOR THE 2-D ALGORITHMS, THE AVERAGE LOSSLESS RATE HAS BEEN RETAINED FOR EACH MODE. THE FILTER IS 5×3, $L = 4$, AND GLOBAL CONDITIONING IS USED IN THE MLZC MODE

	EZW-2D	MLZC-2D (070)	JPEG2000	JPEG-LS	EZW-3D	MLZC (271)
MRI	4.698	4.597	4.651	5.101	**4.456**	4.457
MR-MRI	2.878	2.848	2.954	3.437	2.271	**2.143**

a wider set of measurements is required for the comparative evaluation of the two competing systems.

B. Object-Based Performances

The results given in this section concern the MRI dataset, for which a segmentation mask was available. For convenience of notations, we define *object projection* $(OP(l, j))$ the support of the polyphase representation of the corresponding signal segment in subband (l, j). Then, we call *border voxels* $BV(l, j)$ the set of samples belonging to the generalized projection $GP(l, j)$ and *not* to $OP(l, j)$

$$BV = GP \setminus (GP \cap OP). \tag{9}$$

Fig. 6 gives an example of the generalized projection. The subband is LLH and $l = 1, 2, 3$. Border voxels are represented in gray, while white and black points represent the object and the background, respectively. The number of border voxels determines the overloading in the encoded information. This increases with the decomposition level until a saturation occurs for $l = l_{sat}$ [18], [38]. This means that for all $l \geq l_{sat}$, the relative increase of border voxels is due to the decrease in object voxels. This trend is illustrated in Fig. 7. The horizontal axis represent a composed subband index defined as $i = (l-1) \times 7 + j$, $j = 1, \ldots, 7$ and $l = 1, \ldots, L$ (we recall that $l = 1$ corresponds to the finest scale). The continuous line represents the percentage of the object voxels, while the other lines show the percentage of border voxels for the object of interest in the different subbands. In this example, the majority of the voxels in the coarsest subbands are of the border type. It is worth pointing out that even though the relative number of border voxels in the deep subbands is high, the global percentage of such voxels—namely the ratio between the total number of border voxels and the volume size—is indeed very small. For the MRI dataset, for example, it is about 2.6%.

In the proposed system, the object of interest and the background are encoded *independently*. Each of them generates a self-contained segment of the bitstream. This implies that the *border information* is encoded twice: as side information for *both* the object *and* the background. In this way, each of them can be accessed and reconstructed *as if* the whole set of wavelet coefficients were available, avoiding artifacts along the contours for any quantization of the decoded coefficients.

ROI-based EZW-2D has been assumed as the bench-mark for the object-based functionalities. Despite the availability of ROI-based functionalities, JPEG2000 was not suitable for the purpose. In JPEG2000, ROI-based coding is performed by the MAXSHIFT method [21]. Basically, the subband coefficients within the ROI mask are shifted up (or, equivalently, those outside the ROI are shifted down) so that the minimum value in the ROI is greater than the maximum value in the background. This splits the bitplanes respectively used for the ROI and the background in two disjoint sets. The rate allocation procedure assigns to each layer of each codeblock (in the different subbands) a coding priority which depends on both the semantics (through the MAXSHIFT method) and the gain in terms of rate/distortion ratio. This establishes the relative order of encoding of the ROI subband coefficients with respect to the background. With the implementation described in [46], for the head MRI dataset high priority is assigned to the background layers in the codeblocks, moving the focus of the encoder out of the ROI. The ROI and background codeblocks are mixed up, compromising ROI-based functionalities. This can be easily verified by decoding the portion of the bitstream indicated by the encoder as representing the ROI. The resulting image is composed of both the ROI and the background. A possible solution would be to design an *ad-hoc* rate allocation algorithm optimized for datasets having a background very easy to code, but this was out of the scope of our work. Instead, we independently compressed the ROI and the background with JPEG2000 and compared the respective bitrates to those provided by both our EZW-2D object-based system and ROI-based JPEG2000. Such working conditions emphasize the *implicit* ROI mask encoding by JPEG2000. Even though the mask does not need to be separately coded, its encoding is implied by the exhaustive scanning of the subbands. Results are given in Fig. 8. The global lossless rate in the different conditions are shown as a function of the image index. In particular, the dash-dot line represents ROI-based JPEG2000 and the continuous line is for EZW-2D with independent object coding (IO). The curve represents the sum of the lossless

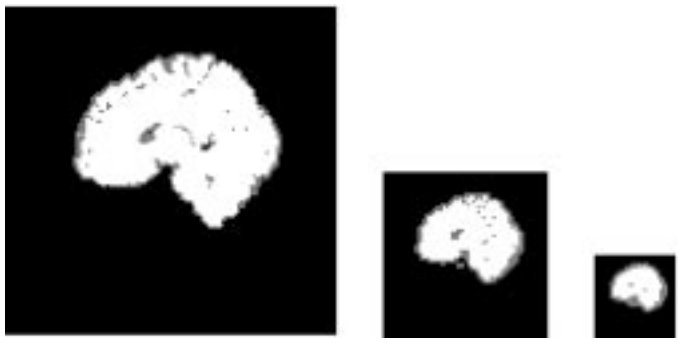

Fig. 6. Three-dimensional generalized projection of the brain in subband LLH. White voxels identify $OP(l, j)$, while gray voxels represent the *border extension*.

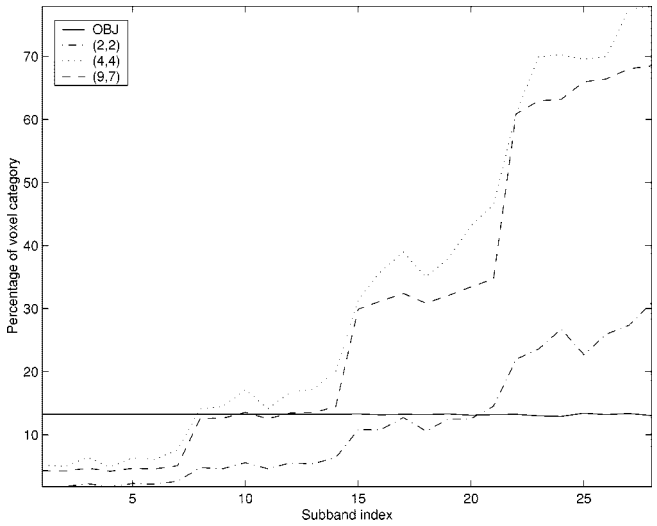

Fig. 7. Percentage of object and border voxels across the subbands. Subband (l, j) corresponds to the index $i = (l-1) \times 7 + j$. The relative amount of *border* voxels in particularly sensible in the higher decomposition levels.

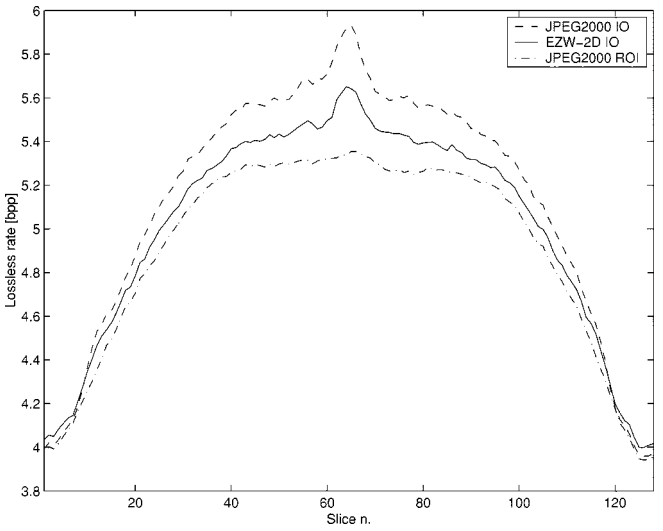

Fig. 8. Lossless rates as a function of the position of the 2-D images along the z axis. Continuous line: EZW-2D; dashed line: JPEG2000 IO (Independent Object); dash-dot line: JPEG2000 ROI.

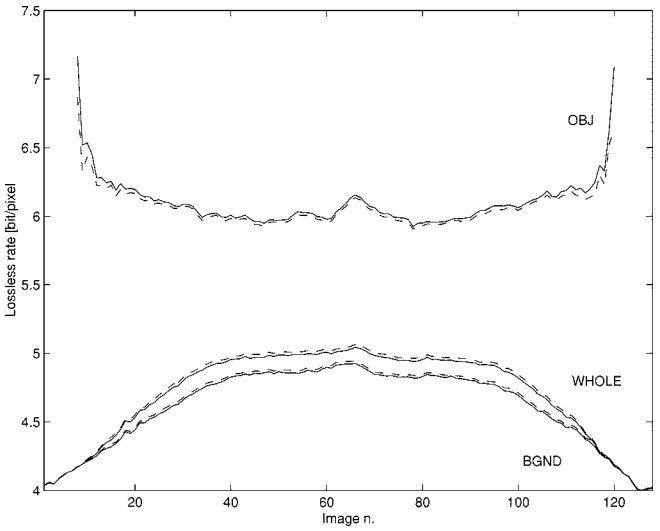

Fig. 9. Lossless rates for the EZW-2D algorithm as a function of the position of the 2-D images along the z axis, for the 5×3 filter. Dashed line: $L = 3$; continuous line $L = 4$.

rates concerning the ROI and the background. Due to the rate allocation policy, JPEG 2000 ROI outperforms EZW-2D in compression efficiency. The drawback is that, as previously mentioned, the codeblocks of the ROI and the background are interlaced in such a way that the ROI-based functionalities are not always achieved. The dashed line represents the total rate needed for independently encoding via JPEG2000 the ROI and the background by (JPEG 2000 IO). The gap between the corresponding curve and the one for EZW-2D IO emphasizes the performance degradation due to the implicit coding of the mask. Fig. 8 points out that the EZW-2D coding scheme represents a good compromise for the tradeoff between coding efficiency and random access to the objects. Fig. 9 shows the lossless rates for the ROI (OBJ), the background (BGND) and the entire image (WHOLE) for EZW-2D. The continuous and dashed lines correspond to $L = 3$ and $L = 4$, respectively. Here, the bitrates are calculated as the ratio between the size of the portion of the bitstream concerning the OBJ(BGND) and the size of the OBJ(BGND). While the curves for WHOLE and BGND are close to each other, the one for OBJ is outdistanced. The volume and the background enclose a large number of black samples, which are simple to compress. Conversely, the region of interest is entirely structured, and necessitates more bit/pixel to be encoded. The steep slope at both ends of the curve representing the object are due to the fact that the ROI takes only very few or no pixels, stretching the curve to infinity. This example points out the importance of the ROI-based approach. For this dataset, only the 19%—on average—of the bitstream corresponding to the entire volume is needed to represent the ROI. The random access to the objects allows fast access to the important information, with considerable improvement in compression *efficiency*. Fig. 10 shows the lossless rates for the object (OBJ) when varying the shape of the support of the conditioning term in layers $(\nu-1)$ and $(\nu+1)$ for MLZC. As was the case for WHOLE, the most performant is the (271) with inter-band conditioning. Results also show that the same conclusion holds for the background.

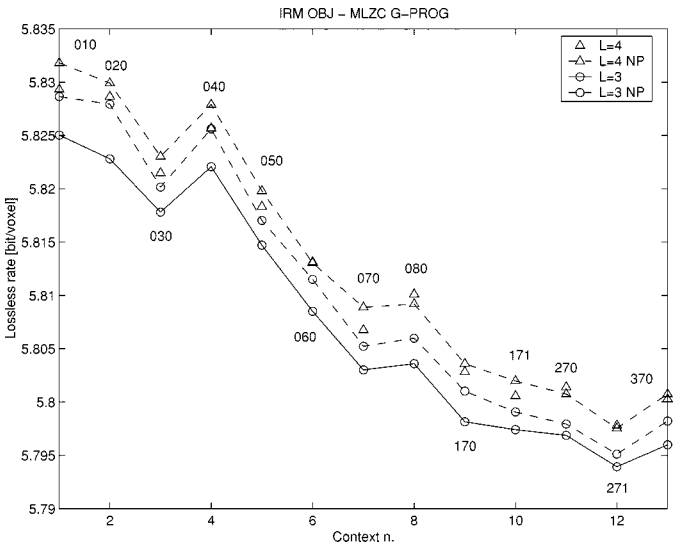

Fig. 10. Lossless rates for the object of interest (OBJ) for $L = 3, 4$ and 5×3 filter. The lossless rate is evaluated as the ratio between the size of the bitstream for OBJ and the size of OBJ. Continuous lines: global conditioning; dashed lines: spatial conditioning.

Table II quantifies the degradation in compression efficiency due to independent object coding. The first two columns (OBJ and BGND) show the lossless rates for the ROI and the background. The third column is the bitrate obtained when encoding the entire volume, and the last one shows the percentage increase of the lossless rates for independent encoding of the objects (OBJ + BGND) with respect to that corresponding to the entire volume (WHOLE). The increase of the lossless rate for independent object coding is measured by the difference between the required rate (OBJ + BGND) and the reference one (WHOLE). The differences between the compression ratios for the cases WHOLE and OBJ + BGND are due to two causes. First, the entropy coder performs differently in the two cases because of the different sources. Second, the total number of coefficients to be encoded is larger for

OBJ + BGND because of the generalized projections of both the object and background. The size of the bitstream increases by about 7% for $L = 4$ in case of separate object handling. According to Table II, the gain in compression efficiency due to the exploitation of the full correlation among data is about 4–5%. The improvement in compression efficiency provided by MLZC over JPEG2000 depends on the working mode. Taking the OBJ + BGND as reference, the corresponding rate reduction is about 2.2%, respectively 6.3%, for JPEG2000 ROI and JPEG2000 IO.

The prioritization of the information inherent to separate object processing leads to a significant improvement in coding efficiency when relaxing the lossless constraint in the background region. In this case, the BGND can be encoded/decoded at a lower resolution and combined with the object of interest—which has been encoded/decoded without loss—in the final composed image. Fig. 11 gives an example. Both the object and the background have been compressed by the MLZC scheme, with context (271) and using interband conditioning. The OBJ has been decoded at full quality (e.g., in lossless mode) while the BGND corresponds to a rate of 0.1 bit/voxels in Fig. 11(a) and 0.5 bit/voxels in Fig. 11(b). The PSNR values for images of Fig. 11(a) and (b) are of 27.76 and 33.53 dB, respectively. Reconstructed images respecting the lossless constraint in the ROI and preserving a good visual appearance in the background can thus be obtained by decoding only the 20% of the information that would be required for a lossless representation of the whole volume.

VI. Conclusions

We have presented two fully 3-D coding systems featuring object-based functionalities. The MLZC and EZW-3D coders have been generalized for ROI processing by restricting the information to be used for coding to the region taken by the object in every subband. Each object is encoded independently to generate a self-contained segment of the bitstream. The implementation of the DWT via the lifting steps scheme in the nonlinear integer version and the inherent embedding of the encoded information resulting from the coding systems allow the reconstruction of each object at a progressive up-to-lossless quality. Border artifacts are avoided by encoding some extra coefficients (for each object). The compression efficiency of the 3-D coding techniques has been evaluated by comparison with the 2-D counterparts—namely EZW-2D and MLZC-2D—as well as the JPEG and JPEG2000 standards. The rate saving provided by the 3-D coding techniques over JPEG2000 are in the range 2–6.3%, depending on the working mode taken as reference. The performances of MLZC and EZW-3D are competitive with those of the others state-of-the-art techniques. The ROI-based processing enables a *pseudo-lossless* regime where the object of interest can be encoded/decoded without loss, and combined with the background that can be represented at a lower quality. In this way, images respecting the lossless constraint in the ROI and preserving a good visual appearance in the background can be obtained at a significantly lower rate.

TABLE II
Lossless Rates (LR) for Head MRI. The Filter is 5×3, $L = 4$. Global Conditioning has Been Used in the MLZC Mode

LR [bpp]	OBJ	BGND	WHOLE	OBJ+BGND	Δ%
EZW-3D	0.9045	3.9012	4.4598	4.8057	+7.75
MLZC (271)	0.9188	3.8868	4.4566	4.8056	+7.83
EZW-2D	0.9327	4.0835	4.6977	5.0162	+6.78
JPEG2000 IO	1.0641	4.0656	4.6511	5.1297	+10.29
JPEG2000 ROI	-	-	4.6511	4.9099	+5.56

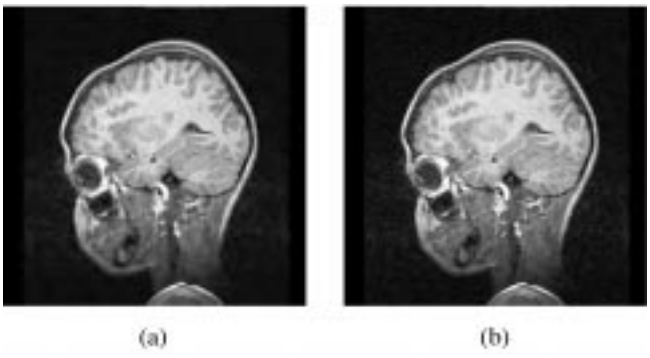

Fig. 11. *Pseudo-lossless* regime for a sample MRI image. The OBJ has been recovered without loss, while the BGND has been decoded at (a) 0.1 bpv and (b) 0.5 bpv. The corresponding PSNR values are of 27.76 and 33.53 dB, respectively.

Acknowledgment

The authors would like to thank Dr. J. Reichel for the fruitful discussions and for reviewing this work.

References

[1] B. Girod, P. Eisert, M. Magnor, E. Steinbach, and T. Wiegand, "3D image models and compression: Synthetic hybrid or natural fit?," in *Proc Int. Conf. Image Processing (ICIP)*, vol. 2, Oct. 1999, pp. 525–529.
[2] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, pp. 572–588, Sept. 1994.
[3] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1988, pp. 1110–1113.
[4] C. Podilchuk, N. S. Jayant, and N. Farvadin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, pp. 125–139, Feb. 1995.
[5] K. Aizawa and T. S. Huang, "Model-based image coding: Advanced video coding techniques for very low bit-rate applications," *Proc. IEEE*, vol. 83, pp. 259–271, Feb. 1995.
[6] E. Chang and A. Zakhor, "Scalable video coding using 3D subband velocity coding and multirate quantization," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 1993.
[7] J.-Y. Tham, S. Ranganath, and A. Kassim, "Highly scalable wavelet-based video codec for very low bit-rate environment," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 12–27, Jan. 1998.
[8] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D sphit)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 1374–1387, Dec. 2000.
[9] J. Wang and H. K. Huang, "Medical image compression by using three-dimensional wavelet transform," *IEEE Trans. Med. Imag.*, vol. 15, pp. 547–554, Aug. 1996.
[10] A. Bilgin, G. Zweig, and M. V. Marcellin, "Three-dimensional image compression with integer wavelet transform," *Appl. Opt.*, vol. 39, no. 11, pp. 1799–1814, 2000.
[11] L. Makris, I. Kamilatos, E. V. Kopsacheilis, and M. G. Strintzis, "Teleworks: A cscw application for remote medical diagnosis support and teleconsultation," *IEEE Trans. Inform. Technol. Biomed.*, vol. 2, pp. 62–73, June 2000.

[12] Z. Yu, S. Yu, G. Zhou, and Y. Mao, "Medical image compression based on wavelet transform," in *Proc. Int. Conf. Signal Processing (ICSP)*, vol. 1, 1998, pp. 811–814.
[13] W. Dajun and T. E. Chong, "Lossless medical image compression algorithm exploring three-dimensional space," in *Proc. Int. Conf. Signal Processing (ICSP)*, 2000, pp. 1062–1064.
[14] Y. Kim and W. A. Pearlman, "Lossless volumetric medical image compression," *Proc. SPIE*, vol. 3808, pp. 305–312, 1999.
[15] A. Manduca and A. Said, "Wavelet compression of medical images with set partitioning hierarchical trees," in *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Soc. (EMBS)*, vol. 3, 1997, pp. 1224–1225.
[16] C. S. Choi and T. Takebe, "Analysis and synthesis of facial image sequences in model-based image coding," *IEEE Trans. Video Technol.*, vol. 4, pp. 257–275, June 1994.
[17] I. Martins and L. Corte-Real, "A video-coder using 3D model-based background for video surveillance applications," in *Proc. Int. Conf. Image Processing (ICIP)*, 1998, pp. 919–923.
[18] G. Menegaz, "Model-based coding of multi-dimensional data with applications to medical imaging," Ph.D. dissertation, Signal Processing Lab., Swiss Federal Inst. Technol., Lausanne, May 2000.
[19] M. G. Strinzis, "Object-based coding of stereoscopic and 3D image sequences," *IEEE Signal Processing Mag.*, pp. 14–28, May 1999.
[20] ISO/IEC JTC 1/SC 29/WG1, "Information technology—JPEG2000 image coding system,", ISO/IEC Int. Std. 15444-1, ITU Recommend. T.800, 2000.
[21] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG 2000 still image coding system: An overview," *IEEE Trans. Consumer Electron.*, vol. 46, pp. 1103–1127, Nov. 2000.
[22] C. Christopoulos, J. Askelof, and M. Larsson, "Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard," *IEEE Signal Processing Lett.*, vol. 7, pp. 247–249, Sept. 2000.
[23] ——, "Efficient region of interest encoding techniques in the upcoming jpeg2000 still image compression standard," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 2, Vancouver, BC, Canada, Sept. 10–13, 2000, pp. 41–44.
[24] V. Vlahakis and R. I. Kitney, "Roi approach to wavelet-based, hybrid compression of mr images," in *Proc. 6th Int. Conf. Image Processing and its Applications*, vol. 2, 1997, pp. 833–837.
[25] G. Minami, Z. Xiong, A. Wang, P/A. Chou, and S. Mehrotra, "3D wavelet coding of video with arbitrary regions of support," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 2, 1999, pp. 1422–1425.
[26] A. Czhio, G. Cazuguel, B. Solaiman, and C. Roux, "Medical image compression using region-of-interest vector quantization," in *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBS)*, vol. 20-3, 1998, pp. 1277–1280.
[27] I. Daubechies and W. Sweldens, "Factoring wavelet transform into lifting steps," *J. Fourier Anal. Applicat.*, vol. 41, no. 3, pp. 247–269, 1998.
[28] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmon. Anal.*, vol. 5, no. 3, pp. 332–369, 1998.
[29] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
[30] V. Vaerman, "Multi-dimensional object modeling with application to medical image coding," Ph.D. dissertation, Signal Process. Lab., Swiss Federal Inst. Technol., Lausanne, 1999.
[31] V. Vaerman, G. Menegaz, and J. Ph. Thiran, "A parametric hybrid model used for multidimensional object representation," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 1, 1999, pp. 163–167.
[32] W. Sweldens and P. Shroeder, "Building your own wavelets at home," Tech. Rep., Univ. of South Carolina, Columbia, 1995.
[33] J. Reichel, "Complexity related aspects of image compression," Ph.D. dissertation, Swiss Federal Inst. Technol., Lausanne, Feb. 2001.
[34] J. Reichel, G. Menegaz, M. Nadenau, and M. Kunt, "Integer wavelet transform for embedded lossy to lossless image compression," *IEEE Trans. Image Processing*, vol. 10, pp. 383–392, Mar. 2001.
[35] J. Reichel, G. Menegaz, and M. Nadenau, "Integer wavelet decomposition for lossy image compression," *Proc. SPIE*, vol. 3808, pp. 257–268, 1999.
[36] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. Int. Conf. Image Processing (ICIP)*, 1997, pp. 596–599.
[37] G. Menegaz and J.-Ph. Thiran, "3D encoding/2D decoding of medical data," *IEEE Trans. Med. Imag.*, submitted for publication.
[38] G. Menegaz, V. Vaerman, and J.-Ph. Thiran, "Object-based coding of volumetric medical data," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 3, 1999, pp. 920–924.
[39] G. A. Triantafyllidis and M. G. Strinzis, "A context based adaptive arithmetic coding technique for lossless image compression," *IEEE Signal Processing Lett.*, vol. 6, pp. 168–170, July 1999.
[40] N. V. Boulgoris, D. Tzovaras, and M. G. Strinzis, "Lossless image compression based on optimal prediction, adaptive lifting, and conditional arithmetic coding," *IEEE Trans. Medical Imag.*, vol. 10, pp. 2–13, Jan. 2001.
[41] W. Pennebacker, J. Mitchell, G. Langdon, and R. Arps, "An overview of the basic principles of the q-coder adaptive binary arithmetic coder," *IBM J. Res. Manage.*, vol. 32, no. 3, pp. 717–726, Nov. 1988.
[42] J.-Ph. Thiran, V. Warscotte, and B. Macq, "A queue-based region growing algorithm for the accurate segmentation of multi-dimensional digital images," *Signal Process.*, vol. 60, no. 1, pp. 1–10, 1997.
[43] Z. Xiong, X. Wu, D. Y. Yun, and W. A. Pearlman, "Progressive coding of medical volumetric data using three-dimensional integer wavelet packet transform," in *Proc. IEEE 2nd Workshop Multimedia Signal Processing*, Piscataway, NJ, 1998, pp. 553–558.
[44] Y. Kim and W. A. Pearlman, "Stripe-based sphit lossy compression of volumetric medical images for low-memory usage and uniform reconstruction quality," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 4, 2000, pp. 2031–2034.
[45] ISO/IEC JTC 1/SC 29/WG1, *ISO Working Doc. ISO/IEC JTC1/SC29/WG1 N399 - WD14495*, July 1996.
[46] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Trans. Image Processing*, vol. 9, pp. 1158–1170, July 2000.

Gloria Menegaz (M'95) received the M.S. degree in electronic engineering from the Politecnico di Milano, Milan, Italy, in 1993, the M.S. degree in information technology from the Research and Education Center in Information Technology (CEFRIEL), Politecnico di Milano, in 1995, and the Ph.D. degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in July 2000.

She is currently a Postdoctoral Fellow with the Audio–Visual Communications Laboratory (LCAV) at the School for Computer and Communication Sciences (I&C) of EPFL. Her main research interests are in the field of multidimensional signal processing and coding, image modeling, visual perception, and computational and computer vision.

Jean-Philippe Thiran (M'96) was born in Namur, Belgium, in 1970. He received the electrical engineering degree and the Ph.D. degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1993 and 1997, respectively. His Ph.D. dissertation was related to 3-D medical image registration.

From 1993 to 1997, he was the Coordinator of the Image Analysis Group, Communications and Remote Sensing Laboratory, UCL, mainly working on medical image analysis. He is currently Senior Lecturer (Maître d'Enseignement et de Recherche) and the Leader of the Computer Vision and Image Analysis Group, Signal Processing Institute, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. His current scientific interests include image segmentation, *a priori* information in image analysis, PDEs in image analysis, multimodal signal processing, medical image analysis, including multimodal image registration, segmentation, computer-assisted surgery, and diffusion MRI. He is Co-Editor-in-Chief of the *Signal Processing Journal*.

Dr. Thiran was an active member of the Organizing Committee of the 3rd IEEE International Conference on Image Processing (ICIP'96). He was also member of the organizing committees of the 2002 IEEE International Conference on Multimedia and Expo (ICME'02) held in Lausanne, and of the 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP'02) held in Martigny, Switzerland.