3-D Scalable Medical Image Compression With Optimized Volume of Interest Coding

Victor Sanchez*, *Student Member, IEEE*, Rafeef Abugharbieh, *Senior Member, IEEE*, and Panos Nasiopoulos, *Member, IEEE*

***Abstract*—We present a novel 3-D scalable compression method for medical images with optimized volume of interest (VOI) coding. The method is presented within the framework of interactive telemedicine applications, where different remote clients may access the compressed 3-D medical imaging data stored on a central server and request the transmission of different VOIs from an initial lossy to a final lossless representation. The method employs the 3-D integer wavelet transform and a modified EBCOT with 3-D contexts to create a scalable bit-stream. Optimized VOI coding is attained by an optimization technique that reorders the output bit-stream after encoding, so that those bits belonging to a VOI are decoded at the highest quality possible at any bit-rate, while allowing for the decoding of background information with peripherally increasing quality around the VOI. The bit-stream reordering procedure is based on a weighting model that incorporates the position of the VOI and the mean energy of the wavelet coefficients. The background information with peripherally increasing quality around the VOI allows for placement of the VOI into the context of the 3-D image. Performance evaluations based on real 3-D medical imaging data showed that the proposed method achieves a higher reconstruction quality, in terms of the peak signal-to-noise ratio, than that achieved by 3D-JPEG2000 with VOI coding, when using the MAXSHIFT and general scaling-based methods.**

***Index Terms*—Embedded block coding with optimized truncation (EBCOT), medical image compression, scalable compression, volume of interest coding, 3D-JPEG2000.**

I. Introduction

VOLUMETRIC medical images, such as magnetic resonance imaging (MRI) and computed tomography (CT) sequences, are becoming a standard in healthcare systems and an integral part of a patient's medical record. Such 3-D data usually require a vast amount of resources for storage and transmission. For example, a single MRI sequence of a human brain, with slices of 512 $\times$ 512 pixels taken at 1 mm intervals, could easily result in over 100 MB of voxel data.

Manuscript received March 27, 2010; revised May 27, 2010; accepted May 31, 2010. Date of publication June 17, 2010; date of current version September 24, 2010. This work was supported in part by the Consejo Nacional de Ciencia y Tecnología (CONACYT), Mexico. *Asterisk indicates corresponding author.*

*V. Sanchez is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: victors@ece.ubc.ca).

R. Abugharbieh and P. Nasiopoulos are with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2010.2052628

With the wide pervasiveness of medical imaging applications in healthcare settings and the increased interest in telemedicine technologies, it has become essential to reduce both storage and transmission bandwidth requirements needed for archival and communication of related data, preferably by employing lossless compression methods. Furthermore, providing random access as well as resolution and quality scalability to the compressed data has become of great utility. Random access refers to the ability to decode any section of the compressed image without having to decode the entire data set. Resolution and quality scalability, on the other hand, refers to the ability to decode the compressed image at different resolution and quality levels, respectively. The latter is especially important in interactive telemedicine applications, where clients (e.g., radiologists or clinicians) with limited bandwidth connections using a remote image retrieval system may connect to a central server to access a specific region of a compressed 3-D data set, i.e., a volume of interest (VOI). The 3-D image is then transmitted progressively within the VOI from an initial lossy to a final lossless representation.

Several compression methods for 3-D medical images have been proposed in the literature, some of which provide resolution and quality scalability up to lossless reconstruction [1]–[6]. These methods are based on the discrete wavelet transform (DWT), whose inherent properties produce a bit-stream that is resolution-scalable. Quality scalability is then achieved by employing bit-plane based entropy coding algorithms that exploit the dependencies between the location and value of the wavelet coefficients, such as the embedded zerotree wavelet coding (EZW), the set partitioning in hierarchical trees (SPIHT), and the embedded block coding with optimized truncation (EBCOT) algorithms [7]–[9]. These compression methods, however, do not provide VOI decoding capabilities, i.e., the ability to reconstruct a VOI at higher quality than the rest of the 3-D image.

Recently, a number of medical image compression methods that support VOI coding have been proposed [10]–[13]. In [10], the authors presented a compression method based on JPEG2000 that supports prioritized VOI coding based on the anatomical tissues depicted in a 3-D medical image. The method employs a one-dimensional DWT (1D-DWT) along the slice direction with JPEG2000 encoding of the resulting transform slices. A priority is assigned to each group of coefficients describing the same spatial region at the same decomposition level according to its intensity level in the spatial domain. The method also allows for the definition of the relative importance of each sub-band in the coding process. In [11], the authors introduced a 3-D medical

image compression technique that supports VOI coding based on 3-D sub-band block hierarchical partitioning (3D-SBHP), a highly scalable wavelet transform based entropy coding algorithm. A number of parameters that affect the effectiveness of VOI coding are studied, including the size of the VOI, the number of decomposition levels, and the target bit-rate. The authors also discussed an approach to optimize VOI decoding by assigning a decoding priority to the different wavelet coefficient bit-planes. In [12], the authors summarized the features of various methods for VOI coding, including the maximum shift (MAXSHIFT) and general scaling-based (GSB) methods supported by the JPEG2000 standard [14]. These particular methods scale up the coefficients associated with a VOI above the background coefficients, by a scaling value. The MAXSHIFT method employs a maximum scaling value so that VOI coefficients are completely decoded before any background coefficients. The GSB method, on the other hand, employs a lower scaling value so that VOI and background coefficients are decoded simultaneously. In [13], the authors presented a VOI coding method for volumetric images based on the GSB method and the shape-adaptive wavelet transform. The method extends the capabilities of the GSB method to 3-D images with arbitrarily-shaped VOIs and allows for coding partial background information in conjunction with the VOI.

The main objective of this paper is to present a 3-D medical image compression method with 1) scalability properties, by quality and resolution up to lossless reconstruction and 2) optimized VOI coding at any bit-rate. We are particularly interested in interactive telemedicine applications, where different remote clients with limited bandwidth connections may request the transmission of different VOIs of the same compressed 3-D image stored on a central server. In this particular scenario, it is highly desirable to progressively transmit the different VOIs without the need to recode the entire 3-D image for each client's request. Furthermore, in order to improve the client's experience in visualizing the data remotely, it is also desirable to transmit the VOI at the highest quality possible at any bit-rate, in conjunction with a low quality version of the background, which is important in a contextual sense to help the client observe the position of the VOI within the original 3-D image [15]–[17]. In this work, the VOI is a cuboid defined in the spatial domain with possibly different values for the length, width and height.

The method presented in this paper employs a 3-D integer wavelet transform (3D-IWT) and a modified EBCOT with 3-D contexts to compress the 3-D medical imaging data into a layered bit-stream that is scalable by quality and resolution, up to lossless reconstruction. VOI coding capabilities are attained after compression by employing a bit-stream reordering procedure, which is based on a weighting model that incorporates the position of the VOI and the mean energy of the wavelet coefficients. In order to attain optimized VOI coding at any bit-rate, the proposed method also employs after compression, an optimization technique that maximizes the reconstruction quality of the VOI, while allowing for the decoding of background information with peripherally increasing quality around the VOI. The proposed method is different from the method in [10], where the VOI coding procedure is tissue-based, the relative importance of a specific sub-band is empirically assigned, and the entropy coding of wavelet coefficients is performed using 2-D contexts. Our proposed method is also different from the VOI coding method proposed in [11], where the background information is only decoded after the VOI is fully decoded, which prevents observing the position of the VOI within the original 3-D image. The proposed method also differs from the method in [13], where the scaling value of the VOI coefficients is empirically assigned and the shape information of the VOI must be encoded and transmitted, which may result in an increase in computational complexity as well as bit rate (due to shape encoding).

The novelties of the proposed method are threefold. First, our method employs the 3D-IWT in conjunction with a modified EBCOT with 3-D contexts to exploit redundancies between slices and improve the coding performance, while at the same time creating a layered bit-stream that is scalable by resolution and quality up to lossless reconstruction. Second, the bit-stream reordering procedure is performed after encoding, thus allowing for the decoding of any VOI without the need to recode the entire 3-D image. Third, the background information that is decoded in conjunction with the VOI allows for placement of the VOI into the context of the 3-D image and enhances the visualization of the data at any bit-rate.

We test the performance of the proposed method on various real 3-D medical images and compare it to 3D-JPEG2000 with VOI coding, using the MAXSHIFT and the GSB methods. Performance evaluation results show that, at various bit-rates, the proposed method achieves a higher reconstruction quality, in terms of the peak signal-to-noise ratio (PSNR), than those achieved by the MAXSHIFT and GSB methods.

The remainder of the paper is organized as follows. In Section II, we describe the proposed compression method. In Section III, we present and discuss the experimental results. We give the concluding remarks in Section IV.

II. Proposed Compression Method

The proposed compression method is depicted in Fig. 1. We first apply a 3D-IWT with dyadic decomposition to an input 3-D medical image. This transform maps integers to integers and allows for perfect invertibility with finite precision arithmetic, which is required for perfect reconstruction of a signal [18]. In this work, we employ the bi-orthogonal Le Gall 5/3 wavelet filter, implemented using the lifting step scheme [19]. Each level of decomposition, r, of the transform decomposes the 3-D image input into eight 3-D frequency sub-bands denoted as $\mathrm{LLL}r$, $\mathrm{LLH}r$, $\mathrm{LHL}r$ $\mathrm{LHH}r$, $\mathrm{HLL}r$, $\mathrm{HLH}r$, $\mathrm{HHL}r$, and $\mathrm{HHH}r$. The approximation low-pass sub-band, LLL, is a coarser version of the original 3-D image, whereas the other sub-bands represent the details of the image. The decomposition is iterated on the approximation low-pass sub-band.

We then group the wavelet coefficients into 3-D groups and compute the mean energy of each group. We encode each group of coefficients independently using a modified EBCOT with 3-D contexts to create a separate scalable layered bit-stream for each group. The coordinates of the VOI in the spatial domain, in conjunction with the information about the mean energy of the grouped coefficients, are then used in a weight assignment model to compute a weight for each group of coded wavelet coefficients. These weights are used to reorder the output bit-

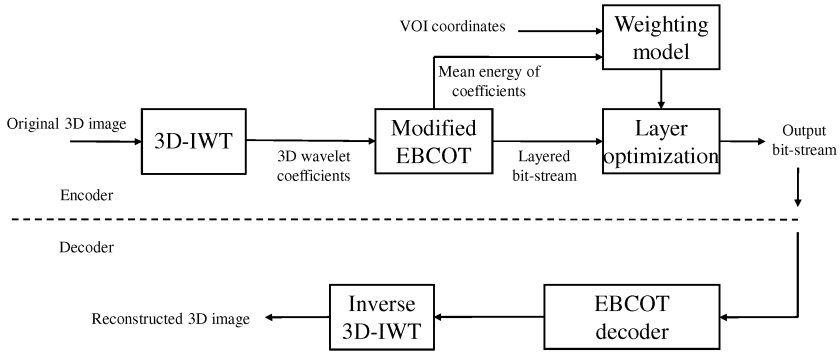

Fig. 1. Block diagram of the proposed scalable lossless compression method. 3D-IWT: three-dimensional integer wavelet transform. EBCOT: embedded block coding with optimized truncation.

stream and create an optimized scalable layered bit-stream with VOI decoding capabilities and gradual increase in peripheral quality around the VOI. At the decoder side, the wavelet coefficients are obtained by applying the EBCOT decoder. Finally, an inverse 3D-IWT is applied to obtain the reconstructed 3-D image. The decoder can also truncate the received bit-stream to obtain a 3-D image at any bit-rate.

It is important to mention that the proposed method attains VOI decoding capabilities after the 3-D medical imaging data is coded. This is particularly advantageous in interactive telemedicine applications, where different clients may request different VOIs of the same compressed 3-D image stored on a central server. The server may then transmit different versions of the same compressed bit-stream by simply performing the bit-stream reordering procedure for each requested VOI, thus saving time in recoding the entire 3-D image for each client's request. Moreover, if a client requests a different VOI while transmission of a compressed bit-stream is taking place, the server only needs to update the coefficient weights according to the newly requested VOI and reorder the untransmitted portion of bit-stream, which also saves time in recoding and retransmitting the entire 3-D image. Note that the bit-stream reordering procedure can take place before transmission since the decoder is capable of decoding any bit-stream regardless of the order it is transmitted (due to the fact that code-cubes are encoded independently). Alternatively, the bit-stream reordering procedure may also be performed at the client side once the image has been fully transmitted. In this particular scenario, the main advantage of the proposed method lies on saving time in recoding the entire 3-D image for different VOIs.

There are three key techniques in the proposed compression method. The first is the modified EBCOT. The second is the weight assignment model. The last is the creation of an optimized scalable layered bit-stream. We will discuss them in the next subsections.

A. *Modified EBCOT*

EBCOT is an entropy coding algorithm for 2-D wavelet-transformed images, which generates a bit-stream that is both resolution and quality scalable [9]. EBCOT partitions each sub-band in small group of samples, called code-blocks, and generates a separate scalable layered bit-stream for each code-block. The algorithm is based on context adaptive binary arithmetic coding and bit-plane coding, and employs four coding passes to code new information for a single sample c in the current bit-plane p. The coding passes are 1) zero coding (ZC), 2) run-length coding (*RLC*), 3) sign coding (SC), and 4) magnitude refinement (MR). A combination of the ZC and RLC passes encodes whether or not sample c becomes significant in the current bit-plane p. A sample c is said to be significant in the current bit-plane p if and only if $|c| \geq 2^p$. The significance of sample c is coded using ten different context models (nine for the ZC pass and one for the RLC pass), which exploit the correlation between the significance of sample c and that of its immediate neighbors. If sample c becomes significant in the current bit-plane p, the SC pass encodes the sign information of sample c using five different context models. The MR pass uses three different context models to encode the value of sample c only if it is already significant in the current bit-plane p.

We may employ EBCOT to code the wavelet coefficients on a slice-by-slice basis. However, in our compression method, the input samples to the entropy coding algorithm are 3D-IWT wavelet coefficients rather than 2D-IWT wavelet coefficients. Therefore, coding 3D-IWT wavelet coefficients on a slice-by-slice basis makes EBCOT less efficient since the correlation between coefficients is not exploited in three dimensions. Consequently, a modified EBCOT algorithm is needed to overcome this problem, which we solve by partitioning each 3-D sub-band into small 3-D groups of samples (i.e., wavelet coefficients), which we call code-cubes, and coding each code-cube independently by using a modified EBCOT with 3-D contexts.

In this work, code-cubes are comprised of $a \times a \times a$ samples and describe a specific region of the 3-D image at a specific decomposition level. We employ a pyramid approach to define the size of code-cubes across the different decomposition levels. In this approach, a code-cube of size $a \times a \times a$ samples and position $\{x, y, z\}$ at decomposition level r is related to a code-cube of size $a/2 \times a/2 \times a/2$ samples and position $\{x, y, z\}$ at decomposition level $r + 1$, where $r = 1$ is the first decomposition level. Fig. 2 shows the 3D-IWT sub-bands of a 3-D image after two levels of decomposition in all three dimensions with a single code-cube in sub-bands HHH_2 and HHH_1. It can be seen that by employing a pyramid approach to define the size of code-cubes, it is possible to access any region of the 3-D image at any resolution, which is essential for VOI coding. In this work, we limit the code-cube dimension, a, to be a power of 2, with $a \geq 2^3$.

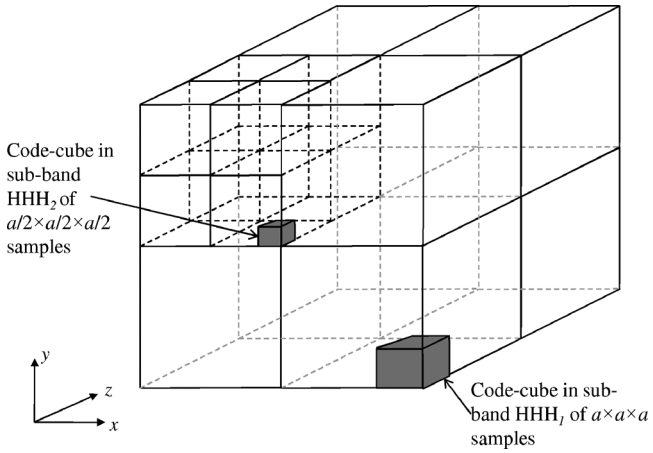

Fig. 2. 3D-IWT sub-bands of a 3-D image after two levels of decomposition in all three dimensions with a single code-cube in sub-bands HHH_1 and HHH_2.

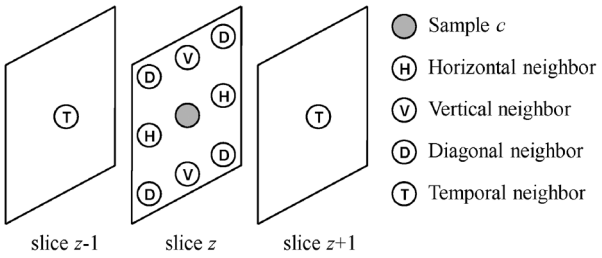

Fig. 3. The immediate horizontal, vertical, diagonal and temporal neighbors of sample c located in slices z, slices $z-1$ and $z+1$.

We code each code-cube independently using a modified EBCOT with 3-D contexts that exploit inter-slice correlations. Coding wavelet coefficients by extending 2-D context modeling to 3-D has been extensively used to improve coding efficiency [1], [2], [20], [21]. Here, we propose a 3-D context model, based on the four coding passes previously discussed, that incorporates information from the immediate horizontal, vertical, diagonal and temporal neighbors of sample c located in slices z, $z-1$ and $z+1$, as illustrated in Fig. 3.

During the ZC pass, we code whether or not sample c becomes significant in the current bit-plane p. As explained by Taubman in [9], the significance of sample c is highly dependent upon the value of its immediate horizontal, vertical and diagonal neighbors. Here, in order to exploit interslice correlations, we also employ the information about the significance of the immediate temporal neighbors to code the significance of sample c. Let h denote the number of significant horizontal neighbors, with $0 \leq h \leq 2$. Let v denote the number of significant vertical neighbors, with $0 \leq v \leq 2$. Similarly, let d denote the number of significant diagonal neighbors, with $0 \leq d \leq 4$; and let t denote the number of significant temporal neighbors, with $0 \leq t \leq 2$. The proposed 3-D context assignment for the ZC pass is summarized in Table I. Note that this 3-D context assignment emphasizes on the neighbors which are expected to present the strongest correlation in a particular sub-band. For example, we expect the strongest correlation amongst horizontally adjacent samples in sub-bands LLL, LLH, LHL, and LHH; therefore, the proposed 3-D context assignment emphasizes on horizontal neighbors for these sub-bands.

For the SC pass, we expect that the sign information of sample c exhibit some correlation with that of its temporal neighbors, in addition to the correlation exhibited with its vertical and horizontal neighbors, as explained in [9]. Therefore, in this pass, we employ the sign and significance information of the temporal, vertical and horizontal neighbors to code the sign information of sample c. Let $X(c)$ denote the sign bit of sample c, so that $X(c) = 1$ if $c \geq 0$; otherwise $X(c) = 0$. Let h_s denote the sign information of the horizontal neighbors, with $h_s = 0$ if both horizontal neighbors are insignificant or both are significant with different sign, $h_s = 1$ if at least one horizontal neighbor is positive, and $h_s = -1$ if at least one horizontal neighbor is negative. Let us define v_s and t_s in a similar fashion for the sign information of the vertical and temporal neighbors, respectively. The proposed 3-D context assignment for the SC pass is summarized in Table II. Note that this 3-D context assignment exploits the fact that the distribution of $X(c)$ given any particular neighborhood should be identical to the distribution of $-X(c)$, given the dual neighborhood with the signs of all neighbors reversed. The binary valued symbol that is coded with respect to the corresponding context is $X(c) \oplus \hat{X}(c)$, where $\hat{X}(c)$ is an auxiliary variable that indicates the sign prediction under a given context.

For the MR pass, we also expect that the magnitude of sample c exhibit some correlation with the magnitude of its immediate temporal neighbors. We thus employ the significance information of the immediate temporal neighbors, in addition to the significance information of the immediate horizontal and vertical neighbors, to code the magnitude of sample c. Let S denote the total number of significant temporal, horizontal and vertical neighbors of sample c, with $0 \leq S \leq 6$. Let σ be a variable that transitions from 0 to 1 after sample c is found to be significant for the for first time; i.e., after the MR pass is first applied to sample c. The proposed 3-D context assignment for the MR pass is summarized in Table III.

Note that for each coding pass, the coding engine maintains a look-up table in order to identify the probability model to be used by the adaptive arithmetic coder under each context.

B. Weight Assignment Model

The purpose of the weight assignment model is to enable the encoder to reorder the output bit-stream, so that the code-cubes that constitute the VOI are included earlier while allowing for gradual increase in peripheral quality around the VOI, under the constraint that the VOI is the main focal point. Techniques that allow gradual increase in peripheral quality around a focal point have been extensively used to improve image and video coding algorithms [22]–[25]. In the proposed compression method, we apply this technique to decode contextual background information with peripherally increasing quality around the VOI, which in turn enhances the visualization of the data at any bit-rate. We achieve this by considering two main factors: 1) the proximity of a code-cube to the VOI and 2) the mean energy of a code-cube. The desired weight assignment for code-cube Cc_i is a function of the form

$$w_{Cc_i}\left(\mathrm{P}_{Cc_i}, \mathrm{B}_{Cc_i}, \rho_{Cc_i}\right) \in [0,1] \tag{1}$$

TABLE I
PROPOSED 3D CONTEXT ASSIGNMENT FOR THE ZERO CODING (ZC) PASS OF SAMPLE c

Sub-bands LLL, LLH, LHL, LHH					Sub-bands HLL, HLH					Sub-bands HHH, HHL			
h	v	d	t	Context	h	v	d	t	Context	d	$h+v$	t	Context
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	≥1	0	1	0	0	≥1	0	1	0	1	0	1
0	≥1	X	0	2	≥1	0	X	0	2	0	≥2	0	2
0	X	X	1	3	X	0	X	1	3	0	X	1	3
0	X	X	2	4	X	0	X	2	4	0	X	2	4
1	0	0	0	5	0	1	0	0	5	1	0	0	5
1	0	≥1	0	6	0	1	≥1	0	6	1	≥1	0	6
1	≥1	X	0	7	≥1	1	X	0	7	1	X	≥1	7
1	X	X	≥1	8	X	1	X	≥1	8	2	0	0	8
2	X	X	0	9	X	2	X	0	9	2	≥1	0	9
2	X	X	≥1	10	X	2	X	≥1	10	2	X	≥1	10

X: don't care

TABLE II
PROPOSED 3-D CONTEXT ASSIGNMENT FOR THE SIGN CODING (SC) PASS OF SAMPLE c

h_s	v_s	t_s	$\hat{X}(c)$	Context
1	1	1	1	0
1	1	{-1,0}	1	1
1	0	1	1	2
1	0	{-1,0}	1	3
1	-1	1	1	4
1	-1	{-1,0}	0	5
0	1	1	1	6
0	1	{-1,0}	1	7
0	0	X	1	8
0	-1	{-1,0}	0	7
0	-1	1	1	6
-1	1	{-1,0}	0	5
-1	1	1	0	4
-1	0	{-1,0}	0	3
-1	0	1	0	2
-1	-1	{-1,0}	0	1
-1	-1	1	0	0

X: do not care

TABLE III
PROPOSED 3-D CONTEXT ASSIGNMENT FOR THE MAGNITUDE REFINEMENT (MR) PASS OF SAMPLE c

σ	S	Context
0	0	0
0	{1,2}	1
0	{3,4}	2
0	{5,6}	3
1	X	4

X: don't care

where P_{Cc_i} is a value in the range (0,1] that depends on the proximity between the center of code-cube Cc_i and the center of the VOI, B_{Cc_i} is a value in the range [0,1] that depends on the mean energy of code-cube Cc_i, and ρ_{Cc_i} is a value in the range [0,1] that depends on the proportion of wavelet coefficients of code-cube Cc_i that contributes to the VOI. We define function $w_{Cc_i}(\mathrm{P}_{Cc_i}, \mathrm{B}_{Cc_i}, \rho_{Cc_i})$ by studying an important feature of 3-D medical images in the spatial and wavelet domains.

In the spatial domain, 3-D medical images usually depict the anatomy of one or more organs (or structures) over an empty background. Furthermore, the areas comprised by the depicted structures typically contain most of the energy of the 3-D image. In the wavelet domain, the original structures depicted in the 3-D medical image are preserved as edge information within each sub-band. Following the grouping of wavelet coefficients into code-cubes, those code-cubes comprising the edge information thus tend to contain most of the sub-band energy.

Based on the above observations, we employ the information about the coordinates of the VOI and the mean energy of a code-cube to determine if a code-cube constitutes the VOI, the non-empty background, i.e., a structure depicted in the 3-D medical image that is not part of the VOI, or the empty background. The main objective is to assign the largest weight $(w_{Cc_i} = 1)$ to those code-cubes within the VOI, a smaller weight to those code-cubes within the non-empty background, and the smallest weight to those code-cubes within the empty background.

We determine which code-cubes constitute the VOI by using the VOI coordinate information and the location of the code-cubes in the spatial domain. The latter is calculated by tracing back the wavelet coefficients to a set of voxels using the footprint of the wavelet kernel used to transform the data. We employ $\rho_{Cc_i} \in [0, 1]$ as a measure of the proportion of wavelet coefficients of code-cube Cc_i that contribute to the VOI, with $\rho_{Cc_i} = 0$ for those code-cubes outside the VOI, $\rho_{Cc_i} = 1$ for those code-cubes that fully contribute to the VOI, and $0 < \rho_{Cc_i} < 1$ for those code-cubes with some contribution to the VOI.

In order to determine which code-cubes constitute the empty background, we use the information about their mean energy, which for code-cube Cc_i is calculated as follows:

$$\hat{\varepsilon}_{Cc_i} = \frac{1}{K} \sum_{k=1}^{K} c_k^2 \tag{2}$$

where c_k is the kth sample of Cc_i, and K is the total number of samples in Cc_i.

We expect the value of $\hat{\varepsilon}_{Cc_i}$ to be zero for those code-cubes within the empty background. However, this may not always be true due to the discrete size of code-cubes and the smearing effects of the wavelet filter, which may result in some code-cubes comprising both structure and empty background information. The simplest possible method to determine if a code-cube is part of the empty background is to use a thresholding approach, where code-cube Cc_i is considered to constitute the

empty background if the mean energy $\hat{\varepsilon}_{Cc_i}$ is below a defined value. Generally, the use of continuous functions, where no hard decision is required to determine if a code-cube is part of the empty background, leads to better results. We, thus, use the following simple continuous, monotonically decreasing function to determine if code-cube Cc_i in sub-band s is part of the empty background

$$\mathrm{B}_{Cc_i} = 1 - \frac{\hat{\varepsilon}_{Cc_i}}{\max\limits_{Cc_i \in s}\{\hat{\varepsilon}_{Cc_i}\}} \in [0,1] \tag{3}$$

where $\max\limits_{Cc_i \in s}\{\hat{\varepsilon}_{Cc_i}\}$ is the maximum mean energy $\hat{\varepsilon}_{Cc_i}$ in sub-band s. A value of B_{Cc_i} close to one means a high probability that code-cube Cc_i is part of the empty background, corresponding to a low mean energy content, whereas a value of B_{Cc_i} close to zero means a low probability that code-cube Cc_i is part of the empty background, corresponding to a high mean energy content. All values B_{Cc_i} are calculated during the encoding process and are stored as header information.

We now define function $w_{Cc_i}(\mathrm{P}_{Cc_i}, \mathrm{B}_{Cc_i}, \rho_{Cc_i})$ to assign weight w_{Cc_i} to code-cube Cc_i. We also employ a continuous, monotonically decreasing function with a range [0,1] as follows:

$$w_{Cc_i}(\mathrm{P}_{Cc_i}, \mathrm{B}_{Cc_i}, \rho_{Cc_i}) = \rho_{Cc_i} + (1 - \rho_{Cc_i})\, e^{-\left(\frac{\mathrm{B}_{Cc_i}}{\mathrm{P}_{Cc_i}}\right)^2} \tag{4}$$

where B_{Cc_i} is as defined in (3), ρ_{Cc_i} is the proportion of wavelet coefficients of code-cube Cc_i that contributes to the VOI, and P_{Cc_i} is the probability that code-cube Cc_i is located peripherally close to the VOI and is calculated by

$$\mathrm{P}_{Cc_i} = 1 - \frac{d_{Cc_i}}{D_{\max}} \in (0,1] \tag{5}$$

where d_{Cc_i} is the radial distance between the center of the VOI and the center of the region represented by code-cube Cc_i in the spatial domain, and $D_{\max}$ is the maximum radial distance in the spatial domain between two samples of the 3-D image

$$D_{\max} = \sqrt{x^2 + y^2 + z^2} \tag{6}$$

where $\{x, y, z\}$ denotes the size of the 3-D image in the spatial domain. Note that P_{Cc_i} may only take values in the range (0, 1], since $0 < d_{Cc_i} < D_{\max}$ for code-cubes outside the VOI with a dimension $a \geq 2^3$. A value of P_{Cc_i} close to one means a high probability that code-cube Cc_i is located peripherally close to the VOI, whereas a value of P_{Cc_i} close to zero means a low probability that code-cube Cc_i is located peripherally close to the VOI.

We employ the function in (4) as it is one of the simplest functions to provide the desired gradual decrease in weight w_{Cc_i} that quickly falls off as the probability that code-cube Cc_i is part of the empty background increases (i.e., as the value of B_{Cc_i} increases) and the probability that is located peripherally close to the VOI decreases (i.e., as the value of P_{Cc_i} decreases), but still leads to weights equal to one for those code-cubes within the VOI (i.e., with a value of ρ_{Cc_i} equal to one). Function $w_{Cc_i}(\mathrm{P}_{Cc_i}, \mathrm{B}_{Cc_i}, \rho_{Cc_i})$ has a simple probabilistic interpretation. The value of weight w_{Cc_i} corresponds to the probability of code-cube Cc_i being within the VOI and thus containing structure information. For code-cubes that fully contribute to the VOI, this probability is equal to one since the underlying assumption is that all code-cubes within the VOI contain structure information. For code-cubes outside the VOI, this probability follows a Gaussian distribution with a peak value of one centered at $\mathrm{B}_{Cc_i} = 0$ and a decaying rate controlled by P_{Cc_i}. For code-cubes that partially contribute to the VOI, this probability depends on their proportion of wavelet coefficients that contribute to the VOI, and a Gaussian distribution controlled by B_{Cc_i} and P_{Cc_i}. Fig. 4 shows the plot of (4) for code-cubes outside the VOI for various values of P_{Cc_i}. It can be seen that the value of w_{Cc_i} slowly decays peripherally around the center of the VOI for small values of B_{Cc_i} and large values of P_{Cc_i}, whereas it quickly approaches zero for large values of B_{Cc_i} and small values of P_{Cc_i}.

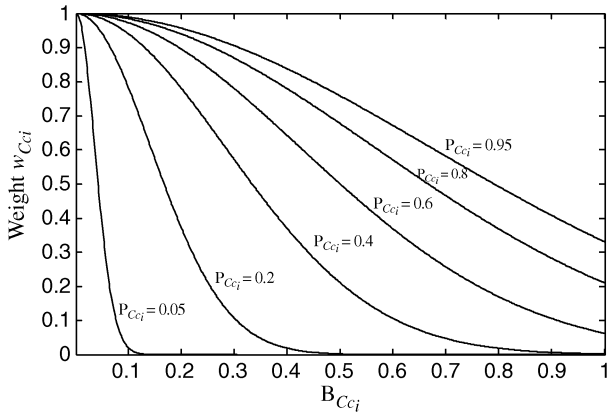

Fig. 4. Weight assignment for code-cube Cc_i according to B_{Cc_i}, its probability of being part of the empty background, for various values of P_{Cc_i}, its probability of being located peripherally close to the VOI.

Note that after the image is coded, the calculation of the code-cube weights for any VOI requires only the recomputation of two values for each code-cube, 1) its probability of being peripherally close to the VOI (i.e., value P_{Cc_i}), and 2) its contribution to the VOI (i.e., value ρ_{Cc_i}). There is no need to recompute the code-cube probabilities of being within the empty background (i.e., values B_{Cc_i}), since these probabilities are independent of the VOI and are calculated only once during the coding process (values B_{Cc_i} are stored as header information).

C. Creation of an Optimized Scalable Layered Bit-Stream

The bit-stream of each code-cube Cc_i may be independently truncated to any of a collection of different lengths, due to the entropy coding process, which is performed using a number of coding passes. We organize these truncated bit-streams into a number of quality layers to create a scalable layered bit-stream. This is done by collecting the incremental contributions from the various code-cubes into the quality layers such that the code-cube contributions result in a rate-distortion optimal representation of the 3-D image, for each quality layer L. The code-cube incremental contributions into each quality layer are stored as header information during the coding process.

After the creation of the scalable layered bit-stream, the main objective is to reorder this bit-stream, so that the code-cubes that constitute the VOI are included earlier in conjunction with contextual background information. We achieve this by finding

the optimal collection of truncated bit-streams that minimizes the overall distortion of the reconstructed 3-D image at quality layer L, while attaining VOI decoding capabilities. We incorporate weight w_{Cc_i}, as calculated in Section II-A, in the optimization process to achieve a gradual increase in peripheral quality around the VOI.

In this work, we employ the mean square error (MSE) to quantify the distortion of code-cube Cc_i at quality layer L

$$\mathrm{M}_{Cc_i}^{L} = \frac{1}{K}\sum_{k=1}^{K}(c_k - \hat{c}_k)^2 \tag{7}$$

where c_k is the kth sample of Cc_i, $\hat{c}_k$ is the quantized representation of the kth sample of Cc_i associated with the truncated bit-stream at quality layer L, and K is the total number of samples in Cc_i. The MSE is easily calculated by using the information about the code-cube contributions into quality layer L stored as header information during the coding process.

The MSE of code-cube Cc_i at quality layer L in sub-band s on a per-voxel basis over the entire 3-D image may then be calculated as

$$\overline{\mathrm{M}}_{Cc_i}^{L} = \frac{g_s}{N_s}\frac{q_s}{Q}\mathrm{M}_{\mathrm{Cc_i}}^{L} = 2^{2r}\frac{g_s}{N_s}\mathrm{M}_{\mathrm{Cc_i}}^{L} \tag{8}$$

where Q is the total number of image voxels, r is the decomposition level to which Cc_i belongs ($r = 1$ corresponds to the first decomposition level), $q_s = Q/2^{2r}$ is the number of coefficients in s, N_s is the number of code-cubes in s (the code-cubes are of equal size), $\mathrm{M}_{Cc_i}^{L}$ is as defined in (7), and g_s is a factor used to compensate for the nonenergy preserving characteristics of the bi-orthogonal Le Gall 5/3 wavelet filter. Factor g_s is a function of the specific wavelet filters used for reconstruction and is calculated from the filter coefficients [26].

In order to attain a gradual increase in peripheral quality around the VOI, we define a weighted MSE for code-cube Cc_i over the entire reconstructed 3-D image as follows:

$$\hat{\overline{\mathrm{M}}}_{Cc_i}^{L} = \frac{1}{(1 + w_{Cc_i})}\overline{\mathrm{M}}_{Cc_i}^{L} \tag{9}$$

where w_{Cc_i} is the weight of Cc_i as defined in (4) and $\overline{\mathrm{M}}_{Cc_i}^{L}$ is as defined in (8). Note that for code-cubes within the VOI, $w_{Cc_i} = 1$ and $\hat{\overline{\mathrm{M}}}_{\mathrm{Cc_1}}^{L} = (1/2)\overline{\mathrm{M}}_{\mathrm{Cc_1}}^{L}$. However, for code-cubes outside the VOI, $w_{Cc_i} < 1$ and $\hat{\overline{\mathrm{M}}}_{\mathrm{Cc_1}}^{L} > (1/2)\overline{\mathrm{M}}_{\mathrm{Cc_1}}^{L}$. The latter translates into a greater distortion, at quality layer L, for those code-cubes with a low mean energy content and located peripherally far from the VOI.

Thus, for a 3-D image coded using a total of I code-cubes, the overall distortion at quality layer L is

$$D^{L} = \sum_{i=1}^{I}\hat{\overline{\mathrm{M}}}_{Cc_i}^{L}. \tag{10}$$

The key to attaining VOI decoding capabilities at quality layer L, is to include only the truncated bit-streams of those code-cubes within the VOI. Under this condition, the output bit-stream at quality layer L is the summation of the truncated bit-streams of the code-cubes within the VOI

$$Y^{L} = \sum_{Cc_i \in \mathrm{VOI}} y_{Cc_i}^{L} \tag{11}$$

where Y^L is the output bit-stream at quality layer L and $y_{Cc_i}^{L}$ is the truncated bit-stream of Cc_i at quality layer L. The bit-rate of Y^L is then the summation of the bit-rates of each $y_{Cc_i}^{L}$

$$R_{Y^L} = \sum_{Cc_i \in \mathrm{VOI}} R_{y_{Cc_i}^{L}} \tag{12}$$

where R_{Y^L} denotes the overall bit-rate of Y^L and $R_{y_{Cc_i}^{L}}$ denotes the bit-rate of $y_{Cc_i}^{L}$.

The overall distortion of the reconstructed 3-D image at quality layer L, assuming the output bit-stream in (11), is thus

$$D^{L} = \sum_{Cc_i \in \mathrm{VOI}}\hat{\overline{\mathrm{M}}}_{Cc_i}^{L} + \sum_{Cc_i \notin \mathrm{VOI}}\hat{\overline{\mathrm{m}}}_{Cc_i}^{L} \tag{13}$$

where $\hat{\overline{\mathrm{m}}}_{Cc_i}^{L}$ denotes the weighted MSE added to the overall distortion D^L if $y_{Cc_i}^{L}$ is not included in layer L, and $\hat{\overline{\mathrm{M}}}_{\mathrm{Cc_1}}^{L}$ is as defined in (9). Using (7)–(9), $\hat{\overline{\mathrm{m}}}_{Cc_i}^{L}$ is calculated by equating $\hat{c}$, the quantized representation of the kth sample of code-cube Cc_i, to zero.

In order to increase the overall quality of the reconstructed 3-D image at quality layer L, while retaining the VOI decoding capabilities and allowing for the decoding of contextual background information, we encode some bit-streams $y_{Cc_i}^{L} \notin$ VOI along with bit-streams $y_{Cc_i}^{L} \in$ VOI. For a maximum bit-rate at quality layer L, some bit-streams $y_{Cc_i}^{L} \in$ VOI in (11) may have to be discarded in order to accommodate bit-streams $y_{Cc_i}^{L} \notin$ VOI. Due to the resolution scalability features of the output bit-stream, bit-streams $y_{Cc_i}^{L} \in$ VOI should be discarded in a sequential order starting with those comprising the first decomposition level (i.e., the highest-frequency sub-bands) and ending with those comprising the last decomposition level (i.e., the lowest-frequency sub-bands). Hence, the distortion in (13) can be expressed as follows:

$$\begin{aligned} D^{L} &= \sum_{i=i}^{I}\hat{\overline{\mathrm{M}}}_{Cc_i}^{L}\delta\left(y_{Cc_i}^{L}\right) + \sum_{i=i}^{I}\hat{\overline{\mathrm{m}}}_{Cc_i}^{L}\left[1 - \delta\left(y_{Cc_i}^{L}\right)\right] \\ &= \sum_{i=i}^{I}\delta\left(y_{Cc_i}^{L}\right)\left[\hat{\overline{\mathrm{M}}}_{Cc_i}^{L} - \hat{\overline{\mathrm{m}}}_{Cc_i}^{L}\right] + \sum_{i=i}^{I}\hat{\overline{\mathrm{m}}}_{Cc_i}^{L} \end{aligned} \tag{14}$$

where $\delta(y_{Cc_i}^{L})$ is 1 if $y_{Cc_i}^{L}$ is included in layer L (otherwise it is zero).

In order to attain the optimal overall reconstruction quality of the 3-D image at quality layer L, we minimize D^L in (14) under two bit-rate constraints

$$\begin{aligned} &\sum_{i=1}^{I} R_{y_{Cc_i}^{L}}\delta\left(y_{Cc_i}^{L}\right) \le R_{Y^L} \\ &\sum_{Cci \notin \mathrm{VOI}} R_{y_{Cc_i}^{L}}\delta\left(y_{Cc_i}^{L}\right) < \sum_{Cci \in VOI} R_{y_{Cc_i}^{L}}\delta\left(y_{Cc_i}^{L}\right) \end{aligned} \tag{15}$$

TABLE IV
3-D TEST MEDICAL IMAGES AND CORRESPONDING VOI COORDINATES AND CODE-CUBE SIZES

Modality	Dimensions Size {x,y,z}(mm) (slices:pixels per slice:bpv)	VOI coordinates		Code-cube size* $a \times a \times a$ (samples)	Scaling value†
		$p(x,y,z)$ (mm)	$m[X,Y,Z]$ (mm)		
1. MRI	{240,240,33} (11:512×512:8)	p(117,24,9)	m[85,140,12]	32×32×32	3
2. MRI	{256,256,100} (100:256×256:8)	p(128,0,0)	m[128,256,100]	16×16×16	4
3. MRI	{272,272,100} (50:512×512:8)	p(44,31,20)	m[170,159,54]	32×32×32	3
4. CT	{270,270,120} (120:512×512:12)	p(122,21,50)	m[111,212,50]	32×32×32	6
5. CT	{270,270,100} (100:512×512:12)	p(48,48,0)	m[185,137,30]	32×32×32	5
6. CT	{480,480,200} (200:512×512:12)	p(47,83,30)	m[353,286,40]	32×32×32	6

* Defined for the first level of decomposition.
† Used in the GSB method to scale up coefficients associated with the VOI above background coefficients.
MRI: magnetic resonance imaging. CT: computed tomography. VOI: volume of interest.

where $R_{y^L_{Cc_i}}$ is the bit-rate of $y^L_{Cc_i}$, R_{Y^L} is the maximum available bit-rate at quality layer L, and $\delta(y^L_{Cc_i})$ is 1 if $y^L_{Cc_i}$ is included in layer L (otherwise it is zero). Note that the constraints in (15) force the bit-rate spent on bit-streams $y^L_{Cc_i} \notin$ VOI to be less than the bit-rate spent on bit-streams $y^L_{Cc_i} \in$ VOI. This guarantees that the VOI is decoded at higher quality than the rest of the 3-D image.

We solve the optimization problem defined in (14), (15) by finding the points that lie on the lower convex hull of the rate-distortion plane corresponding to the possible sets of bit-stream assignments.

III. EXPERIMENTAL RESULTS AND DISCUSSION

We obtained two sets of experimental results. The first set evaluated the performance of the proposed method for VOI decoding at various bit-rates, including lossless reconstruction. The second set evaluated the effect of code-cube sizes on coding performance and size of the decoded VOI. We conclude this section with a discussion on the complexity of the proposed method.

A. Evaluation of VOI Decoding at Various Bit-Rates

Our test data set consisted of three 8-bit MRI and three 12-bit CT sequences of various resolutions. We defined a single VOI comprising clinically relevant information in each of the test sequences. The characteristics of the 3-D test sequences, the corresponding VOI coordinates and code-cube sizes used for entropy coding are summarized in Table IV. Sequence 1 comprises MRI slices (sagittal view) of a human spinal cord; Sequence 2 comprises MRI slices (axial view) of a human head; and Sequence 3 comprises MRI slices (sagittal view) of a human knee. The test CT sequences comprise consecutive slices (axial view) of the "Visible Male" (Sequences 4 and 5) and "Visible Woman" (Sequence 6) data sets maintained by the National Library of Medicine (NLM) [27]. In this work, the VOI is defined in the spatial domain by two sets of values, $p(x,y,z)$ and $m[X,Y,Z]$; where $p(x,y,z)$ denotes the lower-left corner coordinates closest to the coordinate origin and $m[X,Y,Z]$ denotes the dimensions of the VOI.

We compared the performance of the proposed compression method to that of 3D-JPEG2000 with VOI coding, using the MAXSHIFT and GSB methods. 3D-JPEG2000 is the extension of JPEG2000 for compression of 3-D images [28], [29]. 3D-JPEG2000 employs a 3-D discrete wavelet transform across the slices with the resulting 3-D sub-bands being entropy coded by first grouping coefficients into smaller 3-D sections called 3-D code-blocks. As mentioned earlier, MAXSHIFT scales up the coefficients associated with a VOI well above the background coefficients. At the decoder side, the nonzero VOI and background coefficients are identified by their magnitude, and thus, VOI coefficients are completely decoded before any background coefficients. The GSB method, on the other hand, scales up the coefficients associated with a VOI by a certain scaling value. Depending on the scaling value, some of the bits of the VOI coefficients may be encoded in conjunction with the bits of the background coefficients. The GSB method requires, however, the generation of a VOI mask at the encoder and decoder sides, as well as the coding and transmission of the VOI shape information, which may increase the computational complexity and overall bit-rate of the compressed bit-stream [30].

It is important to note that due to the scaling-up process performed by MAXSHIFT, the entropy decoder in 3D-JPEG2000 must be capable of decoding a large number of bit-planes. Current decoder implementations conforming to the JPEG2000 standard may not be capable to decode such large number of bit-planes, which renders the MAXSHIFT method not suitable for lossless reconstruction of 12-bit medical imaging data with VOI decoding capabilities. In this work, we used the *OpenJPEG* implementation of 3D-JPEG2000.[1]

In our proposed compression method, we employed the Le Gall 5/3 wavelet filter implemented using the lifting step scheme to decompose the test images with four levels of decomposition in all three dimensions. We created a layered output bit-stream whose reconstruction quality progressively improves up to lossless reconstruction.

For the case of 3D-JPEG2000, we employed four levels of decomposition in all three dimensions. We losslessly coded the resulting 3-D sub-bands using 3-D code-blocks and precincts to group coefficients describing the same 3-D spatial region at the same decomposition level. The dimensions of the 3-D code-blocks and precincts were selected to match the dimensions of the code-cubes employed in our proposed method, as tabulated in column 4, Table IV. For each test sequence, we created a layered output bit-stream whose reconstruction quality progressively improves up to lossless reconstruction. We employed the MAXSHIFT method (8-bit sequences) and the GSB method (8- and 12-bit sequences) to define a VOI according to the VOI coordinates tabulated in column 3, Table IV.

[1]http://www.openjpeg.org

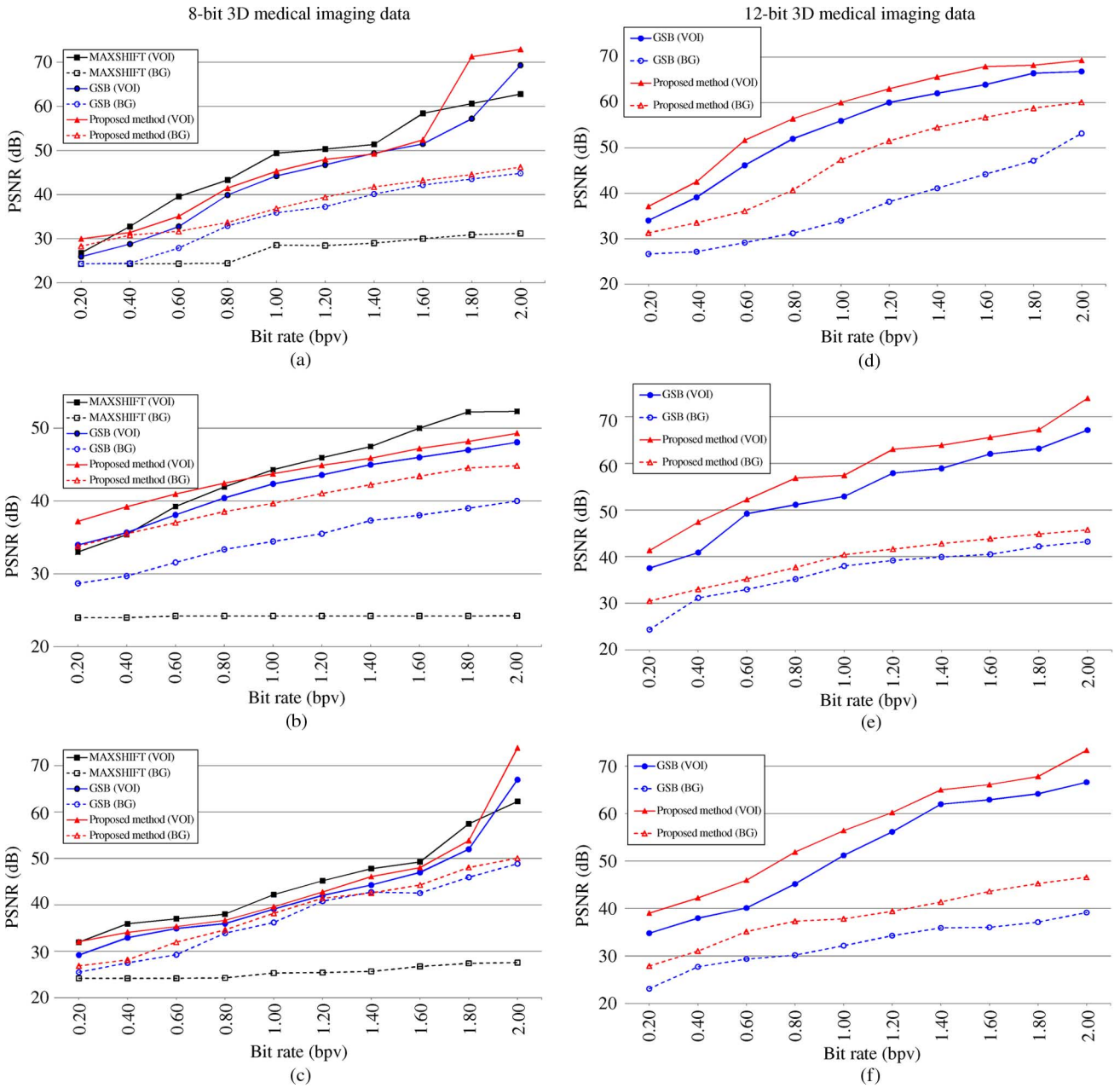

Fig. 5. PSNR values (in dB) for the VOI and background (BG) of 8-bit and 12-bit 3-D medical imaging data decoded at various bit-rates after compression using different methods (see Table IV). (a) Sequence 1, MRI slices (sagittal view) of a human spinal cord. (b) Sequence 2, MRI slices (axial view) of a human head. (c) Sequence 3, MRI slices (sagittal view) of a human knee. (d) Sequence 4 and (e) Sequence 5, consecutive CT slices (axial view) of the "Visible Male" data set maintained by the National Library of Medicine (NLM) [27]. (f) Sequence 6, consecutive CT slices (axial view) of the "Visible Woman" data set maintained by the NLM.

In order to measure the reconstruction quality of the VOI and background at different bit-rates, we employed the PSNR, which for a 3-D image of bit-depth m is defined by

$$\mathrm{PSNR} = 20\log_{10}\frac{(2^m-1)}{\sqrt{\mathrm{MSE}}}$$
$$\mathrm{MSE} = \frac{1}{K}\sum_{k=1}^{K}(c_k-\hat{c}_k)^2 \qquad (16)$$

where MSE denotes the mean square error, (2^m-1) is the maximum voxel value in the 3-D image, K is the total number of voxels in the area to be evaluated (e.g., the VOI), c_k and $\hat{c}_k$ are the original and reconstructed values of the kth voxel, respectively.

In the case of the GSB method, we empirically selected the scaling value that produces the VOI quality (in terms of the PSNR) most similar to the VOI quality attained by our proposed method, while still allowing for decoding of partial background information at different bit-rates. The selected scaling values for the test sequences are tabulated in column 5, Table IV.

Fig. 5 plots the PSNR values of the VOI and background of the 3-D test sequences after decoding at a variety of bit-rates (in

bits per voxel, bpv). It can be seen that, for all test sequences, the proposed method achieves higher PSNR values for the VOI and background than those achieved by the GSB method. Even though the GSB method allows for decoding of partial background information in conjunction with the VOI, this partial information is determined by manually selecting a scaling value for the background coefficients, which may affect the coding performance. The proposed method requires no manual selection of a scaling value, as it employs a weighting model and an optimization technique to determine the optimal amount of background information that minimizes the overall distortion of the image. This is done by reordering the output bit-stream after compression. As a result, in the peripheral regions of the VOI, low-frequency code-cubes with large weights have a greater opportunity to be included earlier in the output bit-stream. In the VOI, both low- and high-frequency code-cubes are included earlier in the output bit-stream because of their large weights.

It can also be seen in Fig. 5 that, for 8-bit sequences, MAXSHIFT achieves higher PSNR values for the VOI than those achieved by the proposed method, especially at low bit-rates (e.g., bit-rates lower than 0.40 bpv, Sequences 1 and 3). This is expected, since MAXSHIFT first decodes the VOI coefficients before decoding any background coefficients. Note that the apparent high PSNR values achieved by MAXSHIFT for the background at low bit-rates are due to the smearing effects of the wavelet filter, which may result in some background areas surrounding the VOI being decoded in conjunction with the VOI. Also note that as the bit-rate increases, the quality of the VOI decoded by the three evaluated methods tends to be very similar, since more bits are decoded and the reconstruction quality approaches the lossless case.

It is important to mention that the plots in Fig. 5(a)–(c) present different behaviors for the case of the MAXSHIFT method. This is mainly due to the size of the VOI, which may affect the number of bits in each bit-plane needed to fully reconstruct the VOI before the background. For small VOIs, e.g., those decoded in Sequences 1 and 3, MAXSHIFT achieves higher PSNR values for the VOI than the proposed method at bit-rates lower than 0.4 bpv [see Fig. 5(a) and (c)]. In this case, the VOI does not include a large proportion of significant coefficients and the number of most significant bit-planes in the VOI and background are similar. MAXSHIFT is therefore capable of decoding the VOI at high qualities at low bit-rates. For larger VOIs, e.g., that decoded in Sequence 2 (where the VOI comprises half of the entire volume), the proportion of significant coefficients in the VOI is larger and, therefore, a larger number of bits is needed to fully recover the VOI before the background. This explains the lower PSNR values achieved by MAXSHIFT for the VOI at bit-rates lower than 0.80 bpv when compared to the proposed method [see Fig. 5(b)].

Lossless compression ratios and bit-rates for the three evaluated methods are tabulated in Table V. The proposed method achieves compression ratios comparable to those achieved by MAXSHIFT and the GSB method, with the additional advantage of allowing for decoding any VOI from the same compressed bit-stream.

TABLE V
LOSSLESS COMPRESSION RATIOS AND BIT-RATES OF 3-D MEDICAL IMAGES USING VARIOUS COMPRESSION METHODS

Modality Size {x,y,z}(mm) (slices:pixels per slice:bpv)	**Compression method ***		
	MAXSHIFT method	GSB method	**Proposed method**
	compression ratio (bit-rate, bits per voxel)		
1. MRI {240,240,33} (11:512×512:8)	2.39:1 (3.34 bpv)	2.46:1 (3.25 bpv)	**2.44:1 (3.27 bpv)**
2. MRI {256,256,100} (100:256×256:8)	1.98:1 (4.04 bpv)	2.02:1 (3.96 bpv)	**2.00:1 (4.00 bpv)**
3. MRI {272,272,100} (50:512×512:8)	4.25:1 (1.88 bpv)	4.31:1 (1.85 bpv)	**4.34:1 (1.83 bpv)**
4. CT {270,270,120} (120:512×512:12)	N/A	3.25:1 (4.92 bpv)	**3.22:1 (4.96 bpv)**
5. CT {270,270,100} (100:512×512:12)	N/A	2.31:1 (6.92.bpv)	**2.36:1 (6.77 bpv)**
6. CT {480,480,200} (200:512×512:12)	N/A	2.47:1 (6.47 bpv)	**2.49:1 (6.41 bpv)**

* A single VOI was defined in each sequence as specified in Table IV.

MRI: magnetic resonance imaging. CT: computed tomography. VOI: volume of interest.

N/A: not applicable.

Fig. 6 illustrates sample reconstructed slices at 0.6 bpv belonging to the VOI of Sequences 1 and 3 (see Table IV). It can be seen that the GSB and the proposed methods are capable to decode the VOI while still including partial background information, which allows placing the VOI into the context of the 3-D image, in this case the sagittal view of a human spinal cord (Sequence 1), and a human knee (Sequence 3). The proposed method, however, decodes the background information peripherally around the VOI according to the mean energy of the code-cubes, which results in a higher reconstruction quality than that attained by the GSB method. Note that the VOIs decoded by the MAXSHIFT and the GSB methods appear to be of different size when compared to the VOI decoded by the proposed method. Let us remember that the MAXSHIFT and the GSB methods work on a coefficient-basis and thus, both methods are able to decode more precisely only those coefficients within the VOI at higher quality than the background coefficients. In other words, the granularity of the MAXSHIFT and GSB methods in representing a VOI in the sub-band domain is one wavelet coefficient, as opposed to the proposed method, where the granularity is one code-cube. Therefore, in the proposed method, the code-cube sizes have a direct impact on the size of the decoded VOI. This will be further discussed in the next subsection.

B. Evaluation of the Effect of Code-Cube Sizes

In this section, we evaluated the trade-off between the size of the decoded VOI and coding performance, for various code-cube sizes. In order to measure how similar the size and location of the decoded VOI are to the size and location of the desired VOI, we employ a VOI shape decoding quality defined by a set of two values, $e_p(x, y, z)$ and v_r; where $e_p(x, y, z)$ denotes the absolute value of the difference between the lower left-hand

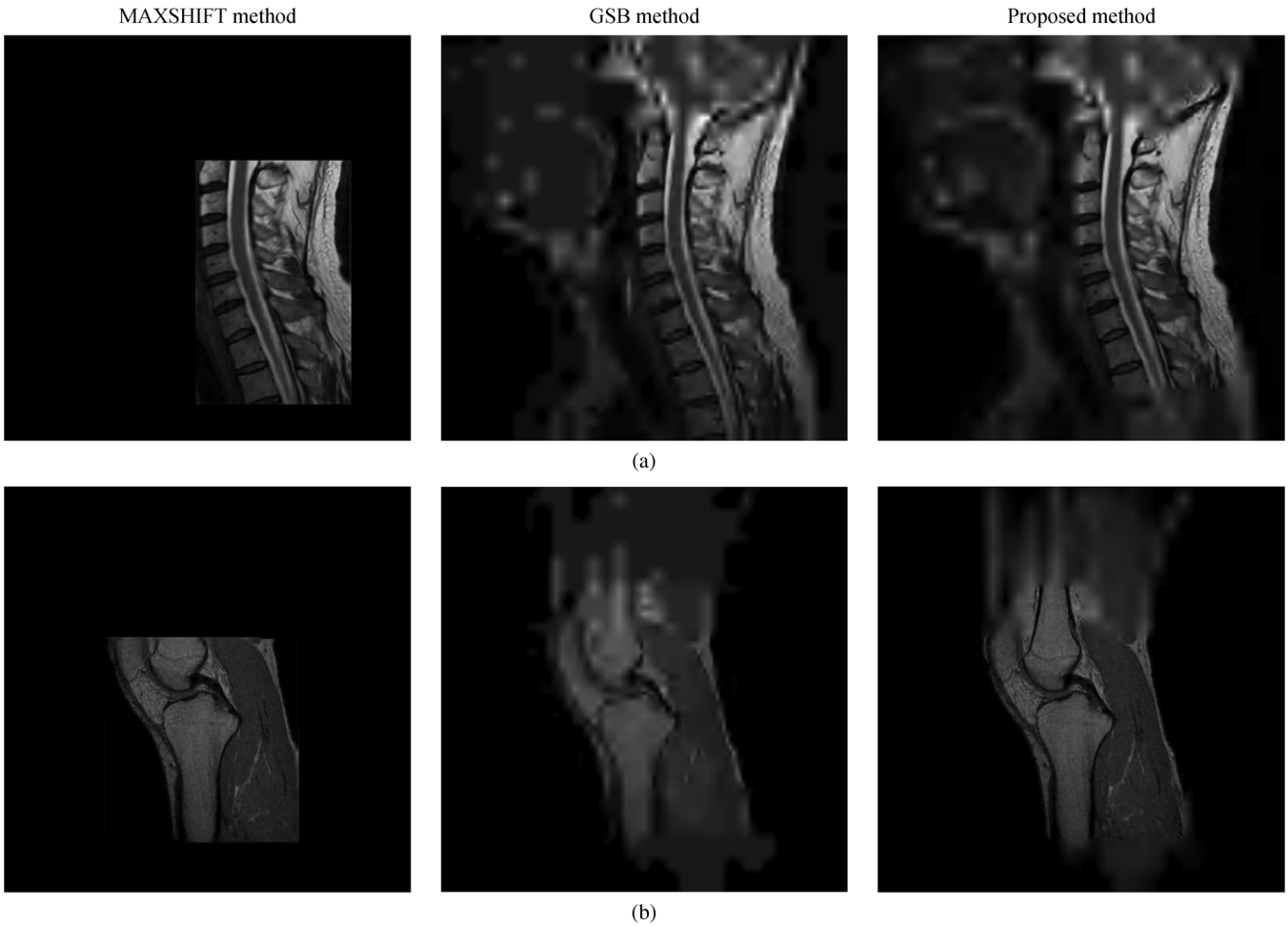

Fig. 6. (a) Slice no. 5 belonging to the VOI of Sequence 1 and (b) slice no. 22 belonging to the VOI of Sequence 3 (see Table IV) reconstructed at 0.6 bpv after compression using the MAXSHIFT method, the GSB method, and the proposed method. Observed PSNR values were (a) 39.56 dB (VOI) and 24.33 dB (background) for MAXSHIFT; 32.97 dB (VOI) and 27.90 dB (background) for the GSB method; and 35.07 dB (VOI) and 31.60 dB (background) for the proposed method; and (b) 39.99 dB (VOI) and 24.16 dB (background) for MAXSHIFT; 34.89 dB (VOI) and 29.25 dB (background) for the GSB method; and 35.31 dB (VOI) and 31.90 (background) for the proposed method.

corner coordinates of the desired VOI and those of the decoded VOI, and v_r is a real number defined by

$$v_r = \frac{\sum_{Cc_i \in \mathrm{VOI}} V_{Cc_i}}{V_{\mathrm{VOI}}} \tag{17}$$

where V_{Cci} is the volume size of the region represented by code-cube Cc_i in the spatial domain (the summation of all $Cc_i \in \mathrm{VOI}$ comprise the decoded VOI), and V_{VOI} is the volume size of the desired VOI in the spatial domain. A value $v_r = 1$ means that the decoded VOI is equal in volume size to the desired VOI, a value $v_r > 1$ means that the decoded VOI is larger in volume size than the desired VOI, whereas a value $v_r < 1$ means that the decoded VOI is smaller in volume size than the desired VOI.

Let us remember that each code-cube is associated with a limited spatial region due to the finite footprint of the wavelet kernel. It is thus expected that small code-cube sizes will result in higher VOI shape decoding qualities (i.e., $e_p(x, y, z)$ values close to (0,0,0) with v_r values close to 1). However, small code-cube sizes may also result in reduced coding performance due to the increased number of independent bit-streams needed to represent all the code-cubes at each quality layer L.

Fig. 7 plots the VOI shape decoding quality and PSNR values for the VOI of Sequences 1 and 4 (see Table IV) after decoding at a variety of bit-rates using different code-cube sizes. Fig. 8 shows sample reconstructed slices at 0.6 bpv of Sequence 1 after encoding using different code-cube sizes.

As expected, results in Fig. 7 show that as the code-cube size is reduced the coding performance decreases, but the VOI shape decoding quality increases. This can be seen in Fig. 8(b), where the VOI seems to be larger than in Fig. 8(d)–(f), because the code-cubes are not small enough for the VOI to be accurately decoded. Also note the blocky artifacts when employing code-cubes of $8 \times 8 \times 8$ samples, which are the result of the low coding performance due to the increased number of independent bit-streams needed to represent all code-cubes. In this case, code-cubes of $32 \times 32 \times 32$ and $64 \times 64 \times 64$ samples (defined for the first decomposition level) present the best trade-off between VOI shape decoding quality and coding performance. If the VOI coordinates are known *a priori* the entropy coding process, large code-cubes may be employed for large VOIs or if the whole volume needs to be decoded; whereas small code-cubes may be employed for small VOIs or if a high VOI shape decoding quality is required.

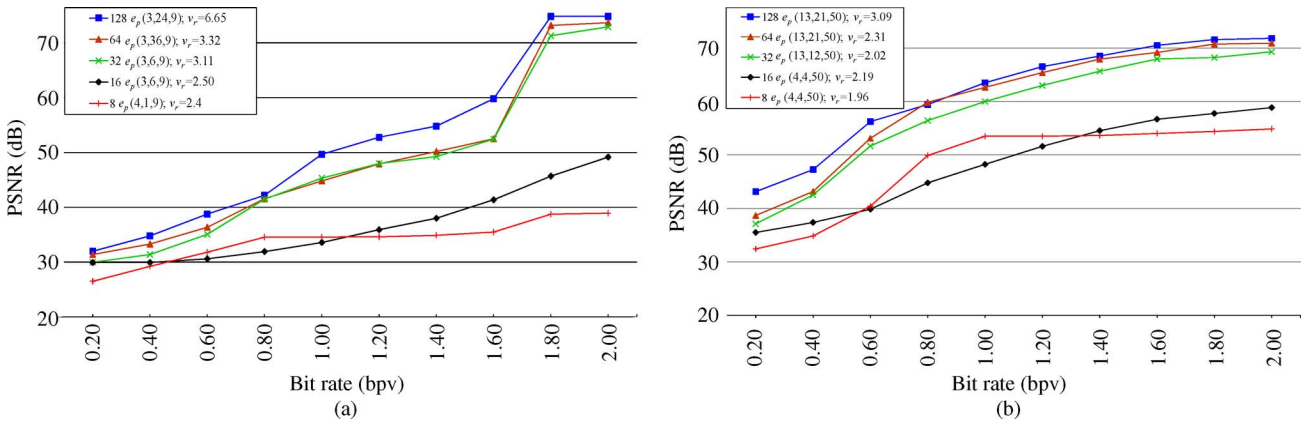

Fig. 7. PSNR (in dB) for the VOI and VOI shape decoding quality values of (a) Sequence 1 and (b) Sequence 4 after decoding at a variety of bit-rates using different code-cube sizes (see Table IV).

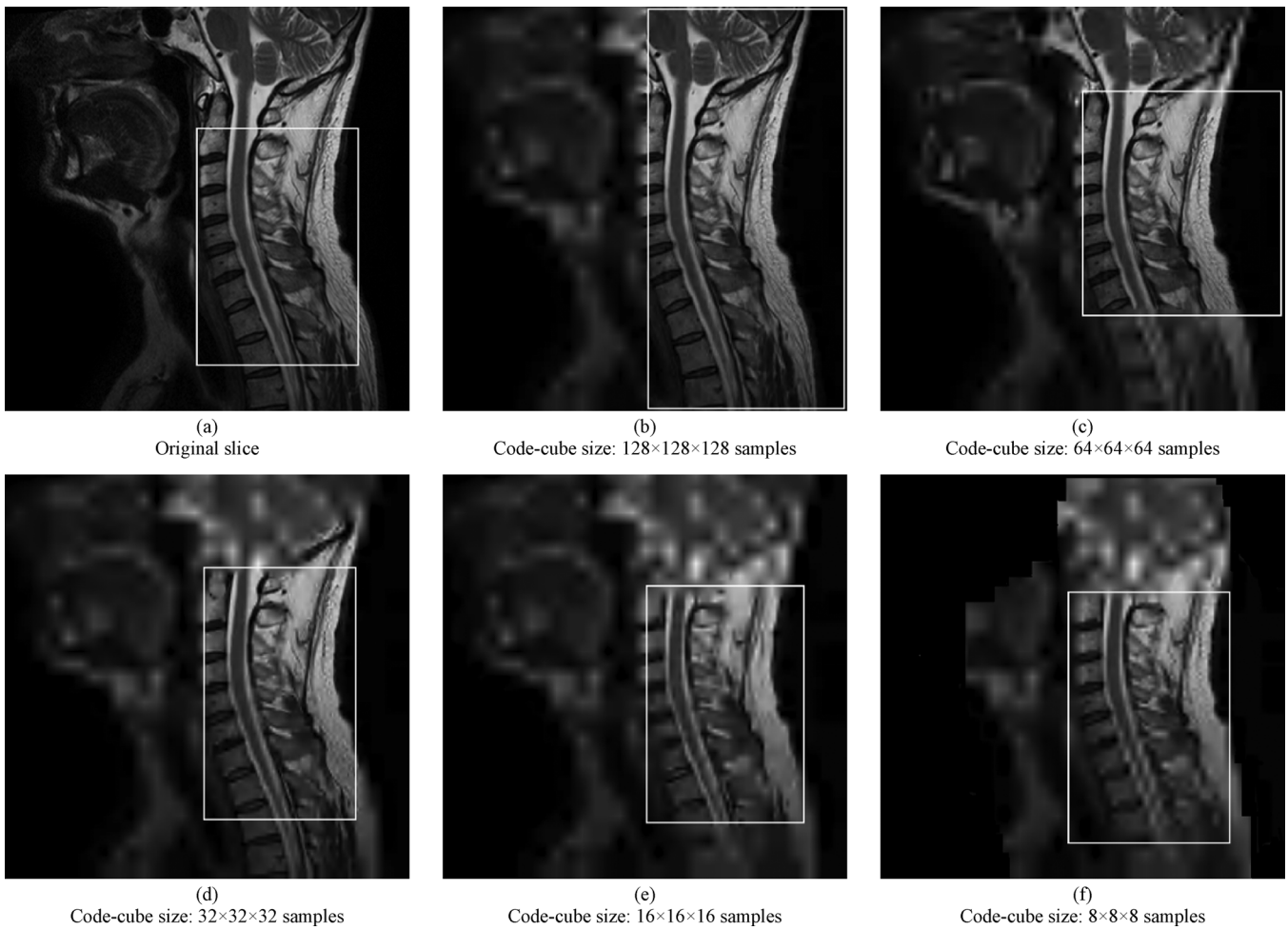

Fig. 8. (a) Original slice no. 5 of Sequence 1. The voxels belonging to the desired VOI are delimited by a square area (see Table IV). (b)–(f) Slice no. 5 of Sequence 1 reconstructed at. 0.6 bpv after coding using various code-cubes sizes with four level of decomposition (code-cube sizes are defined for the first level of decomposition). The voxels belonging to the decoded VOI are delimited by a square area. (a) Original slice; (b) code-cube size: $128 \times 128 \times 128$ samples; (c) code-cube size: $64 \times 64 \times 64$ samples; (d) code-cube size: $32 \times 32 \times 32$ samples; (e) code-cube size: $16 \times 16 \times 16$ samples; (f) code-cube size: $8 \times 8 \times 8$ samples.

C. Computational Complexity Considerations

We conclude our performance evaluation with a brief discussion regarding the complexity of the proposed compression method. Compared to 3D-JPEG2000 with VOI coding (MAXSHIFT and GSB methods), the proposed method presents a higher complexity at the encoder side due mainly to the bit-stream reordering procedure. This augmented complexity is a consequence of the calculation of the code-cube weights and the layer optimization technique, which needs to be performed each time a VOI is to be decoded.

It is important to remember that, in the proposed method, the entropy coding process needs to be performed only once for a 3-D medical image, since the decoding of a VOI simply requires

the reordering of the compressed bit-stream. As mentioned earlier, the calculation of the code-cube weights for a requested VOI simply requires the recomputation of two values for each code-cube. Moreover, the MSE required during the layer optimization technique is easily calculated from the information about the code-cube contributions into each quality layer, which is stored as header information during the coding process.

At the decoder side, the complexity of the proposed method is very similar of that of the MAXSHIFT method, since there is no need for the decoder to reorder the bit-stream prior to decoding. Compared to the GSB method, the decoding complexity of the proposed method is lower, since the GSB method requires the generation of a VOI mask prior to decoding.

Finally, it is important to remark that in the proposed method, all information needed to perform the bit-stream reordering procedure and layer optimization technique is stored and transmitted as header information. In the case of the test sequences evaluated in this work, this additional information represents a mere 0.04%–0.5% of the compressed bit-rate.

IV. Conclusion

We presented a novel scalable 3-D medical image compression method with optimized VOI coding within the framework of interactive telemedicine applications. The method is based on a 3-D integer wavelet transform and a modified version of EBCOT that exploits correlations between wavelet coefficients in three dimensions and generates a scalable layered bit-stream. The method employs a bit-stream reordering procedure and an optimization technique to optimally encode any VOI at the highest quality possible in conjunction with contextual background information from a lossy to a lossless representation. We demonstrated the two main novelties of the method; namely, the ability to decode any VOI from the compressed bit-stream without the need to recode the entire 3-D image; and the ability to enhance the visualization of the data at any bit-rate by including contextual background information with peripherally increasing quality around the VOI. We evaluated the performance of the proposed method on real 8-bit and 12-bit 3-D medical images of various resolutions. We demonstrated that the proposed method achieves higher reconstruction qualities than those achieved by 3D-JPEG2000 with VOI coding at a variety of bit-rates. We also demonstrated that the proposed method attains lossless compression ratios comparable to those attained by 3D-JPEG2000 with VOI coding. Finally, we studied the effect on coding performance and VOI decoding capabilities of the proposed method with different coding parameters.

References

[1] P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro-Nieto, and J. Cornelis, "Wavelet coding of volumetric medical datasets," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 441–458, Mar. 2003.

[2] Z. Xiong, X. Wu, S. Cheng, and J. Hua, "Lossy-to-lossless compression of medical volumetric images using three-dimensional integer wavelet transforms," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 459–470, Mar. 2003.

[3] X. Wu and T. Qiu, "Wavelet coding of volumetric medical images for high throughput and operability," *IEEE Trans. Med. Imag.*, vol. 24, no. 6, pp. 719–727, Jun. 2005.

[4] G. Menegaz and J. P. Thirian, "Three-dimensional encoding/two-dimensional decoding of medical data," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 424–440, Mar. 2003.

[5] R. Srikanth and A. G. Ramakrishnan, "Contextual encoding in uniform and adaptive mesh-based lossless compression of MR images," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1199–1206, Sep. 2005.

[6] V. Sanchez, R. Abugharbieh, and P. Nasiopoulos, "Symmetry-based scalable lossless compression of 3-D medical image data," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1062–1072, Jul. 2009.

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

[8] A. Said and W. Pearlman, "A new fast and efficient image coded based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[9] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[10] K. Krishnan. M. Marcellin, A. Bilgin, and M. Nadar, "Efficient transmission of compressed data for remote volume visualization," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1189–1199, Sep. 2006.

[11] Y. Liu and W. A. Pearlman, "Region of interest access with three-dimensional SBHP algorithm," in *Proc. SPIE*, 2006, vol. 6077, pp. 17–19.

[12] C. Doukas and I. Maglogiannis, "Region of interest coding techniques for medical image compression," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 5, pp. 29–35, Sep.–Oct. 2007.

[13] I. Ueno and W. Pearlman, "Region of interest coding in volumetric images with shape-adaptive wavelet transform," in *Proc. SPIE*, 2003, vol. 5022, pp. 1048–1055.

[14] *JPEG 2000 Part I: Final Draft International Standard (ISO/IEC FDIS15444-1)*, ISO/IECJTC1/SC29/WG1 N1855, Aug. 2000.

[15] J. Strom and P. C. Cosman, "Medical image compression with lossless regions of interest," *Signal Process.*, vol. 59, no. 3, pp. 155–171, Jun. 1997.

[16] A. Signoroni and R. Leonardi, "Progressive ROI coding and diagnostic quality for medical image compression," in *Proc. SPIE*, 1997, vol. 3309, no. 2, pp. 674–685.

[17] X. Bai, J. S. Jin, and D. Feng, "Segmentation-based multilayer diagnosis lossless medical image compression," in *Proc. ACM Int. Conf.*, Jun. 2004, vol. 100, pp. 9–14.

[18] 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.

[19] I. Daubechies and W. Sweldens, "Factoring wavelet transform into lifting steps," *J. Fourier Anal. Appl.*, vol. 41, no. 3, pp. 247–269, 1998.

[20] J. Xu, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Appl. Computat. Harmonic Anal.*, vol. 10, pp. 290–315, 2001.

[21] N. Zhang, M. Wu, S. Forchhammer, and X. Wu, "Joint compression-segmentation of functional MRI data sets," in *Proc. SPIE*, 2003, vol. 5748, pp. 190–201.

[22] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1397–1410, Oct. 2001.

[23] V. Sanchez, A. Basu, and M. K. Mandal, "Prioritized region of interest coding in JPEG2000," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 9, Sep. 2004.

[24] K. J. Wiebe and A. Basu, "Improving image and video transmission quality over ATM with fovea prioritization and priority dithering," *Pattern Recognit. Lett.*, vol. 22, pp. 905–915, 2001.

[25] Z. Wang and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243–254, Feb. 2003.

[26] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1996, pp. 387–395, 1996.

[27] The National Library of Medicine (NLM) [Online]. Available: http://www.nlm.nih.gov

[28] *Information Technology—JPEG 2000 Image Coding System—Part 2: Extensions*, ISO/IEC 15 444-2, 2002.

[29] *Information Technology—JPEG 2000 Image Coding System—Part 10: Extensions for Three Dimensional Data*, ISO/IEC 15 444-10, 2007.

[30] T. Bruylants, P. Schelkens, and A. Tzannes, "JP3D—Extensions for three dimensional data (part 10)," in *The JPEG2000 Suite*, P. Schelkens, A. Skodras, and T. Ebrahimi, Eds. New York: Wiley, 2009, pp. 218–219.