Wavelet-based compression of medical images: Protocols to improve resolution and quality scalability and region-of-interest coding

Peter Schelkens *, Adrian Munteanu, Jan Cornelis

Department of Electronics and Information Processing (ETRO), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
Accepted 02 November 1998

Abstract

The paper describes a methodology to improve the scalability support of an embedded bit stream, generated with a wavelet-based compression algorithm, and a generic protocol to handle multiple regions-of-interest (ROIs). The generic scheme, illustrated for embedded zero-tree wavelet (EZW) coding, exploits the inherent graceful degradation capabilities of wavelet-based compression methods and ensures an optimal trade-off between the image reconstruction quality and the compression ratio. Additionally, an efficient protocol is proposed to handle multiple ROIs in an interactive client-server set-up for telemedicine applications. The processing of the ROIs takes place in the wavelet domain. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Wavelet-based image compression; Region-of-interest coding; Telemedicine; Resolution and quality scalability

1. Introduction

Image processing applications tend to be implemented in heterogeneous environments. To handle the images efficiently, a suitable file format has to be determined, enabling fast and flexible exchange of the desired image data. This constraint calls for an efficient compression algorithm. In the requirement analysis for designing such an algorithm at least the following questions concerning the application domain are raised: (1) What is the source of the image: is it a projection X-ray image, a CT-scan, an MR-image, a PET-scan, etc.? (2) Do we want to store the image? If yes, should this be on a long-term basis or a short-term basis? What is the medium we use to store the image: a CD-ROM, a hard disc, a volatile computer memory, etc.? (3) How are we going to transfer the image to the visualisation device? Over an ISDN or T1 line, or are we operating on a wireless communication channel (e.g. WLAN)? (4) What is the visualisation device: a cathode ray-tube, a LCD screen, a printer, or a video projector? (5) What is the application type? What are the imposed restrictions? (6) Which features do we desire to be supported?

It is obvious that the above analysis results in a broad spectrum of application specific requirements putting high demands on the flexibility and quality of the compression algorithm [9]. We will try to illustrate some of the requirements by considering the following telemedicine application.

On the medical market portable imagery starts to show up. One example is a portable digital X-ray device. These monochromatic images tend to be big (in the range of 2.5k × 2k up to 4.5k × 3k for mammography), and additionally they have a broad contrast res-
olution (10–15 bpp) [9]. This results in a large amount of image data varying between 6.3 and 25.3 MB. Consider for example the technically least demanding situation of an image size equal to 6.3 MB. Assume transmission via the Iridium satellite network, with a bandwidth of 2400 bps, is used to transfer the image to the hospital, which is realistic when the patient is localised in a distant area and one needs a wireless communication facility. Hence, it will take 5 h and 50 min to transfer the uncompressed image to the hospital. Since the X-ray image can on average be losslessly compressed at a bit rate between 2.5 and 4 bpp, at least 1 h and 27 min is still needed. However, at a compression ratio of 100 (0.1 bpp) the hospital will receive a coarse version of the image after 3 min and 30 s. At this ratio it is most likely already feasible to allow the radiologist located at the hospital to produce a first diagnosis in the cases of serious injuries, and to instruct the medical staff at the patient’s location. Further, the encoder should continue to send image information to the guiding radiologist in the hospital allowing a refinement of the visualised X-ray image. The refinement data should be complementary to the previously transmitted image data, which assumes an embedded bit stream, in order to exploit the available bandwidth optimally. Now, two main options are available to the radiologist and/or the medical staff: (1) the image is progressively refined in case the coarse version does not visualise any pathology; (2) the radiologist selects a region-of-interest (ROI), based on the information present in the coarse image version or based on the instruction of the medical staff at the patient’s location. After sufficiently refining this limited region, the rest of image is transferred. The first option is addressed as quality scalability, while the second one is referred to as ROI coding. It is interesting to remark that wireless communications tend to suffer from relatively high bit error rates (BERs), highlighting the error resilience issue. To tackle errors, which may distort the complete image, it appears to be a very successful approach to code the most important information (e.g. the coarse version of the image) with very robust error correcting means. In the least significant parts of the transmitted bit stream a weaker error correcting method can then be applied. Finally, the displaying device in front of the radiologist has to be addressed. This will typically be a high-resolution monitor. A compression/decompression scheme putting specific weights on the different frequency components of the image yields an optimal visual perception of the X-ray image [4] by the radiologist.

For medical applications, in general, it is important to remark that the encoding algorithms should facilitate lossy-to-lossless compression. The lossless mode is very important in situations where legal disputes may pop-up. Significant in this context are the perceptually lossless [10] and semantically lossless paradigms [18] introduced by, respectively, Jayant and Signoroni. The first refers to the fact that the compressed image is visually indistinguishable from the compressed one, while the latter refers to the property that the diagnosis (semantical interpretation) will be identical for both images.

The above telemedicine example illustrates that the compression algorithm should be as flexible as possible, adapting itself to different applications and interactivity demands within a specific application. The need for features like lossy-to-lossless coding, scalability in resolution, scalability in quality, ROI coding and error resilience was demonstrated. Additionally, these algorithms should allow efficient and fast implementations, and of course, an optimal trade-off between the image reconstruction quality and the compression ratio should be ensured.

In this paper we will give an overview of still image compression algorithms (Section 2), and discuss the basics of embedded zero-tree wavelet (EZW) codecs (Section 3). Moreover, a bit-allocation technique (Section 4) and an improved method for obtaining an optimised progressive functionality (Section 5) are proposed. Finally, a generalised ROI coding protocol (Section 6) is introduced.

2. Still image compression algorithms

The current standard for still image compression (JPEG) [26] supports only partially the enumerated requirements. For instance, JPEG is capable of obtaining a reasonable compression performance at high and intermediate bit rates, but for low bit rates quality is poor. The breakdown at high compression is mainly due to the blocking artefacts, caused by the initial partitioning of the image in square blocks within the DCT-based decorrelating module and the block-based quantisation module. With respect to functional-
ity, JPEG supports lossy and lossless compression but not simultaneously: we have to select either the classic JPEG encoder for lossy coding, either the JPEG-LS encoder [25] for lossless coding. In addition, progressive refinement of the image (scalability in quality and resolution) is hard to support. Yet, scalable DCT-techniques are described [12,13], but they do not beat the performance of techniques proposed in following paragraphs.

A promising technique that offers better performance at high compression ratio, because it produces artefacts that are subjectively less annoying for the human visual system (HVS), is segmented image coding. The image is partitioned in segments with variable shapes depending on the image content. When the segment boundaries are coinciding with steep transitions (e.g. edges) in the image, large numerical errors around the segment boundaries in the decompressed image are not readily noticed by the HVS. Moreover, if the segmentation boundaries delineate regions that correspond to meaningful objects, these can be flexibly manipulated (e.g. pre-press applications). The definition of segment hierarchies and the predominance of contour information for visual recognition can lead to efficient progressive transmission schemes. Drawbacks are the complexity of designing robust generic segmentation techniques, the coding of the internal grey value distribution, the high computational complexity.

Fortunately, wavelet-based compression techniques do inherently meet the requirements extracted out of the telemedicine illustration. These methods are currently being subject of study for the new emerging still image compression standard JPEG2000 (ISO/IEC JTC1/SC29/WG1), being the successor of JPEG. The typical blocking artefacts, like the ones occurring in JPEG, are avoided since the image is not compressed in blocks. Similar (or sometimes even higher) quality as for JPEG can be obtained at high and intermediate bit rates. The ringing artefacts occurring at high compression ratios (mainly in the vicinity of edges and/or in textured regions) are generally less disturbing for the HVS than the JPEG blocking artefacts. In terms of functionality, the multiresolution data representation and the spatial locality properties of wavelets are beneficial for different types of flexibility. The computation time also does not increase in a way that it causes an insuperable bottleneck for nowadays processing units.

Studies showed that in terms of hardware complexity, the wavelet transform and the scalable DCT-modules [12,13] do have similar complexity in terms of memory usage and calculation load [11].

Unfortunately, standard wavelet compression techniques [1,16] do not allow a lossless reconstruction of the original image, even when retaining all the coefficients of the wavelet transform. The fact that coefficients are generated as real (floating point) numbers and have to be converted into integers when coded is responsible for this. The alternative is to use the lifting scheme [21] to generate non-linear integer-to-integer wavelet transforms [3,6,7]. The use of these transforms supports the lossy-to-lossless functionality and reduces the computational load for the transform part of the compression algorithm.

Although, the wavelet transform inherently supports scalability, the succeeding quantisation and entropy coding parts must respect this functionality too. A popular image coding technique featuring progressive transmission is the embedded zero-tree wavelet (EZW) coding [16]. An important aspect of EZW is that it successively approximates the image in a bitplane-by-bitplane fashion, while it prioritises the wavelet coefficients according to their magnitude, in contrast to the typical subband coding that prioritises according to the subband frequency. Moreover, it exploits the correspondence of spatially related pixels within the different subbands and within the same bit plane, using the concept of zero-trees (Fig. 1). This method produces significance maps (binary maps) describing the significance of the wavelet coefficients compared to a successively decreasing threshold values.

A more complex technique is set partitioning in hierarchical trees (SPIHT) [15]. The essential difference of the SPIHT coding process with respect to EZW is the way trees of coefficients are partitioned and sorted. An alternative method for the encoding of the significance maps, identified as the SQuare Partitioning (SQP) [4], efficiently encodes the positions of the significant coefficients in the wavelet image using a hierarchical structure of squares that group the insignificant coefficients in blocks of variable width. A partition rule is applied recurrently on the squares, selecting sets of binary elements in the significance map, and correspondingly, sets of coefficients of the wavelet transform matrix.
The current state-of-the-art in wavelet-based still image coding is the embedded conditional entropy coding of wavelet coefficients (ECECOW) [23]. This scheme uses the same scalar uniform quantiser (SAQ) as its predecessors, but the gain in performance is coming from the exploitation of higher order statistics in the coding of the significance maps. This leads to another inter-scale dependency model than the classical tree structures of the previous methods (EZW, SPIHT).

The explained techniques all support the generation of an embedded bit stream by successively approximating the image data. Generally, this is realised through comparison of the wavelet coefficients with gradually decreasing thresholds. However, for a specific threshold the wavelet coefficients of the low-frequency subbands are of course prioritised to those of the high-frequency subbands, but this does not necessarily result in an optimal visual weighting of the subbands. A solution to obtain optimal weighting will be elaborated in the following sections for EZW-coding. However, the demonstrated principles are also applicable to the other wavelet-based compression methods.

3. Embedded zero-tree wavelet coding

For reasons of clarity we will discuss the EZW-algorithm [16]. As mentioned earlier, the wavelet transform is organised in such a way that successive image resolutions are automatically obtained throughout the transformation process. Transmitting successively the subbands, starting from the low-frequency ones, already partially fulfils the graceful degradation (scalability) requirement. This technique is demonstrated in Fig. 2, where the horizontal and vertical axis represent the frequency scale and the bit plane level of the transformed image, respectively. In essence, first all the bit planes of the LL3 image are trans-
mitted, successively followed by the higher subband images. Of course, in this case the relation between the compression ratio and the image reconstruction quality can hardly be called optimal.

Nevertheless, when carefully studying the redundancy between spatially related pixels within the different subbands, a remarkable coherence is revealed. Shapiro figured out that exploiting this property enhances the compression performance [16]. He considered the dependence between spatially related pixels within different subbands as parent–child links. In the context of an \( N \)-level discrete wavelet transform, this means that a pixel in a level \( l \) subimage (either LH, HL or HH) corresponds spatially to four pixels in the level \( l-1 \) subimage with the same type of frequency constellation, i.e. LH, HL or HH (see Fig. 1(b)). The inter-band correlation is exploited using the fact that the probability is high that the pixel values of the children are smaller than a certain threshold, whenever the parent value satisfies this.

The full coding procedure is partitioned in two main repetitive scanning passes: the dominant pass and the subordinate pass. Before initiation of the scanning, three lists are created in the original algorithm: the dominant and subordinate lists, containing all the pixels which should be scanned during the dominant and the subordinate pass respectively, and the zero-tree list initialised before each dominant pass and keeping track of the pixels belonging to a zero-tree. However, if we carefully study the dominant and subordinate lists, we discover that the contained information is complementary. It is therefore more efficient to retain one list, which we address as the significance list, and that does the book-keeping of the pixels classified as significant.

The image is successively scanned starting from the low-frequency subbands up to the high-frequency subbands. During the first pass of the image, called the dominant pass, the pixel values are compared to a threshold level, corresponding to an evaluation of the most significant bit plane (MSB). All the pixel values above that threshold are considered as significant. They are coded either with a P symbol in the case of a positive pixel value, either with an N symbol for a negative value. Additionally, the pixel is appended in the significance list. The non-significant ones are further examined to check whether all the corresponding children are 0 too, and if so, they are all coded at once by a zero-tree (ZT) symbol; if this is not the case an isolated zero (IZ) is coded. Pixels marked in the zero-tree list do not have to be analysed anymore during the current dominant pass. Shapiro also proposed to use a zero symbol (Z) for a zero-value pixel at the subbands of the level one, since they do not have children and cannot be zero-trees or isolated zeros in the explicit sense. However, a statistical evaluation from our side indicated that it does not reduce the coding performance to treat these zeros as zero-trees. In most of the cases it even enhances the compression results, especially for lower bit rates.

After the dominant pass, and reducing the threshold level (dividing it by 2 to go to the next bit plane), a subordinate pass is applied to the image. Only the values of the pixels in the significance list are now coded (to allow refinement of the reconstruction), by applying a two-symbol alphabet: significant (T)rue and not significant (F)alse.

Next, a new dominant pass is applied on the pixels not coded yet as significant by applying the same threshold level as in the preceding subordinate pass. At the beginning of this step the zero-tree list is initialised. The above procedure can be repeated until the last bit plane is coded. The code symbols are transmitted to an arithmetic coder [22]. In contrast to the original method [16], we apply separate models for the entropy coding of the symbols generated in the dominant pass and the subordinate pass, which allows a more efficient variable length encoding.

Progressive transmission capabilities are inherent to the coding scheme, i.e. gradually refining the threshold levels (e.g. coding from the most significant bit plane towards the least significant bit plane) and respecting the order of importance of the subbands (Fig. 3). This way of image scanning rules out the objections against classic schemes (Fig. 2). Remark that the low-pass image is not scanned. In contradiction to Shapiro’s original implementation, our implementation applies a non-wavelet based lossless coding of the average image (JPEG-LS). This is done since the low pass image is of high importance for the image reconstruction quality.

Since the embedded zero-tree wavelet encoding technique proposed by Shapiro utilises scalar quantisation, it partially fails to recognise the intra-subband redundancy. To overcome this shortcoming a progressive vector quantisation based embedded zero-tree
coding was introduced by da Silva [5]. The significance of an image vector – composed out of a set of neighbouring pixels – is consequently evaluated by comparing the magnitude of that vector with a yardstick value, i.e., a vector threshold. This value allows layering the subbands similarly to the bit plane concept in the scalar case. However, the yardsticks do not coincide anymore with the bit plane levels. The vector codebook consists out of a set of normalised lattice-based directional code vectors, allowing successive approximation of the image vector.

4. Bit-allocation of subbands

Given the bit rate, a straightforward implementation of the above-mentioned EZW-methods [5,16] does not lead to an optimal codec in the rate-distortion sense, even if we take into account that the wavelet coefficients are scanned using a path that prioritises the low-frequency coefficients on the high-frequency coefficients for a specific threshold. It appears that this prioritisation protocol is not powerful enough to make a visually optimal reconstruction of the image, for the desired compression ratio. Practice indicates that the subbands do have a different energy and bit range. The energy differences suggest that the amount of the visual information present in the different subbands is not equal. It is therefore advisable to privilege the subbands with the highest energy. Applying subband dependent hard thresholding in a lossy compression scheme allows privileging the subband with the highest energy (Fig. 4).

To find suitable threshold levels for each subband, we have to minimise the quantisation error \( D(b) \) subject to the total bit rate \( R_q(b) \), where the vector \( b \) represents the bit rates allocated to the different subbands [19]:

\[
D(b) = \sum_{k=1}^{M} \alpha_k \omega_k 2^{-2b_k} \sigma_k^2,
\]

\[
R_q(b) = \sum_{k=1}^{M} \alpha_k b_k.
\]

The bit rate is an indication of the number of bits that is being considered for compression, and it does not reflect the effective rate obtained after arithmetic encoding. \( M \) represents the total number of subbands, \( \alpha_k \) is the relative subband size, \( \omega_k \) is the perceptual weighting factor, and \( \sigma_k^2 \) is the subband variance. The latter is a good representative of the subband energy. Remark that while the low-pass subimage at the highest level has a Gaussian distribution, the high-pass images do have a Laplacian distribution. The variance (MSE – mean square error) of the latter can be approx-
imated by the mean absolute error (MAE), allowing a significant reduction of the computational complexity.

Assuming that we want to obtain a certain fixed bit rate \( R_{q,c} \), minimising \( D(b) \) can be solved by applying a method based on Lagrange multipliers:

\[
\sum_{k=1}^{M} \frac{\partial}{\partial b_k} [D(b) + \lambda (R_q(b) - R_{q,c})] = 0.
\]

The differentiation with respect to \( b_k \) delivers

\[
b_k = \frac{1}{2} \log_2 \left( \frac{(2 \ln 2) \omega_k \sigma_k^2}{\lambda} \right).
\]

The fixed bit rate constraint \( R_{q,c} \) poses

\[
\sum_{k=1}^{M} \alpha_k b_k = \frac{1}{2} \sum_{k=1}^{M} \alpha_k \log_2 \left( \frac{(2 \ln 2) \omega_k \sigma_k^2}{\lambda} \right) = R_{q,c}.
\]

This yields then the Lagrange multiplier \( \lambda \):

\[
\lambda = 2 \sum_{k=1}^{M} \alpha_k \log_2 \left[ \frac{(2 \ln 2) \omega_k \sigma_k^2}{\lambda} \right] - 2 R_{q,c}.
\]

Eqs (1) and (2) provide the bit lengths \( b_k \). They are typically used to determine the step sizes of the scalar quantisers, used for building fixed-rate codecs optimal in rate-distortion sense [19]. In our approach, the \( b_k \) values are used to define the lowest bit plane or approximation layer for each subband. For small variances the results can be negative, and therefore they should be truncated to 0. The calculations have then to be repeated for a reduced number of subbands — ignoring the insignificant ones — until all \( b_k \) are bigger or equal to 0, and adjusting the fixed bit rate to

\[
R_{q,c} = \sum_{k=1}^{M} \alpha_k b_k.
\]

The perceptual weighting factors \( \omega_k \) are balancing the different subbands in such a way that the visual perception of the image is optimised. Generally, a dyadic visual weighing is applied for an optimal visual performance:

\[
\omega = (\ldots 3, 2, 2, 2, 2, 1, 2, 1, 2, 0)^t.
\]

However, to determine an efficient visual weighing vector we should take into account the characteristics of the visualisation device.

5. Improved progressive EZW encoding

The scalability issue requires that the coded data stream should be interruptible at any stage and still deliver at each breakpoint a good trade-off between reconstruction quality and compression ratio. One can objectionably argue that the above hard-thresholding technique meets this requirement. The ideal scanning curve through the wavelet data should be associated to a specific thresholding (yardstick) pattern, and retain the intrinsic capacity to continue coding when the transmission channel is not saturated or when devices with different resolutions are mounted to the communication channel.

Although, our solution is valid for both scalar and vector quantisation, from now on, we will use the terminology associated to the successive vector approximation approach [5]. The proposed solution considers the coding bounds between which the respective yardsticks have to progress. A separate yardstick is defined for each subband. The lower bounds of the yardstick variables are determined by the maximum required image quality and registered into the lower bound vector (Fig. 5).

An initial bound vector defines the starting values of the yardstick vector. This vector should be parallel with the lower bound vector, ensuring that the coding process respects the subband hierarchy determined by the subband bit-allocation process. Coding of a specific subband ends at the moment the yardstick vector element equals the lower bound vector element.

It is important to remark that the parent–child relations are evaluated at the respective yardstick levels for the different subbands (Fig. 5). This means that the correlation is exploited between a parent at a certain yardstick value and a set of children at another yardstick value. Practice indicates that the compression quality is not reduced due to this re-adjusted scanning path. If we consider the probabilities for occurrence of zeros for the different yardstick levels, analysis shows that the likelihood of zero occurrence does increase with increasing yardstick level.
Additionally, an upper coding bound is introduced by considering the dynamic ranges of the subimages (Fig. 6). This allows reducing the computational load. A subband is encoded if the yardstick vector element for the specified subband is smaller than or equal to the upper bound vector. If this is not the case, the vector magnitudes of the considered subband are treated as non-significant. Remark that it would be a bad strategy to equalise the initial subband yardsticks to the upper bounds, since the coding process would then neglect the energy hierarchy of the different subimages. This would result in a lower probability of zero-tree occurrence.

Looking carefully to the proposed method, we notice a remarkable resemblance with the soft-thresholding technique. This technique applies as preprocessing step to the image the following function:

$$p_k(i, j) = \left\lfloor \frac{p_k(i, j)}{T_k} \right\rfloor,$$

with $p_k(i, j)$ the wavelet coefficient on position $(i, j)$ in subband $k$ and $T_k$ the respective threshold level. The principle is illustrated in Fig. 7: (a) representing the hard-thresholding principle, and (b) the corresponding soft-thresholding output.

If we compare the soft-thresholding scanning curve (Fig. 7(b)) with the improved scanning curve (Fig. 5), which is obtained for the same compression situation, it is clear that both curves treat the subband data with identical priority. Nevertheless, the improved progressive transmission is an embedded technique, while soft thresholding requires a pre-processing stage. Additionally, the proposed technique supports lossy-to-lossless compression, while soft thresholding is a lossy compression technique.

It is important to remark also that the original hard-thresholding technique obtains an optimal coding performance when reaching the desired prefixed bit rate. However, when the bit stream is interrupted earlier, this requirement is not met, while the improved scanning curve will still satisfy this prerequisite. This is illustrated in Fig. 8, where the classic scalar EZW is compared to the improved technique for lossy-to-lossless coding. We note that our technique results in an average gain of 2 dB for the image of the meniscus. Also in the lossless stage one obtains better results with the adapted scanning (for “meniscus” approximately 5% compression gain). To illustrate the progressiveness in quality of the compression method,
6. Region-of-interest (ROI) coding

To determine a generalised framework capable of handling ROIs, we do have to consider the requirements mentioned in Section 1. Basically, we distinguish two situations: (1) the ROI is determined before/during encoding and (2) the ROI is selected after encoding. Both criteria put specific constraints on how the bit stream should be organised. The former requires markers in the bit stream that specify the location and the shape of ROI, the latter imposes the random bit stream access paradigm that is not as easily guaranteed as marker insertion. Typically, random bit stream access is supported through image tiling. Nevertheless, inferior compression results might be the result if small tiles are selected. Indeed, tiling enhances the random access functionality, but introduces blocking artefacts. Still, tiling is necessary to enable the hardware to process large images, and techniques are available to mask the disturbing effects at the tile.
Fig. 9. Progressive transmission with the improved scalar EZW coder of the image "meniscus", displayed for different bit rates (bpp) and corresponding PSNR (dB) values.
borders (e.g. [8]). However, tile sizes should be large to avoid as much as possible the visually disturbing blocking artefacts. If ROI selection takes place after encoding, still a decoding phase is necessary to select the tiles-of-interest (TOI), and re-encode the information belonging to the ROI. This exercise of mind guides us indirectly to limit our analysis to situations of the ROI selection during or before the encoding. Within this paper, we propose a methodology to handle the interactive selection of ROIs during the encoding process. Hence we will not consider tiling aspects, since these are only of interest if the bit stream was already coded. We assume that the data input are images, and do ignore the fact whether they are full-size images or a selection of TOIs.

When wavelet-based image compression is considered, two main approaches need to be discussed: (1) the ROI is defined and coded in the spatial domain and (2) the ROI is defined and treated in the wavelet domain. The spatial approach has the advantage that the ROI information coincides with the physical perception of the ROI. This approach works well if one is only interested in the ROI information without having to visualise the other image parts. If this is not the case, and it is seldom for an interactive implementation, computational complexity explodes. Let us explain this by using the simple example already illustrated in Section 1. Assuming that in a first stage the first refinement layer of the image was transmitted without prioritisation of a spatial location. After a certain degree of refinement the radiologist decides to select an ROI. To be able to transmit the ROI without causing retransmission of ROI data, we have to perform the inverse wavelet transform on the remaining wavelet coefficient data (i.e. remaining refinement levels). Next, in the spatial domain we select the ROI data and code this up to the requested refinement level. Afterwards, a further refinement of the background data might be requested. It should be clear that going for the spatial approach introduces a large overhead on computational load. In the above example, we did not even consider the case of parallel multiple ROI coding whereas the refinement levels do differ.

Hence it is obvious that one has to consider the solution dealing with ROIs within the wavelet domain [2,14,17,18,20,24]. The link with the related spatial location exists but it is not straightforward, due to the properties of the wavelet transform and the wavelet filters. The dyadic decomposition of the wavelet transform spreads spatially related coefficients over different subbands (Fig. 10(a)). Additionally, if a lossless reconstruction of the complete ROI is required, neighbouring wavelet coefficients are to be considered too (Fig. 10(b)). The extension of the spatially related ROI area is depending, as shown in Fig. 10(b), on the width of the low-pass and high-pass synthesis filters. Analysing the filter behaviour allows generating a bitmap to mark wavelet coefficients relevant for the considered ROI [2,14]. Applying a region-growing operator to the original mask, whereas the operator conforms to the filter size parameters, is an easy and efficient way to generate the lossless mask. Additionally, the mask can be improved in combination with weighing the subbands based on the human visual perception [18]. The complement of the ROI map marks the coefficients to be coded if the background is refined after encoding the ROI. Both encoder and decoder are able to construct this map, if both possess the information related to the position and the shape of the ROI together with the wavelet filter size parameter. For the encoder this information is trivial. However, the decoder has to be informed in order to locate the received information correctly. The above procedure proves that tracking both the refinement levels of ROI and background allows elegant image coding without imposing the need for back and forth wavelet transforming.

Nevertheless, if multiple ROI support is required with possible co-existence during a certain time-frame, the above procedure has to be extended. Our proposal embodies two parts: (a) the determination of the bit masks and (b) the insertion of markers into the bit mask.
Fig. 11. Image transmission with multiple ROI definition. (a) The different ROI regions are displayed for one subband. The ROIs are numbered in the order they are defined. Since ROI2 and ROI3 overlap, a fourth ROI is generated: ROI4 being the intersection for the former two. The latter ROI is subtracted from ROI2 and ROI3 and all of them are treated as independent regions during transmission. The background contains the complement of the union of all ROIs. (b) The interactive coding is visualised in function of a nominative time axis. The possible co-existence of different ROIs is illustrated.

stream to identify the ROI (or background) and the refinement level. The first (a) was partly discussed in previous paragraphs, though if more ROIs occur it becomes slightly more complicated. As long as the ROIs do not overlap, no conflicts do occur and the background is determined as complement of the union of the ROIs. If, however, ROIs do overlap we propose to define the overlapping regions as independent ROIs (see also [24] where a similar approach is followed for a multi-client telemedicine set-up to allow one ROI to be specified by each client). This is illustrated in Fig. 11(a) where three ROIs are defined. The first ROI1 does not overlap, while ROI2 and ROI3 do overlap in the wavelet domain. Since the ROIs might have different constraints concerning the refinement end-level, the overlapping region ROI4 is treated as a separate ROI and is appointed the finest end-level. Since the ROIs are defined at different moments and sometimes do overlap, we propose to use markers (b) to identify the coded information sent to the decoder. For each refinement step, the marker should specify the ROI label. Additionally, if an ROI is initiated, also position and shape information should accompany the identification label. Since based on this, the decoder can reconstruct exactly the ROI mask, and it can easily track the refinement level for each ROI, full flexibility is delivered to the encoder to schedule the transmission of the image, making use of the different ROIs. If only rectangular and circular regions are considered, we propose that the decoder checks for overlapping regions too and generates new ROIs using the same protocol as the encoder to stay synchronised, to avoid the shape encoding of complex ROIs. Fig. 11(b) illustrates a generic case where the encoder starts transmitting the first refinement level (background). After this, the client specifies ROI1 and the involved masks for the ROI and background are, respectively, generated and adapted. When ROI1 reaches refinement level 4 a non-overlapping ROI2 is initiated and further refined together with ROI1. At the specification of ROI3, a collision is detected with ROI2 and the overlapping region is identified as a separate ROI4. All four ROIs are further refined until the client indicates he is no longer interested in ROI2. Consequently, transmission of ROI2 is interrupted, and also the refinement of ROI4 since the region should now follow the refinement of ROI3 and ROI4 is slightly ahead in terms of refinement compared to ROI5. Next, ROI1 reaches its maximal precision level and its coding is stopped. When ROI3 reaches level 6, also ROI4 is refined again up to the maximal precision level. Since all ROIs are now coded to their desired refinement level, the remaining information (the background data starting at level 2, and the ROI2 coefficients starting at level 6) is transmitted. Fig. 12 illustrates the selection of one or multiple ROIs on the meniscus image.

It is obvious that the above transmission scheme is very flexible, and that the complexity of scheduling is maximally shifted towards the encoder. The decoder only has to interpret the markers and correctly locate the refinement data for the wavelet coefficients. The proposed transmission scheme only slightly decreases the compression efficiency in terms of compression ratio and related reconstruction quality. The main cause is that entropy coding is performed for each ROI separately, causing reductions in efficiency at the ROI bor-
ders. Secondly, embedding markers in the bit stream introduces overhead information although this is reduced to a strict minimum.

7. Conclusions

Within this paper two generic schemes were proposed for wavelet-based image coding. The first scheme improves the classic progressive schemes, since it combines both the progressive and the thresholding concepts without paying off on the reconstruction quality and compression ratio parameters. In the second part of the paper we describe a protocol to interactively handle multiple ROI coding. Both methodologies are applicable to a large set of state-of-the-art wavelet-based compression techniques and can contribute to the efficiency of telemedicine image processing applications.

Acknowledgements

The authors would like to thank for the support given by the IWT-project: IT-TeleVision (ITA/950202).

References


Peter Schelkens, born in Wilsebroek, Belgium in 1969, obtained the degree of Electrotechnical Industrial Engineer in 1991 (Industrial College IHAM Mechelen), Electrotechnical Civil Engineer in 1994 (University of Brussels – VUB) and Medical Physicist in 1995 (VUB). Since 1994, he is working as assistant in the Department of Electronics and Information Processing at the University of Brussels (VUB), Belgium. He is associated as visiting researcher to IMEC (Interuniversity Microelectronics Center) in Leuven, Belgium, where he is active in the VLSI Systems Design Methodology (VSDM) division, as a member of the Multimedia and Image Compression Systems (MICS) group. His PhD research concerns wavelet-based image and video compression, and aspects as memory, power and speed optimisation for image processing algorithms on different types of hardware platforms. Earlier research concerned medical imaging, where he studied motion compensation in cardiac MR imaging. He is a member of the ISO/IEC JTC1/SC29/WG1 (JPEG2000) standardisation committee.

Adrian Munteanu, born in Constanta, Romania in 1970, obtained his Bachelor degree in Electronics and Telecommunications from the “Politehnica” University of Bucharest in 1994. He received the M.Sc. degree in Biomedical Engineering from the Technical University of Patras in 1996. Since 1996 he is working as a PhD student in the Department of Electronics and Information Processing at the University of Brussels (VUB), Belgium. His main fields of interest are the compression of images using the wavelet transform and multiscale approaches for edge detection in medical images.

Jan Cornelis, born in Wilrijk, Belgium, in 1950, obtained the degree of Mechanical & Electrotechnical Civil Engineer (1973 at the VUB) and the PhD in Applied Sciences (1980 at the VUB). Since then he has been working on several aspects of signal and image processing. He is currently Professor in Electronics and Digital Image Processing at the Faculty of Applied Sciences (VUB). He is director of the research group VUB-IRIS, member of the Board of Directors of IMEC (Interuniversity Microelectronics Center – Leuven, Belgium) and co-ordinator of the “Image Processing Systems” research community (FWO – National Foundation of Scientific Research). His current main research emphasis is on image processing and machine vision – with applications in medicine, remote sensing, anti-personnel mine detection and motion analysis.