Future Generation Computer Systems 15 (1999) 223–243

Wavelet-based medical image compression

Eleftherios Kofidis[1,*], Nicholas Kolokotronis, Aliki Vassilarakou, Sergios Theodoridis, Dionisis Cavouras

Department of Informatics, Division of Communications and Signal Processing, University of Athens, Panepistimioupolis, TYPA Buildings, GR-15784 Athens, Greece

Abstract

In view of the increasingly important role played by digital medical imaging in modern health care and the consequent blow up in the amount of image data that have to be economically stored and/or transmitted, the need for the development of image compression systems that combine high compression performance and preservation of critical information is ever growing. A powerful compression scheme that is based on the state-of-the-art in wavelet-based compression is presented in this paper. Compression is achieved via efficient encoding of wavelet zerotrees (with the embedded zerotree wavelet (EZW) algorithm) and subsequent entropy coding. The performance of the basic version of EZW is improved upon by a simple, yet effective, way of a more accurate estimation of the centroids of the quantization intervals, at a negligible cost in side information. Regarding the entropy coding stage, a novel RLE-based coder is proposed that proves to be much simpler and faster yet only slightly worse than context-dependent adaptive arithmetic coding. A useful and flexible compromise between the need for high compression and the requirement for preservation of selected regions of interest is provided through two intelligent, yet simple, ways of achieving the so-called selective compression. The use of the lifting scheme in achieving compression that is guaranteed to be lossless in the presence of numerical inaccuracies is being investigated with interesting preliminary results. Experimental results are presented that verify the superiority of our scheme over conventional block transform coding techniques (JPEG) with respect to both objective and subjective criteria. The high potential of our scheme for progressive transmission, where the regions of interest are given the highest priority, is also demonstrated.

Keywords: Medical imaging; Image compression; Wavelet transform; Lifting; Reversible transforms; Wavelet zerotrees; Entropy coding; Selective compression; Progressive transmission

1. Introduction

Medical imaging (MI) [20,37] plays a major role in contemporary health care, both as a tool in primary diagnosis and as a guide for surgical and therapeutic procedures. The trend in MI is increasingly digital. This is mainly motivated by the advantages of digital storage and communication technology. Digital data can be easily archived, stored and retrieved quickly and reliably, and used in more than one locations at a time. Furthermore digital data do not suffer from aging and moreover are suited to image postprocessing operations [11,57]. The rapid evolution in both research and clinical practice of picture archiving and communication systems (PACS) [22,40] and digital teleradiology [18,27] applications, along with the existence

* Corresponding author.

1 Present address: Département Signal et Image, Institut National des Télécommunications, 9 rue Charles Fourier, F-91011 Évry cedex, France.

0167-739X/99/$ – see front matter
PII: S0167-739X(98)00066-1

of standards for the exchange of digital information (e.g., ACR/NEMA [2,36], DICOM [17]) further promote digital MI.

However, the volume of image data generated and maintained in a computerized radiology department, by far exceeds the capacity of even the most recent developments in digital storage and communication media. We are talking about Terabytes of data produced every year in a medium-sized hospital. Moreover, even if the current disk and communication technologies suffice to cover the needs for the time being, the cost is too large to be afforded by every medical care unit and it cannot provide a guarantee for the near future in view of the increasing use of digital MI. *Image compression* comes up as a particularly cost-effective solution to the data management problem by reducing the size of the images for economic representation and transmission, keeping at the same time as much of the diagnostic relevant information as possible. The unique nature of medical imagery, namely, the high importance of the information it contains along with the need for its fast and reliable processing and transmission, imposes particularly stringent requirements for an image compression scheme to be acceptable for this kind of application:

– No loss of relevant information is allowed, though the term ‘relevant’ can take different meanings depending on the specific task. At present, the use of ‘lossy’ compression, i.e. with the reconstructed image being different from the original, is very limited in clinical practice, especially in primary diagnosis, due to the skepticism of the physicians about even the slightest data loss, which might, even in theory, induce some critical information loss. However, efficient compression schemes have been derived which have shown satisfying behavior in preserving features (e.g., edges) that are critical in medical diagnosis and interpretation [1,11,13,33,46]. Moreover, there is a constant increase in the production of lossy imaging systems, though at a lower percent as compared to the lossless ones, as reported in [69].
– Progressive coding is a very desirable feature in a medical compression scheme, since it allows the transmission of large images through low-capacity links with the image being gradually built in the receiving display workstation [67]. The physician can, for example, browse over a set of images belonging to a patient and decide on the basis of rough (small-sized and/or low-contrast) approximations of the images on which of them are useful at that time without having to wait for the reception of the whole image set.
– Compression has to be fast enough to accommodate the significant burden imposed on a radiology department. Higher emphasis has to be given on the speed of decompression; the time between the retrieval command and the display on the screen should not exceed 2 s [69]. Although this time constraint is too stringent to be met by an implementation in a general-purpose computer system, there is however an imperative need for as fast decompression as possible.
– There are cases where the diagnosis can be based on one or more regions of the image. In such cases, there must be a way of *selective compression*, that is, allocating more bits to the regions of interest so as to maintain a maximum fidelity whereas the rest of the image can be subjected to a higher compression, representing only a rough approximation of the structures surrounding the selected portions. This can considerably reduce the image size for storage and/or transmission, without sacrificing critical information.

Compression techniques based on the *wavelet decomposition* of the image have received much attention in the recent literature on (medical) image compression and can meet to a large extent the requirements imposed by the application. This is mainly due to the unique ability of the wavelet transform to represent the image in such a way that high compression is allowed preserving at the same time fine details of paramount importance. Fast algorithms, based on well-known digital signal processing tools, have been derived for the computation of such decompositions, leading to efficient software and hardware implementations which add to the practical value of these methods.

In this paper, we propose using a compression scheme composed of a wavelet decomposition in conjuction with a modern quantization algorithm and a lossless entropy encoder. The quantization module is based on the embedded zerotree wavelet (EZW) algorithmic scheme, which constitutes the state-of-the-art in wavelet-based compression. One of its main features is the exploitation of the characteristics of the wavelet representation to provide a sequence of embedded compressed versions with increasing fidelity and resolution, thus leading to a particularly efficient solution to the problem of progressive transmission (PT). A simple yet effective way of exploiting the generally nonuniform probability distribution of the significant wavelet coefficients is proposed and shown to lead to improved quality of reconstruction at a negligible side-information cost. Our scheme allows both lossy and lossless (i.e., exactly reversible) compression, *for the whole of the image or for selected regions of interest.* By exploiting the priority given to the large magnitude coefficients by the EZW algorithm, we provide the possibility of protecting selected regions of interest from being distorted in the compression process, via a perfectly reversible (lossless) or a more accurate than their context lossy representation.

A recently introduced wavelet implementation (lifting) that combines speed and numerical robustness is also investigated in both the general compression context, as well as in the lossless region of interest representation, where it proves to be an efficient means of decorrelation that in combination with sophisticated entropy coding yields high compression performance.

In the entropy coding stage, we have employed arithmetic coding (AC) in view of its optimal performance. Moreover, the small size of the alphabet representing EZW output makes adaptive AC well-suited for this application. However, the high complexity of AC might not be affordable by a simple and fast implementation. A much simpler and more efficient lossless coder that exploits, via run-length encoding (RLE), the occurrence of long sequences of insignificant wavelet coefficients in natural images, is proposed here and shown to be only slightly worse than context-based adaptive AC.

The rest of this paper is organized as follows. Section 2 contains a short presentation of the general framework for image compression including measures of compression performance and criteria of evaluating quality in reconstructed images. The technical details of the compression module we propose are given in Section 3. The wavelet transform is compared against the more commonly used DCT, underlying the Joint Photographic Experts Group (JPEG) standard, and is proposed as a viable alternative which meets the stringent needs of medical image management. In addition to the classical wavelet decomposition, the lifting realization of the wavelet transform is also investigated and claimed to be of particular interest in the present context. The quantization algorithm is presented in detail and its advantages for these applications are brought out. The various choices for the selection of the final stage (i.e. entropy encoder) are also discussed, with emphasis on the proposed RLE-based coder. Two different methods for selective compression are discussed and a novel application of wavelets to lossless compression is presented. Section 4 contains experimental results on a real medical image, that demonstrate the superiority of the proposed approach over JPEG with respect to both objective and subjective criteria. Conclusions are drawn in Section 5.

2. Image compression: Framework and performance evaluation

Image compression [24,25] aims at removing or at least reducing the redundancy present in the original image representation. In real world images, there is usually an amount of correlation among nearby pixels which can be taken advantage of to get a more economical representation. The degree of compression is usually measured with the so-called *compression ratio (CR)*, i.e., the ratio of the size of the original image over the size of the compressed one in bytes. For example, if an image with (contrast) resolution of 8 bits/pixel (8 bpp) is converted to a 1 bpp representation, the compression ratio will be 8:1. The average number of bits per pixel (bpp) is referred to as the *bit rate* of the image. One can categorize the image compression schemes into two types:

(i) *Lossless.* These methods (also called *reversible*) reduce the inter-pixel correlation to the degree that the original image can be exactly reconstructed from its compressed version. It is this class of techniques that enjoys wide acceptance in the radiology community, since it ensures that no data loss will accompany the compression/expansion process. However, although the attainable compression ratio depends on the modality, lossless techniques cannot give compression ratios larger that 2:1 to 4:1 [48].

(ii) *Lossy.* The compression achieved via lossless schemes is often inadequate to cope with the volume of image data involved. Thus, lossy schemes (also called *irreversible*) have to be employed, which aim at obtaining a more compact representation of the image at the cost of some data loss, which however might not correspond to an equal amount of information loss. In other words, although the original image cannot be fully reconstructed, the degradation that it has undergone is not visible by a human observer for the purposes of the specific task. Compression ratios achieved through lossy compression range from 4:1 to 100:1 or even higher.

In general terms, one can describe an image compression system as a cascade of one or more of the following stages [12,70]:

– *Transformation.* A suitable transformation is applied to the image with the aim of converting it into a different domain where the compression will be easier. Another way of viewing this is via a change in the basis images composing the original. In the transform domain, correlation and entropy can be lower, and the energy can be concentrated in a small portion of the transformed image.
– *Quantization.* This is the stage that is mostly responsible for the 'lossy' character of the system. It entails a reduction in the number of bits used to represent the pixels of the transformed image (also called transform coefficients). Coefficients of low contribution to the total energy or the visual appearance of the image are coarsely quantized (represented with a small number of bits) or even discarded, whereas more significant coefficients are subjected to a finer quantization. Usually, the quantized values are represented via some indices to a set of quantizer levels (codebook).
– *Entropy coding (lossless).* Further compression is achieved with the aid of some entropy coding scheme where the nonuniform distribution of the symbols in the quantization result is exploited so as to assign fewer bits to the most likely symbols and more bits to unlikely ones. This results in a size reduction of the resulting bit-stream on the average. The conversion that takes place at this stage is lossless, that is, it can be perfectly cancelled.

The above process is followed in the encoding (compression) part of the coder/decoder (codec) system. In the decoding (expansion/decompression) part, the same steps are taken in reverse. That is, the compressed bit-stream is entropy decoded yielding the quantized transform coefficients, then 'de-quantized' (i.e., substituting the quantized values for the corresponding indices) and finally inverse transformed to arrive at an approximation of the original image. The whole process is shown schematically in Fig. 1.

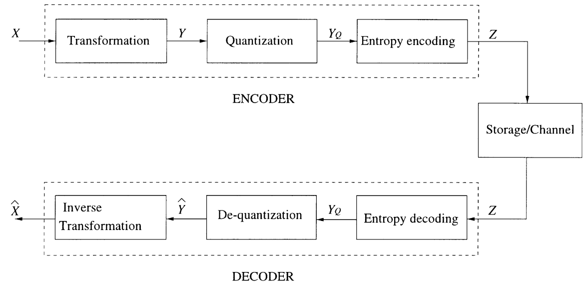


Fig. 1. Block representation of a general transform coding system.

To compare different algorithms of lossy compression several approaches of measuring the loss of quality have been devised. In the MI context, where the ultimate use of an image is its visual assessment and interpretation, subjective and diagnostic evaluation approaches are the most appropriate [12]. However, these are largely dependent on the specific task at hand and moreover they entail costly and time-consuming procedures. In spite of the fact that they are often inadequate in predicting the visual (perceptual) quality of the decompressed image, objective measures are often used since they are easy to compute and are applicable to all kinds of images regardless of the application. In this study we have employed the following two distortion measures:

(i) *Peak signal-to-noise ratio (PSNR)*:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{P^2}{\mathrm{MSE}}\right)\ \mathrm{dB}$$

where

$$\mathrm{MSE} = \frac{1}{MN}\|X - \hat{X}\|_2^2 = \frac{1}{MN}\sum_{i,j}|X_{i,j} - \hat{X}_{i,j}|^2$$

is the *mean squared error* between the original, X, and the reconstructed, $\hat{X}$, $M \times N$ images, and P is the maximum possible value of an element of X (e.g., 255 in an 8 bpp image).

(ii) *Maximum absolute difference (MAD)*

$$\mathrm{MAD} = \|X - \hat{X}\|_\infty = \max_{i,j}|X_{i,j} - \hat{X}_{i,j}|.$$

At this point it is of interest to note that a degradation in the image might also be present even in a lossless transform coding scheme (i.e., where no quantization is present). This is because the transformation process is perfectly invertible only in theory, and if directly implemented, it can be a cause of distortion due to finite precision effects. In this work, an implementation of the wavelet transform that enjoys the perfect reconstruction property in the presence of arithmetic errors is being investigated.

3. The proposed compression scheme

The compression scheme proposed here and shown in the block diagram of Fig. 2 is built around the general process described in Section 2 (Fig. 1). In the sequel, we elaborate on the specific choices for the stages of the general transform-coding process made in our system.

3.1. Transformation

3.1.1. Block-transform coding – discrete cosine transform

Although a large variety of methods have been proposed for medical image compression, including predictive coding, vector quantization approaches, segmentation-based coding schemes, etc. (e.g., [8,11,29,55]) the class of techniques that are based on linear transformations dominates the field [4,7,9,32–34,42,47,51,60,72]. The main advantage of transform coding techniques is the ability of allocating a different number of bits to each transform coefficient so as to emphasize those frequency (or scale) components that contribute more to the way the image is percepted and de-emphasize the less significant components, thus providing an effective way of quantization noise shaping and masking [26].

As noted already, the goal of the transformation step is to decorrelate the input samples and achieve compaction of energy in as few coefficients as possible. It is well known that the optimum transform with respect to both these criteria is the Karhunen–Loève transform (KLT) [26], which however is of limited practical importance due to its dependence on the input signal statistics and the lack of fast algorithms for computing its basis functions. Therefore, suboptimal yet computationally attractive transforms are often used in practice, with the *discrete cosine transform (DCT)* being the most prominent member of this class. DCT owes its wide acceptance to its near-optimal performance (it approaches KLT for exponentially correlated signals with a correlation coefficient approaching one, a model often used to represent real world images) and the existence of fast algorithms for its computation, stemming from its close relationship to the discrete Fourier transform (DFT) [31,44]. Moreover, 2D-DCT is separable, that is, it can be separated into two 1D DCTs, one for the rows and one for the columns of the image (row–column scheme), a feature that further contributes to its simplicity.

The usual practice of DCT-based coding (that used also in the JPEG standard for still continuous-tone image compression baseline mode) is to divide the image into small blocks, usually of 8×8 or 16×16 pixels, and transform each block separately. This block-transform approach is followed mainly for computational complexity and memory reasons. Moreover, it allows adaptation of the spectral analysis action of the transform to the local characteristics of a nonstationary image. However, in low bit-rate compression, where several of the high-frequency DCT coefficients are discarded or coarsely quantized, this approach leads to annoying blocking artifacts in the reconstructed image [31,64]. This is due to the poor frequency localization of DCT, that is, the dispersion of spatially short active areas (e.g., edges) in a large number of coefficients of low energy [69]. This interdependency among adjacent blocks can be alleviated by simply applying the DCT to the whole of the image, considering it as a single block. This so-called full-frame DCT (FFDCT) has been widely applied in the medical imaging area (e.g., [9,32]), where the slightest blocking effect is deemed unacceptable. However, in view of the high spatial and contrast resolution of most of the medical images [28], the computational and storage requirements of such a scheme prevent its implementation in general-purpose software or hardware systems, requir-

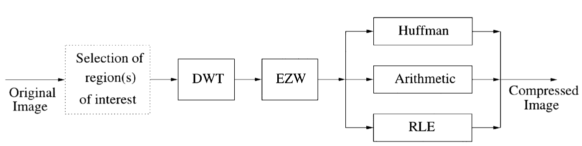


Fig. 2. Block diagram of the proposed coder. Details on selective compression are omitted.

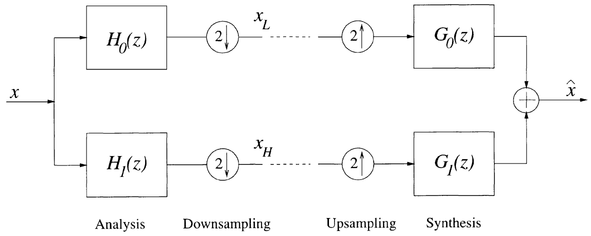


Fig. 3. Two-band multirate analysis/synthesis system.

ing specialized multiprocessor modules with external frame buffers for fast and efficient operation [9].

3.1.2. *The discrete wavelet transform*

A more realistic solution to the blocking problem is offered by *subband coding* schemes [58,61,64], that is, generalized linear transforms where block overlapping is intrinsic in the method. This property has a smoothing effect that mitigates the sharp transition between adjacent blocks. Viewed from the point of view of a multirate maximally decimated filter bank, this is equivalent to the filters (basis functions) being longer than the number of subbands (basis size). This is in contrast to the DCT case, where the inadequate size of the basis vectors results in poor frequency localization of the transform. A multiresolution (multiscale) representation of the image that trades off spatial resolution for frequency resolution is provided by the *discrete wavelet transform (DWT).* The best way to describe DWT is via a filter-bank tree. Fig. 3 depicts the simple 2-band filter bank system for the 1D case. The input signal is filtered through the lowpass and highpass analysis filters H_0 and H_1, respectively, and the outputs are subsampled by a factor of 2, that is, we keep every other sample. This sampling rate alteration is justified by the halving of the bandwith of the original signal. After being quantized and/or entropy coded, the subband signals are combined again to form a full-band signal, by increasing their sampling rate (upsampling by a factor of 2) and filtering with the lowpass and highpass synthesis filters G_0 and G_1 that interpolate the missing samples. It is possible to design the analysis/synthesis filters [2] in such a way that in the absence of quantization of the subband signals, the reconstructed signal coincides with the original, i.e., $\hat{x} = x$. Such analysis/synthesis systems are said to have the *perfect reconstruction (PR)* property.

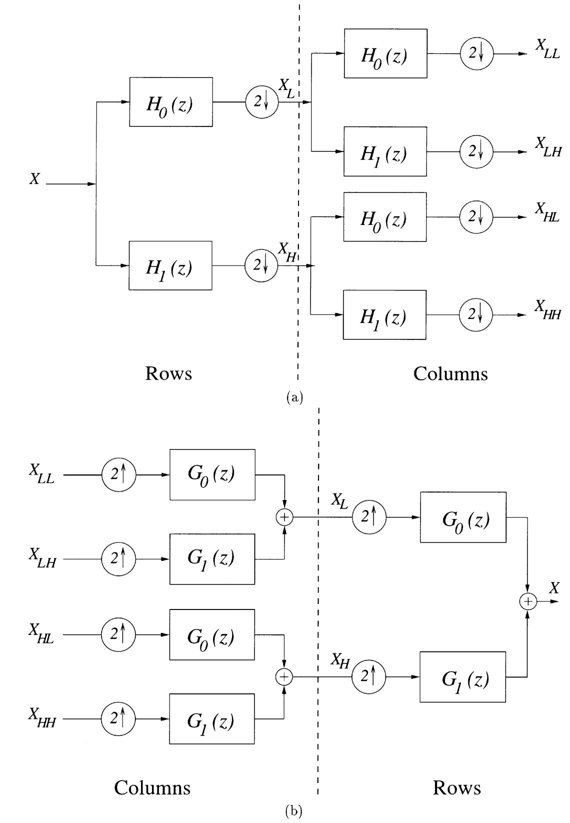


Fig. 4. Separable 2D 4-band filter bank: (a) analysis; (b) synthesis.

A direct extension to the 2D case, which is also the more commonly used, is to decompose separately the image into low and high frequency bands, in the vertical and horizontal frequency [63]. This is achieved via the separable 2D analysis/synthesis system shown in Fig. 4, which ideally results in the decomposition of the frequency spectrum shown in Fig. 5.

The analysis into subbands has a decorrelating effect on the input image. Nevertheless, it might often be necessary to further decorrelate the lowpass subband. This is done via iterating the above scheme on the lowpass signal. Doing this several times leads to a filter bank tree such as that shown in Fig. 6. The result of such a 3-stage decomposition can be seen in Fig. 7. This decomposition of the input spectrum into subbands which are equal in a logarithmic scale is nothing but the DWT representation of the image [45,58,61,64,72].

[2] Here we deal only with FIR filters, that is, functions $H_i(z)$ and $G_i(z)$ are (Laurent) polynomials in z.

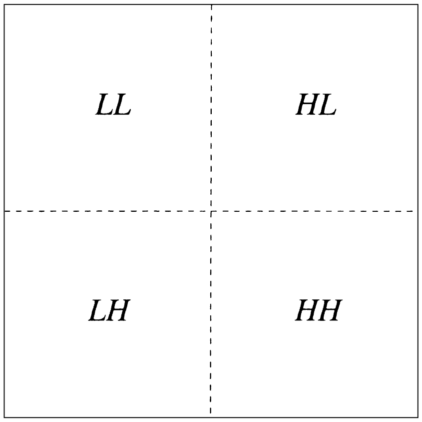


Fig. 5. Ideal decomposition of the 2D spectrum resulting from the filter bank of Fig. 4(a).

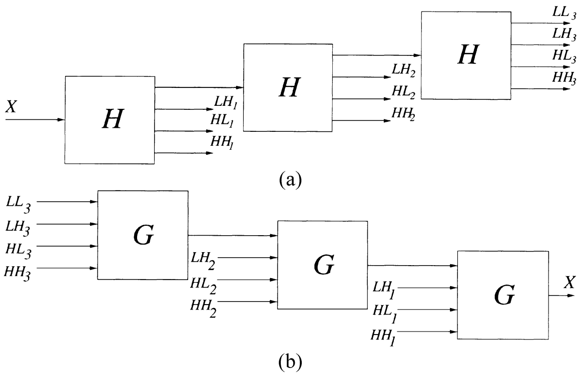


Fig. 6. 2D DWT: (a) forward; (b) inverse. (H-and G-blocks represent the 4-band analysis and synthesis banks of Fig. 4(a) and (b), respectively.)

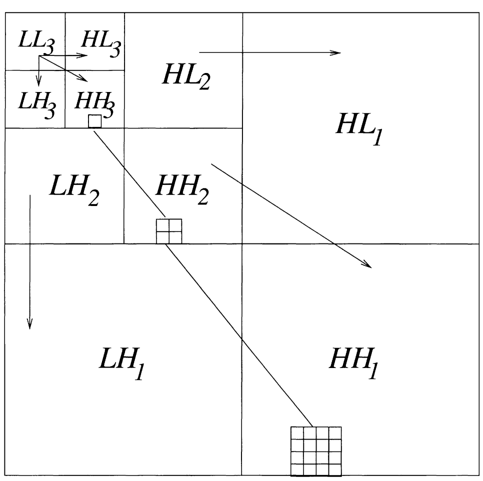


Fig. 7. Wavelet decomposition of the image spectrum (3 levels).

Provided that the 2-band filter bank satisfies the perfect reconstruction condition, the structure of Fig. 6 represents an expansion of the image with the basis functions being given by the impulse responses of the equivalent synthesis filter bank and the coefficients being the outputs of the analysis bank, referred to also as DWT coefficients. In this expansion, the basis functions range from spatially limited ones of high frequency that capture the small details in the image, e.g., edges, lines, to more expanded of low frequency that are matched to larger areas of the image with higher inter-sample correlation. In contrast to what happens in the block-transform approach, a small edge, for example, in the image is reflected in a small set of DWT coefficients, with its size depending on how well localized the particular wavelet function basis is. It is mainly for these good spatial-frequency localization properties that the DWT has drawn the attention of the majority of the MI compression community, as a tool for achieving high-compression with no visible degradation and preservation of small-scale details [34,47,51,60,72].

3.1.3. *The lifting scheme*

As already noted, direct implementations of theoretically invertible transforms usually suffer from inversion errors due to the finite register length of the computer system. Thus, one can verify that even in a double precision implementation of the wavelet tree of Fig. 6, where the 2-band filter bank satisfies the conditions for PR, the input image will not be exactly reconstructed. This shortcoming of the direct DWT scheme gets even worse in a practical image compression application where the wavelet coefficients will first have to be rounded to the nearest integer before proceeding to the next (quantization or entropy encoding) stage. The presence of this kind of error would prove to be an obstacle in the application of the DWT in lossless or even near-lossless compression. In this section we focus on the notion of *lifting*, which represents a generic and efficient solution to the perfect inversion problem [6,10,16,59].

To describe the idea underlying the lifting scheme, we first have to reformulate the analysis/synthesis structure of Fig. 3, translating it into an equivalent, more compact scheme. Decomposing the analysis and synthesis filters into their *polyphase components*:

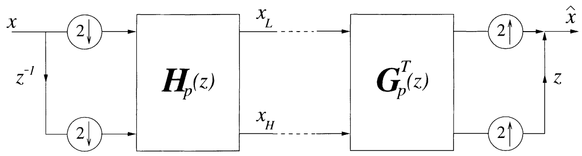


Fig. 8. Polyphase implementation of Fig. 3.

$$H_i(z) = H_{i,0}(z^2) + z^{-1}H_{i,1}(z^2)$$

$$G_i(z) = G_{i,0}(z^2) + zG_{i,1}(z^2),$$

and making use of some well-known multirate identities, one obtains the polyphase structure of Fig. 8 [61]. The *polyphase matrix* $\boldsymbol{H}_p(z)$ is defined as

$$\boldsymbol{H}_p(z) = \begin{bmatrix} H_{0,0}(z) & H_{0,1}(z) \\ H_{1,0}(z) & H_{1,1}(z) \end{bmatrix}$$

with a similar definition for $\boldsymbol{G}_p(z)$. It is apparent from Fig. 8 that PR is ensured if and only if the identity

$$\boldsymbol{G}_p^{\mathrm{T}}(z)\boldsymbol{H}_p(z) = \boldsymbol{I}$$

holds. The idea behind lifting is to *realize these two transfer matrices in such a way that the above identity is preserved regardless of the numerical precision used.*

The term 'lifting' refers to a stepwise enhancement of the band splitting characteristics of a trivial PR filter bank, usually the split–merge filter bank corresponding to the choice $\boldsymbol{H}_p(z) = \boldsymbol{G}_p(z) = \boldsymbol{I}$ in Fig. 8. This is done via multiplication of the analysis polyphase matrix by polynomial matrices that are of a special form: they are triangular with unit diagonal elements. The synthesis polyphase matrix is of course also multiplied by the corresponding inverses which are simply determined by inspection, namely by changing the sign in the nondiagonal elements. The important point with respect to the realization of an analysis/synthesis system with the aid of a lifting scheme is that the polyphase matrix of any PR FIR system, which without loss of generality may be considered as having unity determinant, can be factorized in terms of such elementary factor matrices. It can readily be verified that the effect of rounding-off the multiplication results in the analysis bank of such a structure can be exactly cancelled by an analogous rounding in the synthesis stage, resulting in a realization of the DWT that maps integers to integers with the PR property being structurally guaranteed. It should be noticed that synthesis involves exactly the same operations with the analysis bank, except in the reverse order and with the signs changed.

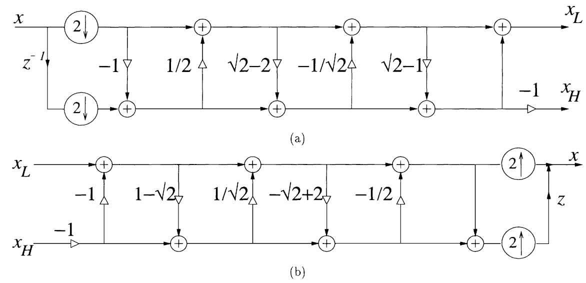


Fig. 9. Lifting realization of the Haar filter bank: (a) analysis; (b) synthesis.

A simple but informative example of the lifting scheme is provided by Fig. 9 illustrating the corresponding realization of the forward and inverse second-order Haar transform:

$$\mathcal{H} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

In addition to its potential for providing a lossless multiresolution of an image, lifting can also yield significant savings in both computational and storage complexity. As shown in [14], for sufficiently long filters, the operations count in the direct wavelet filter bank can drop by about one half when lifting steps are used. Moreover, the triangular form of the factor matrices allows the computations to be performed in place, in contrast to the classical algorithm [59].

3.2. Quantization

The role of the DWT in the compression scheme presented here is not merely that of a decorrelating and energy compacting tool. It also yields a means of revealing the self-similarity [53] and multiscale structure present in a real world image, making efficient encoding and successive approximation of the image information possible and well suited to the needs of progressive transmission. These properties of the DWT are exploited in a very natural manner by the quantization stage employing an EZW scheme.

The EZW algorithm is a particularly effective approach to the following twofold problem: (1) achieving the best image quality for a given compression ratio (bit-rate) and (2) encode the image in such a way that all lower bit rate encodings are embedded at the beginning of the final bit-stream. This embedding property provides an attractive solution to the problem of progressive transmission, especially when transmission errors due to channel noise have also to be dealt with. The basic phases of EZW are very similar to those of the DCT-based coding used in JPEG, that is, first information on which of the coefficients are of significant magnitude is generated and then those coefficients that are significant are encoded via some quantization. However, the multiresolution characteristics of the wavelet representation allow a much more efficient implementation of these two phases. First, the *significance map*, that is, the set of decisions as to whether a DWT coefficient is to be quantized as zero or not, is encoded taking advantage of the self-similarity [3] across scales (subbands) inherent in the wavelet decomposition of natural images and the frequently satisfied hypothesis of a decaying spectrum. The quantization of the coefficients is performed in a successive approximation manner, via a decreasing sequence of thresholds. The first threshold is taken to be half the magnitude of the largest coefficient. A coefficient x is said to be *significant* with respect to a threshold T if $|x| \geq T$, otherwise it is called *insignificant*. Due to the hierarchical nature of the wavelet decomposition, each coefficient (except in the highest frequency subbands) can be seen to be related to a set of coefficients at the next finer scale of similar orientation and corresponding spatial location. In this way, a tree structure can be defined as shown for example in Fig. 7. If the hypothesis stated above is satisfied, it is likely that the descendants of an insignificant coefficient in that tree will also be insignificant with respect to the same threshold. In such a case, this coefficient is termed a *zero-tree root (ZTR)* and this subtree of insignificant coefficients is said to be a *zero-tree* with respect to this threshold. In contrast to related algorithms, e.g. [30], where the insignificance of a tree is decided upon a global statistical quantity, a ZTR here implies that *all* of its descendants are of negligible magnitude with respect to the current threshold, thus preventing statistical insignificance from obscuring isolated significant coefficients. A consequence of this point is that the descendants of a zero-tree root do not have to be encoded; they are *predictably insignificant.* Moreover, the fact that the encoder follows a specific order in scanning the coefficients, starting from low-frequency and proceeding to higher frequency subbands with a specific scanning order within each subband, excludes the necessity of transmission of side position information as it is done for example in threshold DCT coding. In the case that not all descendants are insignificant, this coefficient is encoded as *isolated zero (IZ).* Insignificant coefficients belonging to the highest frequency subbands (i.e., with no children) are encoded with the *zero (Z)* symbol. The symbols *POS* and *NEG*, for positive and negative significant, respectively, are used to encode a significant coefficient.

During the encoding (decoding) two separate lists of coefficients are maintained. The *dominant* list keeps the coordinates of the coefficients that have not yet been found to be significant. The quantized magnitudes of those coefficients that have been found to be significant are kept in the *subordinate* list. For each threshold T_i in the decreasing sequence of thresholds (here $T_i = T_{i-1}/2$) both lists are scanned. In the dominant pass, every time a coefficient in the dominant list is found to be significant, it is added at the end of the subordinate list, after having recorded its sign. Moreover, it is removed from the dominant list, so that it is not examined in the next dominant pass. In the subordinate pass a bit is added to the representations of the magnitudes of the significant coefficients depending on whether they fall in the lower or upper half of the interval $[T_i, T_{i-1})$. The coefficients are added to the subordinate list in order of appearance, and their reconstruction values are refined in that order. Hence the largest coefficients are subjected to finer quantization, enabling embedding and progressive transmission. To ensure this ordering with respect to the magnitudes, a reordering in the subordinate list might be needed from time to time. This reordering is done on the basis of the reconstruction magnitudes known also to the decoder, hence no side information needs to be included. The two lists are scanned alternately as the threshold values decrease until the desired file size (average bit rate) or average distortion has been reached. A notable feature of this algorithm is that the

[3] Interestingly, this property has been recently shown to be closely related to the notion of self-similarity underlying fractal image representation [15].

size or distortion limits set by the application can be exactly met, making it suitable for rate or distortion constrained storage/transmission. [4]

In the decoding stage, the same operations take place, where the subordinate information is used to refine the uncertainty intervals of the coefficients. The reconstruction level for a coefficient can be simply the center of the corresponding quantization interval. However, this implies a uniform probability distribution, assumption which in some cases might be quite inaccurate. Therefore, in our implementation, *we reconstruct the magnitudes on the basis of the average 'trend' of the real coefficients towards the lower or upper boundary of the interval, as estimated in the encoding stage.* This information is conveyed to the decoder in the form of a number in the range (0,1) representing the relative distance of the average coefficient from the lower boundary. This in fact provides a rudimentary estimate of the centroid in each interval for a nonuniform probability density function.

The compressed file contains a header with some information necessary for the decoder such as image dimensions, number of levels in the hierarchical decomposition, starting threshold, and centroid information. We have seen that subtracting the mean from the image before transforming it, and adding it back at the end of the decoding phase yielded some improvement in the quality of the decompressed image. [5] The mean subtraction has the effect of bias reduction (detrend) in the image and moreover, as reported in [62], ensures zero transmittance of the high-pass filters at the zero frequency, thus avoiding the occurrence of artifacts due to insufficiently accurate reconstruction of the lowpass subband. The image mean is also stored in the header of the compressed file.

It must be emphasized that, although the embedding property of this codec may be a cause of suboptimality [19], it nevertheless allows the truncation of the encoding or decoding procedure at any point with the same result that would have been obtained if this point was instead the target rate at the starting of the compression/expansion process [52,53]. Progressive coding with *graceful degradation* as required in medical applications can be benefited very much by the above property. The embedding property of our compression algorithm closely matches the requirements for progressive transmission in fidelity (PFT), i.e., progressive enhancement of the numerical accuracy (contrast resolution) of the image. The sequence of reconstructed images produced by the corresponding sequence of thresholds (bit-planes) with the inherent ordering of importance imposed by EZW on the bits of the wavelet coefficients – precision, magnitude, scale, spatial location, can be used to build an efficient and noise-protected PT system effectively combining PFT and progressive (spatial) resolution transmission (PRT).

3.3. Entropy coding

The symbol stream generated by EZW can be stored/transmitted directly or alternatively can be input to an entropy encoder to achieve further compression with no additional distortion. Compression is achieved by replacing the symbol stream with a sequence of binary codewords, such that the average length of the resulting bit-stream is reduced. The length of a codeword should decrease with the probability of the corresponding symbol. It is well known that the smallest possible number of bits per symbol needed to encode a symbol sequence is given by the *entropy* of the symbol source [35]:

$$H = -\sum_i p_i \log_2 p_i ,$$

where p_i denotes the probability of the ith symbol. In an optimal code, the ith symbol would be represented by $-\log_2 p_i$ bits. *Huffman coding* [23] is the most commonly used technique. This is due to its optimality, that is, it achieves minimum average code length. Moreover, it is rather simple in its design and application. Its main disadvantage is that it assigns an integer number of bits to each symbol, hence it cannot attain the entropy bound unless the probabilities are powers of 2. Thus, even if a symbol has a probability of occurrence 99.9%, it will get at least one bit in the Huffman representation. This can be remedied by using block Huffman coding, that is, grouping the

[4] This characteristic should be contrasted to what happens in JPEG implementations, where the user specification for the amount of compression in terms of the *Quality Factor* cannot in general lead to a specified file size.

[5] An analogous operation takes place in the JPEG standard as well, where an offset is subtracted from the DCT coefficients [66].

symbols into blocks, at the cost of an increase in complexity and decoding delay.

A more efficient method, which can theoretically achieve the entropy lower bound even if it is fractional, is *arithmetic coding (AC)* [21,39,41,69]. In this approach, there is no one-to-one correspondence between symbols and codewords. It is the message, i.e. the sequence of symbols, that is rather assigned a codeword, and not the individual symbols. In this way, a symbol may be represented with less that 1 bit. Although there is a variety of arithmetic coders that have been reported and used, the underlying idea is the same in all of them. Each symbol is assigned a subinterval of the real interval [0,1), equal to its probability in the statistical model of the source. Starting from [0,1), each symbol is coded by narrowing this interval according to its allotted subinterval. The result is a subinterval of [0,1) and any number within it can uniquely identify the message. In practice, the subinterval is refined incrementally, with bits being output as soon as they are known. In addition to this incremental transmission/reception, practical implementations employ integer arithmetic to cope with the high-precision requirements of the coding process [39,69].

AC achieves H bits/symbol provided that the estimation of the symbol probabilities is accurate. The statistical model can be estimated before coding begins or can be adaptive, i.e., computed in the course of the processing of the symbols. An advantage of the adaptive case is that, since the same process is followed in the decoder, there is no need for storing the model description in the compressed stream. One can improve on the simple scheme described above, by employing higher-order models, that is, utilizing conditional probabilities as well [39]. We have used such a coder in our experiments. The model was adaptively estimated for each dominant and subordinate pass, that is, the model was initialized at the start of each phase. This choice was made for two reasons: first, the ability of meeting exactly the target rate is thus preserved, and second, the statistics of the significance symbols and the refinement bits would in general differ.

In spite of its optimality, this method has not received much attention in the data compression area (though there are several reported applications of it in medical image compression [28,49,51]) mainly because of its high complexity which considerably slows down the operation of the overall system. Of course, it should be noted that specialized hardware could be used to tackle this bottleneck problem [41]. Furthermore, approximations to pure AC can be used to speed up the execution at the cost of a slight degradation in compression performance [21,41]. In this work, we have tested a much faster and cost-effective entropy coding technique, based on a rudimentary binary entropy coding of the significance map symbols and an RLE of the ZTR and Z symbols. Based on extended experimentation we have confirmed the fact that the ZTR and Z symbols are much more frequent than the rest of the symbols. Thus, we use a single bit for representing each of these symbols, and a few more for the rest of them. For the lower frequency subbands we use the codes ZTR:0, IZ:10, POS:110, NEG:1110, while the symbols in the highest frequency subbands are coded as Z:0, POS:10, NEG:110. Notice that since the decoder knows exactly in which subband it is found at any time, no ambiguity results from the fact that the same codes are assigned to different symbols in low and high frequency subbands. We use an RLE code to exploit the occurrence of long bursts of zero coefficients. RLE merely represents a long series of consecutive symbols by the length of the series (run-length) and the symbol [39]. The codeword $111\ldots10$ where there are k 1's in total represents a run of $2^{k-1}-1$ zero symbols. To avoid confusion with single-symbol codewords, k is restricted to exceed 2 and 3 for the highest and low subbands, respectively. We have not used this technique on the subordinate list because of the larger fluctuation in the quantization bit sequences. As it will be seen in the evaluation of our experimental results, *the performance of this RLE coder is only slightly worse than that of AC, in spite of its simplicity and computational efficiency.*

3.4. Selective compression

A compromise between the need for high compression imposed by the vast amount of image data that have to be managed and the requirement for preservation of the diagnostically critical information with limited or no data loss, is provided in our compression scheme by the *selective compression (SC)* option. This means that the doctor can interactively select one or more (rectangular) regions of the image that he/she would like to be subjected to less degradation than their context in the compression process. Depending on his/her judgement on the importance of these regions of interest (RoI), he/she can specify that they be losslessly compressed or provide for each a quantitative specification of its usefulness relative to the rest of the image so that it can be more finely quantized.

(i) *Lossy SC.* The relative importance of the RoI as compared to its context is quantified via a weight factor, e.g., 4 or 16. The spatial localization of the DWT is taken advantage of in this case to allow the translation of the spatial coordinates of the RoI to those in the wavelet domain. In other words, we make use of the fact that each pixel depends on only a few wavelet coefficients and vice versa. The RoI is determined at each level by simply halving the coordinates of its representation at the previous (finer scale) level. The corresponding regions in the image subbands are multiplied by the weight factor yielding an amplification of the wavelet coefficients of interest. These same coefficients are divided by this factor in the reconstruction phase to undo the emphasis effect. [6] This trick exploits the high priority given to the coefficients of large magnitude by the EZW algorithm [56]. These coefficients are coded first, so the bit budget is mostly spent on the RoI whereas the rest of the image undergoes a coarser quantization. The larger the weight assigned to a RoI, the better its reconstruction with respect to its context. The amount of degradation suffered from the background image depends on the target bit rate, the number and areas of the RoIs selected. Obviously, if a large number of small RoIs are selected and assigned large weights, the quality of the rest of the image might get too low if a high compression ratio is to be achieved. The selection process must guarantee that the appearance of this part of the image is sufficiently informative to aid the observer appreciate the context of the RoI. The parameters of a successful selection are largely task-dependent and the doctor might have to reevaluate his/her choices in the course of an interactive trial-and-error process.

Our approach of amplifying the wavelet coefficients corresponding to the selected RoI has shown better results than the method proposed by Shapiro [54] where the weighting is applied directly to the spatial domain, that is, before wavelet transforming the image. The latter approach makes use of the linearity of the DWT to translate the weighting to the subband domain. The same reasoning followed in the wavelet-domain method is also adopted here with respect to the prioritization of the large coefficients in the EZW coding. However, the amplification of a portion of the image generates artificial steps which result in annoying ringing effects when filtered by the wavelet filter bank. The introduction of a margin around the selected region to achieve a more gradual amplification with the aid of a smooth window [54] was not sufficient to solve the problem. The visually unpleasing and perhaps misleading borders around the RoI do not show up in the wavelet approach even when no smoothing window is used.

(ii) *Lossless SC.* Although the above method for SC can in principle be used for lossless compression as well, a more accurate and reliable approach for reversibly representing the RoIs has also been developed. It is based on the idea of treating the RoIs in a totally different way than the rest of the image. After the selection has been done, the RoIs are separately kept and input to a lossless encoder while the image is compressed with the normal approach. Prior to its entropy encoding, the RoI can undergo a lowering in its entropy with the aid of a suitable transformation. This operation will normally permit a further compression gain since it will reduce the lower bound for the achievable bit-rate. Extended studies of the performance of several transform-entropy coder combinations for reversible medical image compression have been reported [28,48]. *In this work, we have used the DWT in its lifting representation as the decorrelation method. Apart from the advantages emerging from the hierarchical decomposition implied by this choice, the structural enforcement of the exact reversibility of the lifting scheme naturally meets the requirement for purely lossless compression.* In the expansion stage, the RoIs are 'pasted' onto their orig-

[6] The weight factor along with the coordinates of the RoI are included in the compressed file header.

inal places. To save on the bit-rate for the rest of the image, the corresponding portions are first de-emphasized in the encoding stage by attenuating the corresponding coefficients. This is in effect the exact counterpart of the trick used in the lossy SC approach, where now the selected wavelet coefficients are given the lowest priority in the allocation of the bit-budget. There is also the possibility of performing the de-emphasis in the spatial domain by also introducing a sufficiently wide margin (whose width depends on the length of the filters used) to avoid leakage of ringing effects outside the protected region.

As it will become apparent in the experimental results presented in the next section, this SC approach, despite its conceptual simplicity, is *particularly effective with respect to providing safe representation of critical information and visually pleasing results, without sacrificing compression performance.*

4. Results

This section presents typical results from the application of our compression scheme to real medical imagery, aiming to demonstrate its applicability to the demanding problems inherent in the medical image compression area and its superior performance over that of more conventional approaches in terms of both objective and subjective criteria. The test image used in the experiments presented here is the $512 \times 512 \times 8$ bpp chest X-ray image shown in Fig. 10. The filter banks used in our simulations include the Haar as well as the (9,7) pair developed by Antonini et al. [3]. Both filter banks were ranked among the best ones for image compression in the study of [65]. Both systems enjoy the property of linear phase which is particularly desirable in image processing applications. It has been shown to contribute to a better preservation of the details in low bit-rate compression, apart from overcoming the need for phase compensation in trees of nonlinear-phase filters. Orthogonality, a property of the transform that allows the direct translation of distortion reductions from the transform to the spatial domain in transform coding schemes, is satisfied by both filter pairs, albeit only approximately by the (9,7) system.[7] Despite the simplicity and computational efficiency of the Haar wavelet and its lower susceptibility to ringing distortions as compared to the longer (9,7) filter bank, the latter is deemed more suitable for such applications in view of its better space-frequency localization which battles against the blocking behavior often experienced in Haar-based low-bit rate compression.

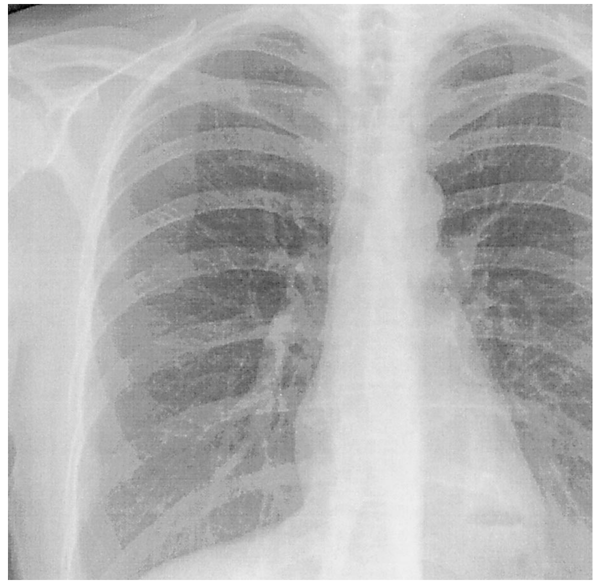


Fig. 10. Test image used in our experiments ($512 \times 512 \times 8$ bpp).

We have tested the performance of our scheme versus the JPEG codec in its Corel 7 implementation. The DWT stage employed a 5-level tree using the (9,7) filter bank. A practical matter comes up when wavelet transforming an image, namely that of coping with the finite input size, i.e., performing the filtering with the appropriate initial conditions that do not harm the PR property. In the experiments discussed here, we used symmetric periodic extension to cope with this problem, that is, the image is extended as if it were viewed through a mirror placed at its boundaries. This choice has the advantage of alleviating distortions resulting from large differences between its opposing boundaries. The direct DWT scheme was used. The results with the lifting scheme are not included since for lossy compression they are comparable to those of the di-

[7] This is in fact a *biorthogonal* analysis/synthesis system, meaning that the analysis filters are orthogonal to the synthesis ones but not to each other.

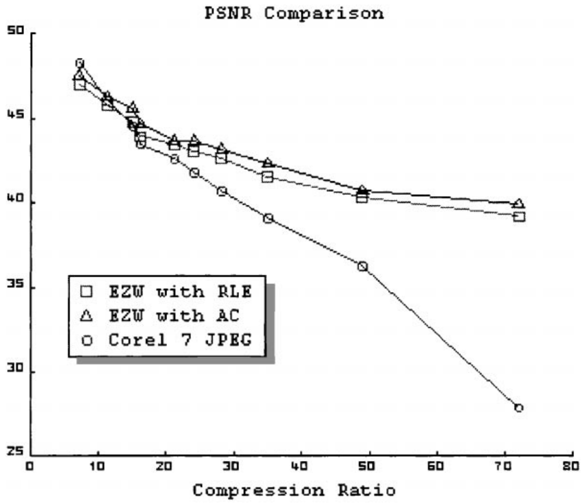


Fig. 11. PSNR comparison of EZW-AC, EZW-RLE, and JPEG.

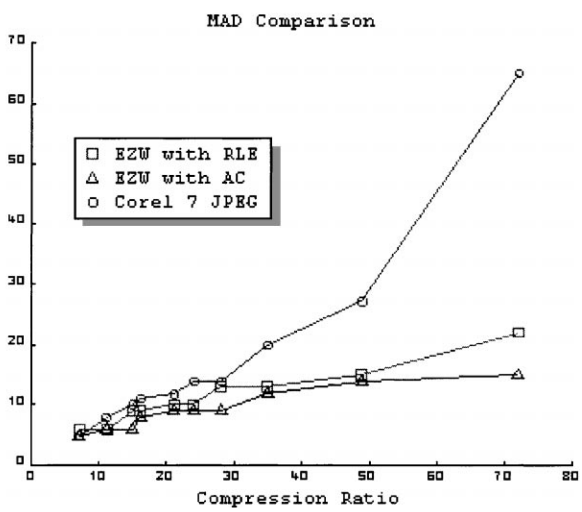
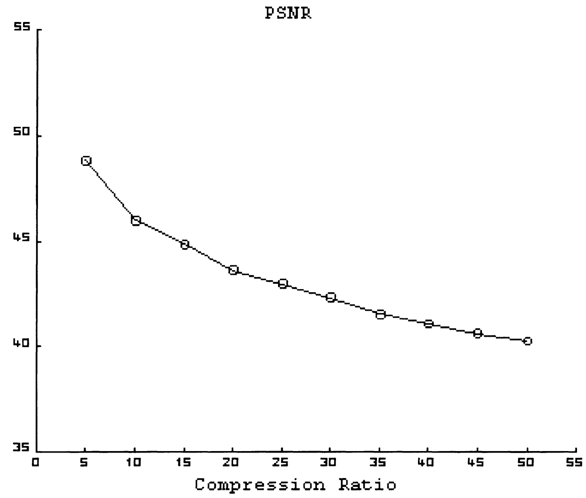
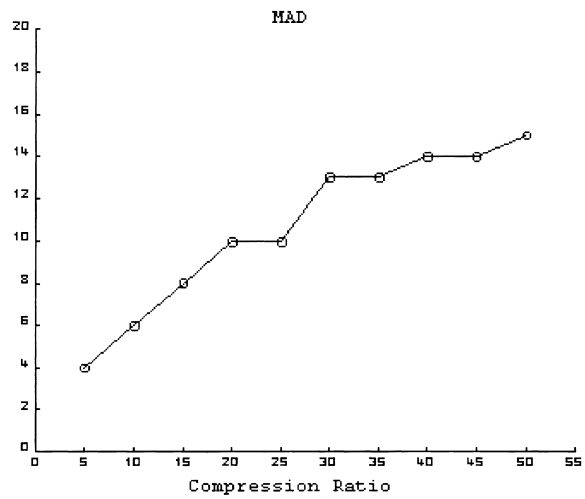


Fig. 12. MAD comparison of EZW-AC, EZW-RLE, and JPEG.



(a)



(b)

Fig. 13. Rate-distortion curves for EZW-RLE: (a) PSNR; (b) MAD.

rect scheme, usually slightly worse especially for high compression ratios. We have tested both the AC and our RLE-based lossless coding algorithms in the entropy coding stage. The resulting PSNR for the three codecs and for a wide range of compression ratios is plotted in Fig. 11. We observe that the curves for the DWT-based schemes are well-above that of JPEG for compression ratios larger that 10:1. Furthermore, the loss from employing the simpler and faster RLE coder in comparison to the much more sophisticated first-order model-based AC is seen to not exceed 1 dB. These results enable us to propose the RLE-based coder as an attractive alternative to more complex entropy coding schemes, since it is proven to combine conceptual and implementational simplicity with high coding efficiency.

The same ranking of the three compression approaches is preserved also in terms of a comparison of the resulting MAD values, as shown in Fig. 12. Moreover, it can be seen that, with respect to this figure of merit, the DWT-based compressor consistently outperforms JPEG, for all compression ratios examined. The results of Fig. 12, albeit reflecting an objective performance evaluation, might also be of use in

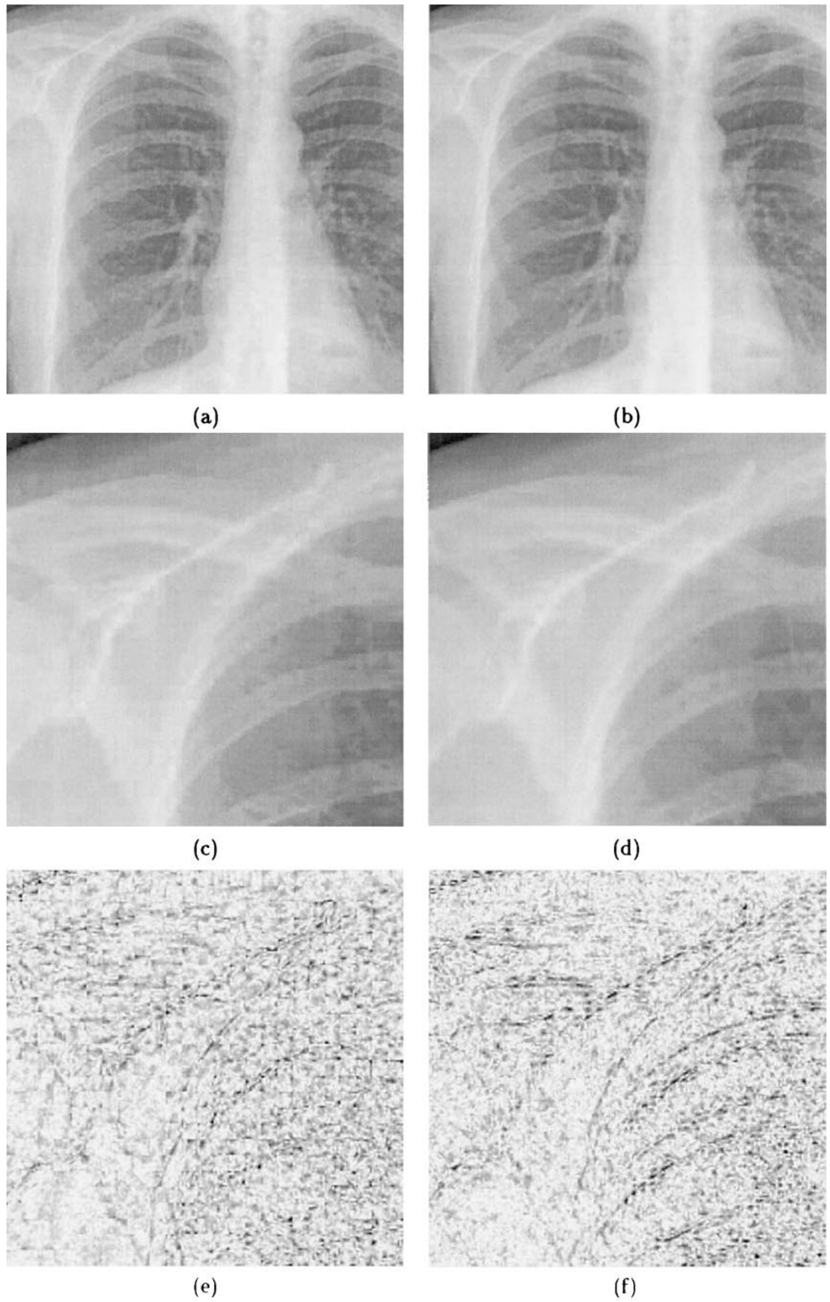


Fig. 14. Results of 32:1 compression with (a) JPEG and (b) EZW-RLE. (c), (d) Details of (a), (b). Corresponding errors are shown in (e), (f).

conjecturing on the subjective visual quality of the reconstructed images [60]. Besides, isolated local errors might prove to be more severe, for the purposes of diagnosis, than distortions that have been spread out on larger areas of the image.

We should note that, as it is evident in both of the above comparisons, the reconstruction quality degradation, albeit increasing in all three algorithms with increasing compression, is far more pronounced in the JPEG case.

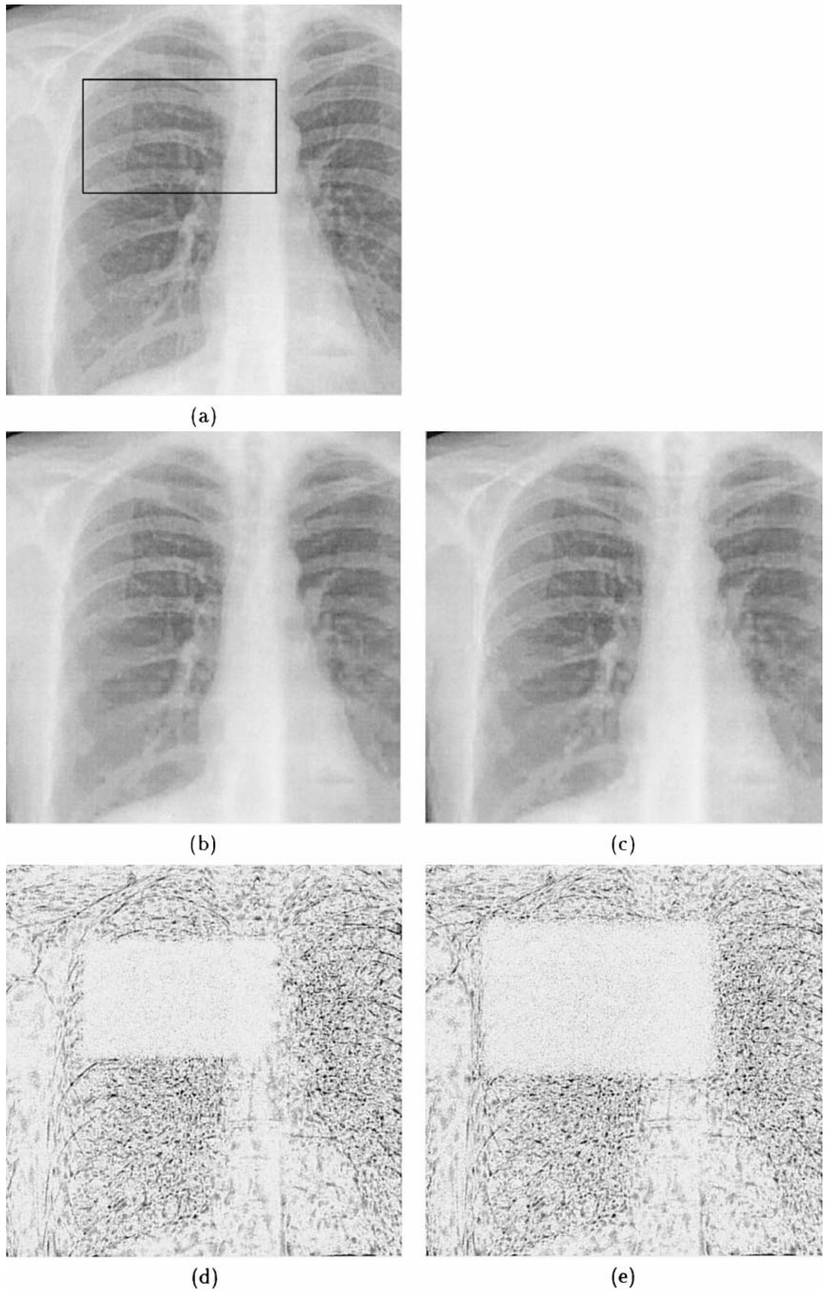


Fig. 15. Lossy SC: weight=16, CR=32:1. (a) RoI shown in border. Reconstructed image using wavelet (b) and spatial (c) domain weighting. (d), (e) Error images for (b), (c).

We have also tested the DWT compressor in a denser set of compression ratios, from 5:1 to 50:1. The results for the PSNR and MAD are plotted in Fig. 13 and should be indicative of the rate-distortion behavior of this scheme.

Fig. 14 shows the compressed/decompressed chest image at a ratio of 32:1 using JPEG and the proposed scheme with 5-level (9,7) DWT and RLE coding. The latter achieves a MAD of 13, and a PSNR of 41.888 dB, as compared to 18 and 39.769 dB for

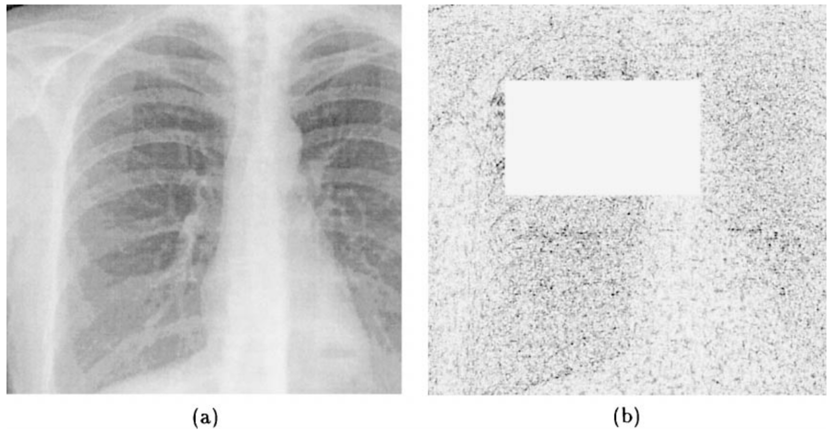


Fig. 16. Lossless SC: (a) reconstructed and (b) error images for the RoI of Fig. 15(a). A CR of 32:1 was specified for the rest of the image.

the MAD and the PSNR furnished by the JPEG coder. Moreover, no visible artifact is present in the DWT-compressed image, whereas the blocking effect is evident in the result of JPEG, as can be observed in the magnified images shown in the same figure. Viewing the corresponding (normalized to the range [0,255]) error images further reveals the prevalence of the blocking effect in the JPEG image.

The SC capability of our system has also been tested in both its lossy and lossless modes. The rectangular region shown with the black border in Fig. 15(a) is the RoI selected for this experiment. For the test of the lossy SC, a weight factor of 16 was chosen and the target rate was specified to correspond to a compression ratio of 32:1. Fig. 15(b,c) shows the reconstructed images where amplification has been performed in the wavelet and spatial domains, respectively. In the latter case, a raised-cosine window has been employed to smooth-out the sharp transition generated. In spite of the smoothing used, the image in Fig. 15(c) exhibits an annoying ringing effect around the RoI, in contrast to the satisfying appearance of the region and its context in Fig. 15(b). The error images shown in Fig. 15(d,e) are also informative on the comparison of the two approaches.

The result of the application of the lossless SC technique to the same RoI is shown in Fig. 16. A compression ratio of 32:1 was again used for the background of the image. Notice that, contrary to what one would expect from such a simple 'cut-and-paste' technique, the result is a visually pleasant image showing no bordering effects around the selected region. In the example of Fig. 16 we used an AC to losslessly compress the RoI. We have also evaluated the performance that a 1-level Haar lifting scheme would have in decorrelating this region. The choice of the Haar transform was made on the basis of its simplicity and furthermore, to exploit the fact that the matrix $\mathcal{H}$ coincides with the KLT of order-2, which is the optimum decorrelating 2×2 transform [26]. In effect, the use of the lifting scheme in this example corresponds to a lossless version of the second-order KLT. We have found that the effect of the Haar lifting transform on the entropy of the RoI was a decrease from 6.8123 bpp to 4.8744 bpp, or equivalently a compression ratio of 1.4:1 if an optimal entropy coder were used. Better results could be obtained by using a deeper lifting tree. For example, with a 3-level tree a reduction in entropy of 3:1 is achieved. Moreover, the multiresolution nature of the DWT adds the possibility for progressive transmission of the RoI. We feel that the applicability of the lifting scheme in reversible image compression deserves further investigation, in view of our preliminary results.

We have also investigated the potential of our scheme for PT applications. A few frames of a progressive decompression using the above technique are shown in Fig. 17, where it is also assumed that an RoI has been selectively compressed. Notice how the inherent prioritization of EZW shows up with the RoI being received and gradually refined first.

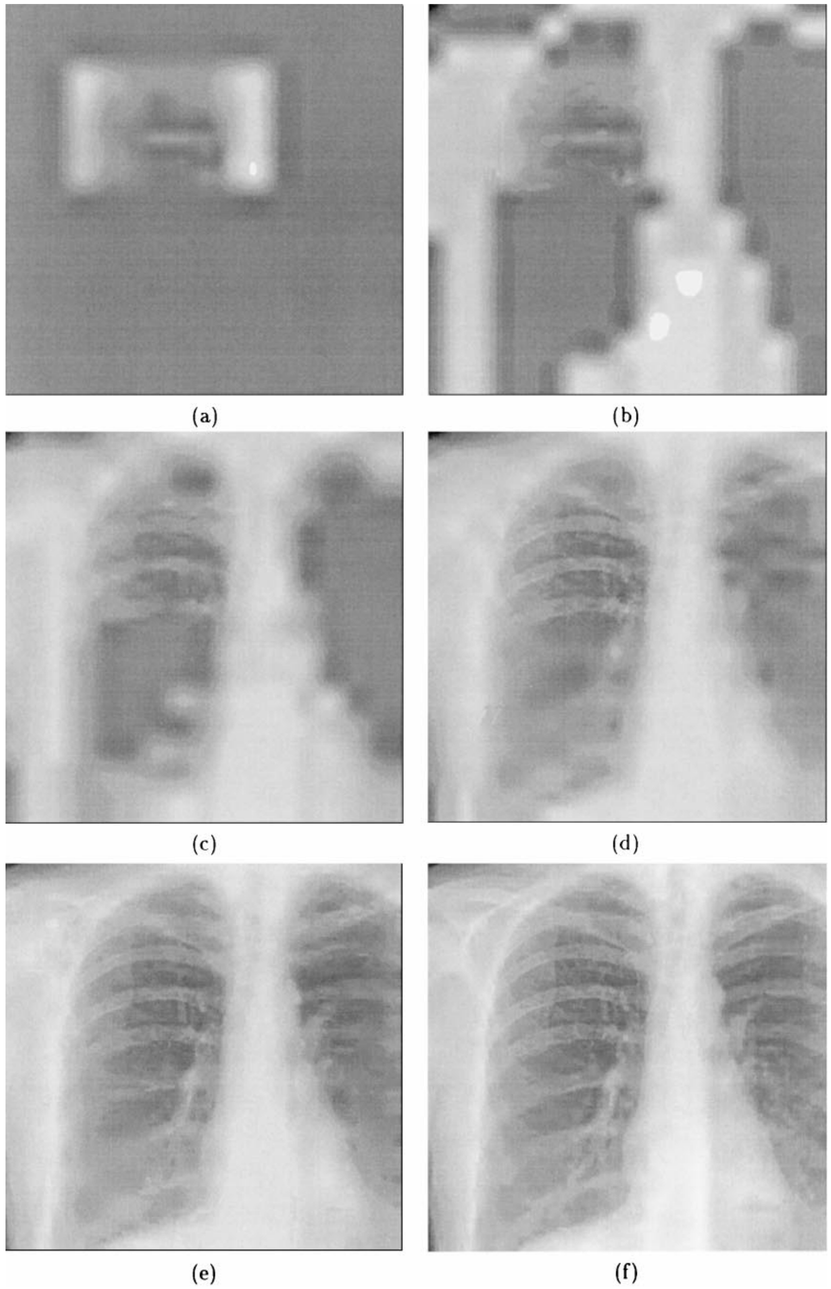


Fig. 17. Progressive decompression corresponding to Fig. 15(b). (Not all frames are shown.)

5. Concluding remarks

The conflicting demands for high-quality and low bit-rate compression imposed by the nature of the medical image compression problem are too difficult to be satisfied by the traditional coding techniques, including the common JPEG standard. To come up with a scheme that would seem promising in meeting the requirements for image management in contemporary PACS and teleradiology applications, we have tried to employ the state-of-the-art techniques in the current image compression palette of tools. Our system is based on a combination of the DWT, EZW compression with nonuniform coefficient distribution taken into account, and effective entropy coding techniques including a particularly efficient RLE-based coder. A notable feature of the proposed codec is the selective compression capability, which considerably increases its usefulness for MI applications, by allowing high compression and critical information preservation at the same time.

We have tested our codec in a variety of images and compression ratios and evaluated its performance in terms of objective measures as well as subjective visual quality. The unique features of multiresolution (or hierarchical) decomposition with enhanced spatial-frequency localization provided by the DWT, along with the clever exploitation of the DWT properties in the EZW algorithm are responsible for most of the performance gain exhibited by our scheme in comparison to JPEG. The ordering of the bits in terms of their importance in conjuction with the embedding characteristic increase the potential for fast and robust progressive transmission, in a manner that effectively combines the spectral selection and successive approximation methods of the progressive mode as well as the hierarchical mode of the JPEG algorithm [5,43]. Nevertheless, despite the good results obtained in our experiments, it should be emphasized that the performance of a compression algorithm is in general task-dependent and a period of thorough and extended testing and objective and subjective measurement has to be experienced prior to applying this methodology to clinical practice.

The compression scheme presented in this paper constitutes a powerful palette of techniques including high-performance approaches for traditional compression tasks as well as innovative features (such as SC) and novel configurations (e.g., lifting decorrelation) that pave the way to further research and development. One could also substitute for the various modules in the general scheme so as to adapt the system to the performance and/or functional requirements of a specific application area. For example, instead of the basic version of the EZW algorithm used here, one could try other variations, such as the SPHT algorithm [50], versions using adaptive thresholds [38], or extensions using statistical rate-distortion optimizations [71]. The use of the lifting realization of the DWT in lossless compression, with its multiresolution, low complexity and structurally guaranteed perfect inversion characteristics, has given interesting results in our study and it is perhaps a worthwhile way of extending the basic system.

The existence of powerful and cost-effective solutions to the problem of MI compression will be of major importance in the envisaged telemedical information society. We feel that the proposed framework is in the right direction towards the achievement of this goal.

References

[1] D.R. Aberle, et al. The effect of irreversible image compression on diagnostic accuracy in thoracic imaging, Invest. Radiology 28 (5) (1993) 398–403.

[2] American College of Radiology (ACR)/National Electrical Manufacturers Association (NEMA) Standards Publication for Data Compression Standards, NEMA Publication PS-2, Washington, DC, 1989.

[3] M. Antonini, et al. Image coding using wavelet transform, IEEE Trans. Image Process. 1(2) (1992) 205–220.

[4] A. Baskurt, I. Magnin, Adaptive coding method of X-ray mammograms, in: Y. Kim (Ed.), Medical Imaging V: Image Capture, Formatting, and Display, Proc. SPIE 1444 (1991) 240–249.

[5] V. Bhaskaran, K. Konstantinides, Image and Video Compression Standards: Algorithms and Architectures, Kluwer, Dordrecht, 1995.

[6] A.R. Calderbank et al., Wavelet transforms that map integers to integers, Technical Report, Department of Mathematics, Princeton University, 1996.

[7] J. Chen, M.J. Flynn, B. Gross, D. Spizarny, Observer detection of image degradation caused by irreversible data compression processes, in: Y. Kim (Ed.), Medical Imaging V: Image Capture, Formatting, and Display, Proc. SPIE 1444 (1991) 256–264.

[8] K. Chen, T.V. Ramabadran, Near-lossless compression of medical images through entropy-coded DPCM, IEEE Trans. Med. Imag. 13 (3) (1994) 538–548.

[9] P.S. Cho, K.K. Chan, B.K.T. Ho, Data storage and compression, in: Ref. [22], pp. 71–81.

[10] R. Claypoole et al., Nonlinear wavelet transforms for image coding, Technical Report, Bell Laboratories, Lucent Technologies, 1997.

[11] P.C. Cosman, et al. Thoracic CT images: Effect of lossy image compression on diagnostic accuracy, Radiology 190 (1994) 517–524.

[12] P.C. Cosman, R.M. Gray, R.A. Olshen, Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic quality, Proc. IEEE 82 (6) (1994) 919–932.

[13] G.G. Cox, et al. The effects of lossy compression on the detection of subtle pulmonary nodules, Med. Phys. 23 (1) (1996) 127–132.
[14] I. Daubechies, W. Sweldens, Factoring wavelet transforms into lifting steps, Technical Report, Bell Laboratories, Lucent Technologies, 1996.
[15] G.M. Davis, A wavelet-based analysis of fractal image compression, IEEE Trans. Image Process. 7(2) (1998) 141–154.
[16] S. Dewitte, J. Cornelis, Lossless integer wavelet transform, IEEE Signal Process. Lett. 4 (6) (1997) 158–160.
[17] Digital Imaging and Communication in Medicine (DICOM) version 3, American College of Radiology (ACR)/National Electrical Manufacturers Association (NEMA) Standards Draft.
[18] J.H. Dripps, K. Boddy, G. Venters, TELEMEDICINE: Requirements, standards and applicability to remote case scenarios in Europe, in: J. Noothoven van Goor, J.P. Christensen (Eds.), Advances in Medical Informatics, IOS Press, 1992, pp. 340–348.
[19] W.H.R. Equitz, T.M. Cover, Successive refinement of information, IEEE Trans. Inform. Theory 37(2) (1991) 269–275.
[20] R.A. Greenes, J.F. Brinkley, Radiology systems, in: E.H. Shortliffe, L.E. Perreault (Eds.), Medical Informatics – Computer Applications in Health Care, Addison-Wesley, Reading, MA, 1990.
[21] P.G. Howard, J.S. Vitter, Arithmetic coding for data compression, Proc. IEEE 82 (6) (1994) 857–865.
[22] H.K. Huang et al., Picture Archiving and Communication Systems (PACS), NATO ASI Series, vol. F74, PACS in Medicine, Springer, Berlin, 1991.
[23] D.A. Huffman, A method for the construction of minimum redundancy codes, Proc. IRE 40 (1952) 1098–1101.
[24] A.K. Jain, P.M. Farrelle, V.R. Algazi, Image data compression, in: M.P. Ekstrom (Ed.), Digital Image Processing Techniques, Academic Press, New York, 1984, pp. 171–226.
[25] N. Jayant, Signal compression: Technology targets and research directions, IEEE J. Selected Areas in Commun. 10 (5) (1992) 796–818.
[26] N.J. Jayant, P. Noll, Digital Coding of Waveforms, Prentice-Hall, Englewood Cliffs, NJ, 1984.
[27] R.G. Jost et al., High resolution teleradiology applications within the hospital, in: R.G. Jost (Ed.), Medical Imaging V: PACS Design and Evaluation, Proc. SPIE 1446, (1991) 2–9.
[28] G.R. Kuduvalli, R.M. Rangayyan, Performance analysis of reversible image compression techniques for high-resolution digital teleradiology, IEEE Trans. Med. Imag. 11(3) (1992) 430–445.
[29] H. Lee, et al. A predictive classified vector quantizer and its subjective quality evaluation for X-ray CT images, IEEE Trans. Med. Imag. 14 (2) (1995) 397–406.
[30] A.S. Lewis, G. Knowles, Image compression using the 2-D wavelet transform, IEEE Trans. Image Process. 1(2) (1992) 244–250.
[31] J.S. Lim, Two-Dimensional Signal and Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1990.
[32] S.-C.B. Lo et al., Full-frame entropy coding for radiological image compression, in: Y. Kim (Ed.), Medical Imaging V: Image Capture, Formatting, and Display, Proc. SPIE 1444 (1991) 265–271.
[33] H. MacMahon, et al. Data compression: Effect on diagnostic accuracy in digital chest radiography, Radiology 178 (1991) 175–179.
[34] A. Manduca, Compressing images with wavelet/subband coding, IEEE Eng. Med. Biol., Sept./Oct. 1995, pp. 639–646.
[35] M. Mansuripur, Introduction to Information Theory, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[36] K.M. McNeill, et al. Evaluation of the ACR-NEMA standard for communications in digital radiology, IEEE Trans. Med. Imag. 9 (3) (1990) 281–289.
[37] A. Macovski, Medical Imaging Systems, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[38] A. Munteanu et al., Performance evaluation of the wavelet based techniques used in the lossless and lossy compression of medical and preprint images, IRIS, TR-0046, 1997.
[39] M. Nelson, J.-L. Gailly, The Data Compression Book, M&T Books, 1996.
[40] R. Passariello et al., PACS-IMACS: Operation evaluation and basic requirements for prospective evolution of PACS technology, in: J. Noothoven van Goor, J. P. Christensen (Eds.), Advances in Medical Informatics, IOS Press, 1992, pp. 295–299.
[41] W.B. Pennebaker et al., Papers on arithmetic coding, IBM J. Res. Develop. 32 (6) (1988).
[42] A. Ramaswamy, W.B. Mikhael, A mixed transform approach for efficient compression of medical images, IEEE Trans. Med. Imag. 15(3) (1996) 343–352.
[43] K.R. Rao, J.J. Hwang, Techniques and Standards for Image, Video and Audio Coding, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[44] K.R. Rao, P. Yip, Discrete Cosine Transform-Algorithms, Advantages, Applications, Academic Press, New York, 1990.
[45] O. Rioul, M. Vetterli, Wavelets and signal processing, IEEE Signal Process. Mag., Oct. 1991, pp. 14–38.
[46] E.A. Riskin, et al. Variable rate vector quantization for medical image compression, IEEE Trans. Med. Imag. 9(3) (1990) 290–298.
[47] O. Rompelman, Medical image compression: Possible applications of subband coding, in: Ref. [70], pp. 319–352.
[48] P. Roos, et al. Reversible intraframe compression of medical images, IEEE Trans. Med. Imag. 7(4) (1988) 328–336.
[49] P. Roos, M.A. Viergever, Reversible image data compression based on HINT decorrelation and arithmetic coding, in: Y. Kim (Ed.), Medical Imaging V: Image Capture, Formatting, and Display, Proc. SPIE 1444 (1991) 283–290.
[50] A. Said, W.A. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, IEEE Trans. Circuits and Systems for Video Techn. 6(3) (1996) 243–250.
[51] P. Saipetch et al., Applying wavelet transforms with arithmetic coding to radiological image compression, IEEE Eng. Med. Biol., Sept./Oct. 1995, pp. 587–593.
[52] J.M. Shapiro, An embedded wavelet hierarchical image coder, Proc. ICASSP'92, pp. IV-657–IV-660.

[53] J.M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, IEEE Trans. Signal Process. 41(12) (1993) 3445–3462.

[54] J.M. Shapiro, Smart compression using the embedded zerotree wavelet (EZW) algorithm, Proceedings of the 27th Annual Asilomar Conference on Signals, Systems and Computers, 1993, pp. 486–490.

[55] L. Shen, R.M. Rangayyan, A segmentation-based lossless image coding method for high-resolution image compression, IEEE Trans. Med. Imag. 16 (3) (1997) 301–307.

[56] A. Signoroni, R. Leonardi, Progressive medical image compression using a diagnostic quality measure on regions-of-interest, Proc. EUSIPCO'98.

[57] R.L. Smathers, W.R. Brody, Digital radiography: current and future trends, British J. Radiology 58 (688) (1985) 285–307.

[58] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley–Cambridge Press, 1996.

[59] W. Sweldens, The lifting scheme: A new philosophy in biorthogonal wavelet construction, in: A.F. Laine, M. Unser (Eds.), Wavelet Applications in Signal and Image Processing III, Proc. SPIE 2569 (1995) 68–79.

[60] G.R. Thoma, L.R. Long, Compressing and transmitting visible human images, IEEE MultiMedia, April–June 1997, pp. 36–45.

[61] P.P. Vaidyanathan, Multirate Systems and Filter Banks, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[62] L. Vandendorpe, Optimized quantization for image subband coding, Signal Process. Image Commun. 4 (1991) 65–79.

[63] M. Vetterli, Multi-dimensional subband coding: Some theory and algorithms, Signal Process. 6 (1984) 97–112.

[64] M. Vetterli, J. Kovačevič, Wavelets and Subband Coding, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[65] J.D. Villasenor, B. Belzer, J. Liao, Wavelet filter evaluation for image compression, IEEE Trans. Image Process. 4 (8) (1995) 1053–1060.

[66] G.K. Wallace, The JPEG still picture compression standard, IEEE Trans. Consumer Electr. 38 (1) (1992) 18–34.

[67] P.H. Westerink, J. Biemond, D.E. Boekee, Progressive transmission of images using subband coding, Proc. ICASSP'89, pp. 1811–1814.

[68] I.H. Witten, R.M. Neal, J.G. Cleary, Arithmetic coding for data compression, Commun. ACM, 30 (6) (1987) 520–540.

[69] S. Wong, et al. Radiological image compression – A review, Proc. IEEE 83(2) (1995) 194–219.

[70] J.W. Woods (Ed.), Subband Image Coding, Kluwer, Dodrecht, 1991.

[71] Z. Xiong, N.P. Galatsanos, M.T. Orchard, Marginal analysis prioritization for image compression based on a hierarchical wavelet decomposition, Proc. ICASSP'93, pp. V-546–V-549.

[72] Z. Yang et al., Effect of wavelet bases on compressing digital mammograms, IEEE Eng. Med. Biol., Sept./Oct. 1995, pp. 570–577.