Neural Network-Assisted Effective Lossy Compression of Medical Images

NIKOS G. PANAGIOTIDIS, STUDENT MEMBER, IEEE, DIMITRIS KALOGERAS, MEMBER, IEEE, STEFANOS D. KOLLIAS, MEMBER, IEEE, AND ANDREAS STAFYLOPATIS, MEMBER, IEEE

A neural network architecture is proposed and shown to be very effective in performing lossy compression of medical images. A novel ROI-JPEG technique is introduced as the coding platform, in which the neural architecture adaptively selects regions of interest (ROI's) in the images. By letting the selected ROI's be coded with high quality, in contrast to the rest of image areas, high compression ratios are achieved, while retaining the significant (from medical point of view) image content. The performance of the method is illustrated by means of experimental results in real life problems taken from pathology and telemedicine applications.

I. INTRODUCTION

Image compression and coding techniques usually consist of solid mathematical models tuned so as to globally minimize the reconstruction error. Consequently, they reach decisions according to metrics and statistics extracted globally on a large variety of input data categories, while in most applications any *a priori* available knowledge is disregarded.

In contrast, in most real life applications, data is distributed over few categories, for which already available knowledge may be exploited in order to develop an optimum application specific coding platform. This is particularly the case in image encoding models, where one may have available significant information concerning, for example, the most frequently utilized colors, the significance of information contained in the chrominance versus luminance components, the desired compression ratios, or the target transmission times.

The scheme presented in this paper stems from the idea that the majority of medical images consist of areas of minimal contribution to the information finally perceived (e.g., background), and of foreground objects or regions which are of extreme interest to the expert/end-user examining the images. An efficient coding and compression scheme is developed, which takes advantage of the difference in visual importance between areas of the same image and exploits it by coding regions of interest (ROI's) with maximum precision, while implementing a tolerably lossy reconstruction of the low-interest areas.

Medical image compression has been up to now mainly concerned with lossless coding techniques [1], [2], which provide compression ratios of around 2:1, ensuring that all significant information for medical purposes is retained in the reconstructed images. The Digital Image and Communications in Medicine proposed standard (DICOM3), that has been derived and adopted by the American College of Radiology and National Electrical Manufacturing Association (ACR-NEMA), includes lossless coding and compression of medical images. However, recent studies concerning, for example, the amount of quantization error in digitized images indicate that lossy compression can be adopted if the reconstruction error does not significantly affect the image quality from a medical point of view. Based on such results, in 1995 the ACR-NEMA announced a call for proposals for lossy image compression techniques that are to be included in DICOM3.

The definition of ROI's can be used to achieve variable spatial reconstruction of the original medical images. Depending on the ratio of high- to low-importance regions, substantial saving can be obtained in the time/space required for transmitting or storing the image, while causing an unperceivable degradation in the image quality. A big gain is generally expected in images where information is concentrated in relatively small portions of the image, a case which is met in various disciplines, such as pathology imaging, radiology examinations (e.g., computed tomography, magnetic resonance imaging (MRI), nuclear imaging, and mammography), and X-rays (breast, chest, bone, and skull). Especially when transmission of such images is required, for example in telemedicine applications, high-compression ratios are necessary, together with a very good quality rendition of the areas containing certain parts of the images.

Manuscript received October 1, 1995; revised February 1, 1996. This work was supported in part by the Greek Telemedicine System and NIKA Projects EKBAN-504 and STRIDE-141.

The authors are with the Division of Computer Science, the Department of Electrical and Computer Engineering, the National Technical University of Athens, Zografou 15773, Athens, Greece (e-mail: npanag@softlab.ece.ntua.gr, dkalo@softlab.ece.ntua.gr, stefanos@softlab.ece.ntua.gr, and andreas@softlab.ece.ntua.gr).

Publisher Item Identifier S 0018-9219(96)07233-7.

Variable spatial reconstruction is, however, possible, if *a priori* levels of reconstruction quality for relevant information can be guided by the contents of the images. This implies that a mechanism is required, which examines the medical images and assists the doctor performing the examination, by automatically providing him with the "suggested" ROI's in the images. An appropriate neural network architecture is proposed in this paper as an efficient and powerful technique for performing this task.

Several neural network techniques have been applied to the problem of image compression [2]. They include nonlinear predictors for predictive coding, principal component extraction of image data for block transform coding, as well as vector quantization approaches. Nonlinear predictors are mainly based on multilayer perceptrons [3], [4], without excluding recurrent network architectures. Several approaches for the calculation of principal components have been proposed, which are typically established upon some form of Hebbian learning algorithm [5], [6]. As far as vector quantization is concerned, the self-organizing map (SOM) [7] has formed the basis of several algorithms for codebook generation. Although most of the above techniques have been claimed to yield promising results, it must be admitted that a thorough and comprehensive evaluation is needed to assess the benefit of using neural networks in place of well-established conventional approaches to image compression. Nevertheless, the availability of efficient hardware implementations exploiting the massively parallel nature of neural networks constitutes a promising perspective in this direction [8].

On the other hand, neural network methods have shown an admittedly good performance in solving problems associated with pattern recognition and classification tasks [9]–[11]. Several issues relating to the potential and limitations of widely used network models, such as multilayer perceptrons and radial basis function networks, have been investigated during the last several years by means of theoretical results and real-world experimentation [3], [12], and [13]. Other neural network approaches, including the categories of probabilistic neural networks (PNN's) [14] and learning vector quantizers (LVQ's) [15], have been applied to a variety of problems. If properly designed, the above approaches can yield near-optimal performance in pattern classification, in most cases.

In this paper, the ability of neural networks to provide effective solutions to image classification and recognition problems is exploited in order to adaptively select ROI's in medical images. More specifically, a neural network module is generated and embedded in a novel ROI-JPEG approach to lossy compression of medical images that uses standard coding techniques (e.g., JPEG, MPEG, and H-26x).

The ROI-JPEG scheme presented next is composed of two distinct algorithmic modules. The first one corresponds to the segmentation stage, which is neural-network-assisted and performs ROI selection and localization. The second stage performs lossy adaptive compression using a coder derived from the JPEG baseline [16], where the ROI-based approach generates spatial reconstruction of suitably nonhomogeneous quality. Following the selection of areas of high/low importance in the image, important regions are coded with appropriately selected high-quality quantization tables, in contrast to the rest of the image which is coded with low-quality quantization tables (i.e., fewer bits per block). Section II describes the coding platform, focusing on the introduction of ROI's in the JPEG coder. Section III presents the proposed hierarchical neural network architecture for performing the ROI selection task, while the specific neural network models included in this architecture are described in Section IV. Experimental studies using real life medical image data, which illustrate the performance of the method, are given in Section V, while conclusions are provided in Section VI.

II. ROI in JPEG Image Coding

A. The JPEG Standard

The JPEG standard [16] for still image coding is based on the discrete cosine transform (DCT) [17]. Color images represented in the component (R, G, B) color space are transformed to the luminance/chrominance Y, C_r, C_b space prior to coding. According to the baseline encoder model, the input image is divided into blocks, usually consisting of 8×8 pixels, which are transformed into the DCT domain. The transform coefficients are then quantized using a user-specifiable quantization matrix and the quantized coefficients go through a lossless coding procedure using either arithmetic or Huffman coding. Following this structure, it is possible to design quantization matrices that take into account image-dependent information according to specific properties of the human visual system. In this case, the coder may provide images of subjectively better quality than the respective quality of images provided by a coder which uses the default matrices included in the JPEG standard. Nevertheless, the above procedure can still work only at a global level, since the quantization matrices will be fixed within the whole image.

A progressive coding scheme is possible if the coefficients of the DCT table are divided into groups, which can be subsequently handled sequentially. A frequently met case generates groups of coefficients corresponding to the low, medium, and high-frequency content of the original image; such groups are designated as L, B, and H in Fig. 1. The boundaries of any group may be modified to describe the appropriate frequency band. Moreover, since coefficients corresponding to very high frequencies generally concentrate a small fraction of the total image energy, coefficients from groups B or H can be set to zero, yielding imperceptible errors in conjunction with a significant reduction of code length for compression purposes.

B. The ROI-JPEG Compression Algorithm

The ROI-JPEG coder provides the means for encoding regions of low/high interest in the image by differentiating the quantization tables among these areas. This goal is achieved

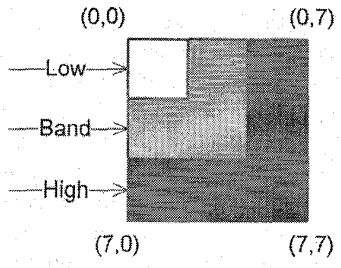


Fig. 1. Spectral decomposition in DCT domain.

by using different quantization quality factors (QF), as defined in the baseline JPEG algorithm, for each category of image regions. Thus while blocks belonging to important regions are coded with high-quality quantization tables, a substantial reduction in bitstream volume is achieved by quantizing the rest of the image blocks with low-quality quantization tables. Further reduction in the volume of information transmitted may be achieved by further filtering of the low-importance regions. This is achieved through the DCT transform, and does not affect the perceived quality of these regions, since coarse quantization already incorporates a low-pass filtering process. Additionally, for off-line storage applications, visually optimal quantization tables on a bits/pixel target rate basis can be computed for both the high interest and background regions of the image [18].

A block diagram of the ROI-JPEG procedure is shown in Fig. 2. It includes the typical components of the JPEG system, i.e., the DCT transform, the quantizer, and the entropy coder (Huffmann or arithmetic), applied to each block of the image. A decision step is added, which classifies the image block either to the window category (ROI) requiring high reconstruction quality, or to the category of relatively low importance. Using more than two categories is possible; however in most applications of interest, two categories seem to be enough for achieving high compression ratios. Let us assume first, that the decision is based on information which an expert interactively gives to the system; this can be, for example, performed by marking the important areas on the captured image, before applying the compression procedure to it. Since this is infeasible, when dealing with large volumes of data, a neural network architecture is proposed in the next section as a means for assisting the expert in automatically selecting the regions of high importance. The coordinates of high importance regions are stored in the coded image header, thus modifying the standard JPEG header structure. Apart from the coordinates of these regions, the decision block provides the quantizer with the specific value of the QF that is required to define the quantization matrices for each ROI category. It is generally desired that the differences between the QF's used are not large, so that blocking artifacts at the borders of the areas are not perceivable.

The image is coded on a block-based scheme, as is the case with baseline JPEG, which defines horizontal and vertical sampling factors for each color component. The sampling factors specify the number of samples of each component relative to the other components in the frame, in the corresponding direction (horizontal or vertical). For example, in a 4:1:1 Y, C_r, C_b image, the sampling ratios are equal to two in each dimension for the luminance component and to one for the chrominance components, as shown in Fig. 3. Naturally, for a single component input image (e.g., grayscale picture) all sampling factors are equal to one.

The selection of regions of interest results in a classification map of the image blocks; this map uses one bit (when classifying blocks in two categories of high/low importance) per block to denote whether it belongs to a ROI or not. The encoder stores this classification map in the image header bit stream according to the JPEG standard, so that the decoder is capable of recognizing the category of each decoded block. The quantization tables, which are used for coding and reconstructing blocks belonging to ROI's, are also stored in the image header, according to the JPEG specifications. Different quantization tables can be defined by letting the user specify a QF value for low-importance regions, and a window quality factor (WQF) for high-quality regions (ROI's); the use of more than two categories of regions is possible, if respective QF's are defined for each category. The properties of QF and WQF are similar to those of the standard JPEG QF; both are used for the derivation of quantization tables from the standard templates incorporated into the JPEG baseline. In general, QF and WQF lie in the intervals [30], [60], [70], [85], respectively.

In summary, the encoding process consists of the following steps. The image is divided into blocks of 8×8 pixels, which are transformed in the DCT domain. If the block currently processed belongs to a ROI, it is quantized using the high-quality quantization tables, if not, low-quality quantization and low-pass filtering are performed. The quantized block is subsequently routed to the lossless coding unit, Huffman or arithmetic coding having been chosen by the user during initialization.

On the decoder side, first the bitstream header is processed in order to extract the stored quantization tables and the classification map, including a description of the category of each image block. According to the previous decision, the corresponding inverse quantization tables are used to generate the signal in the DCT domain. Subsequent inverse-DCT transformation generates the decoded image block.

III. The Neural Network Architecture

The ROI selection stage is a time consuming process in real-time applications, where large amounts of medical image data are considered. In the following a neural network system is proposed for efficiently performing the ROI selection task in an adaptive way.

The proposed system is composed of an hierarchical two-level architecture. The first level of this architecture automatically selects all edges appearing in the examined

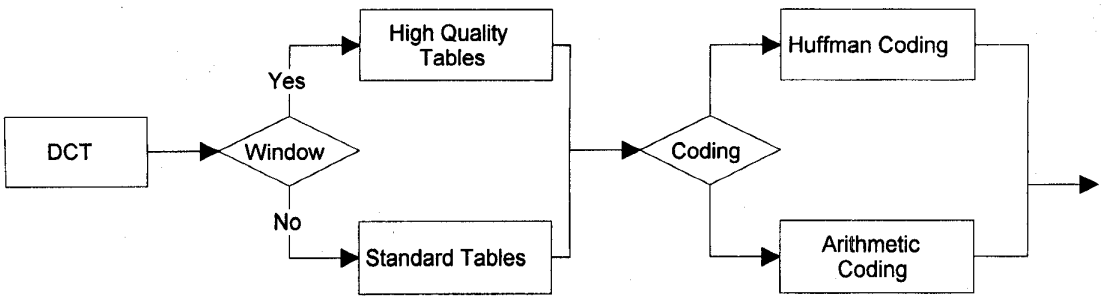


Fig. 2. ROI-JPEG coder block diagram.

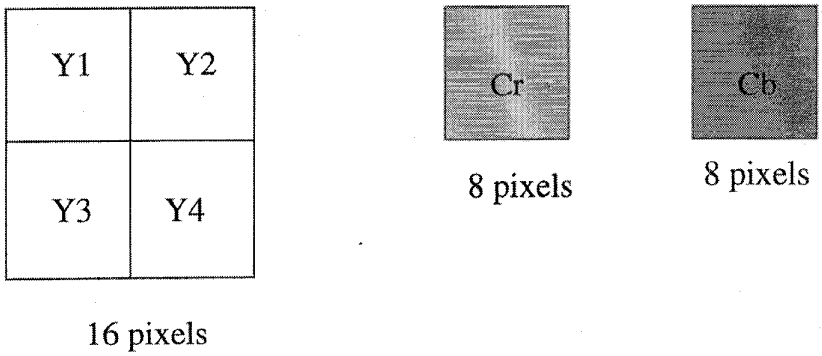


Fig. 3. Y, C_r, C_b coding and subsampling.

medical images and classifies the corresponding blocks to the ROI category. This is due to the fact that, in most applications, the edges existing in the image belong to regions that are of major importance for recognition or classification purposes. This level consists of a feedforward network which performs the frequency dependent edge detection task, classifying each image block to a "shade" or "edge" category. "Shade" blocks correspond to homogeneous areas containing no significant edges, while "edge" blocks generally include significant high-frequency content. Due to the fact that edge blocks have more energy in the high-frequency components, a higher QF should be used in order to keep the same quality in the reconstructed block. To accomplish this task, the network accepts at its input the computed DCT coefficients of each image block and exploits high-frequency information appearing in all three components, i.e., luminance (Y) and chrominance (C_r, C_b), of the image. Usage of all three components is necessary because in medical imaging and in other applications edge information depends not only on luminance, but also on color spatial variations.

To let the network detect edges along all different orientations, most, or even all, DCT coefficients from the three components of each image block may be required and should be therefore presented at the network input; consequently, the number of network input units can be up to 192 (3×64), while the number of network outputs is assumed to be equal to two, which correspond to the above defined two categories. In Section V, we consider the number of DCT coefficients which are necessary for coding medical images in real life applications.

Supervised learning has been adopted for training the network to perform edge detection. According to it, a predefined training set of characteristic images is selected, to which conventional spatial edge detection operators, such as Sobel or gradient methods, or more advanced morphological operators are applied; the results of these operators are further examined by experts to improve the quality of detection. Following this selection the images are divided into blocks which are DCT transformed and labeled as "shade" or "edge" ones. The block DCT coefficients for all image components and the corresponding labels for each block are then used to train the network.

The operation of this network is, therefore, similar to that of an edge detector, enriched due to training in accordance with the expert's advice. Furthermore, direct application on the DCT coefficients computed within the JPEG coding scheme provides real-time operation, eliminating the need for preprocessing the images with conventional edge detection masks or filters.

After training, the network is able to classify each block of an image similar to the ones used for training, to an "edge" or "shade" category. In the former case, the block is automatically selected to belong to a region of high visual importance (ROI). If, however, the block is found to belong to an homogeneous region, no decision is taken, but the block is subsequently fed as input to the second level of the proposed architecture, which consists of another network that finally classifies it to a ROI or not. This step is necessary in order to accurately render features like color or amplitude levels, which, although not corresponding to high-frequency information, can be

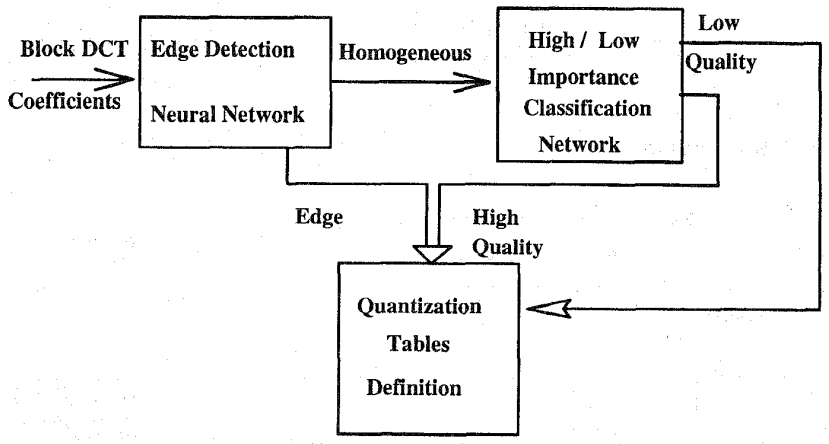


Fig. 4. The proposed neural network architecture.

of high importance in medical, as well as in other applications.

The second network is similar to the first one, in the sense that it also uses the computed DCT coefficients of the block as input features. The number, however, of these features is generally smaller than the corresponding number of the first network input units, which had to detect edges, i.e., high-frequency information across all possible orientations. In particular, the input features that are chosen to feed the second network are the DC coefficient and a small number of AC coefficients following the well-known zigzag DCT scanning of each image component. A supervised learning algorithm has also to be followed in order to train the network, using images from the specific application under consideration. Whenever a block is classified as one of high importance, i.e., belongs to a ROI, fine quantization of its pixel values is used when coded, whereas coarser quantization is applied to pixels belonging to blocks of lower importance. A diagram of the hierarchical network architecture is shown in Fig. 4.

It should be mentioned that in cases where edges do not play an important role, e.g., in texturelike medical images, it is possible to bypass the first level of the hierarchical architecture, focusing on the results of the second level.

The above-described architecture finally provides a classification map per image, which is a binary image indicating whether each image block belongs to a ROI or not; the maps are generated by joining together the outputs of both levels of the proposed architecture. Since classification maps are constructed on a 8×8 pixel block basis, their size is 1/64 of the original images (1/8 in each dimension). Further compression of these maps is possible in bits per pixel basis, due to their binary (bitmap) nature. The classification maps, as well as the corresponding quantization tables, are stored in the image header, so as to be used for the subsequent coding and reconstruction procedure.

Various learning algorithms can be used for training each level of the above-described hierarchical feedforward architecture. Such algorithms include variants of backpropagation [3], [19], the learning vector quantization method with optimized learning rate [15], or probabilistic

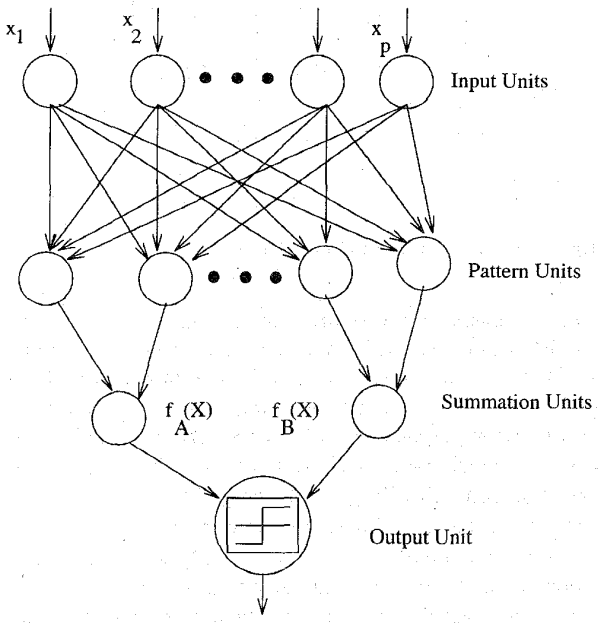


Fig. 5. A PNN.

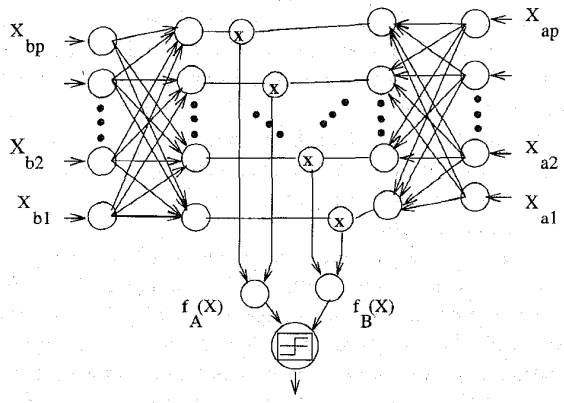


Fig. 6. A hierarchical PNN.

networks [14], [20]. An important issue for selecting the most appropriate learning algorithm is the retraining of

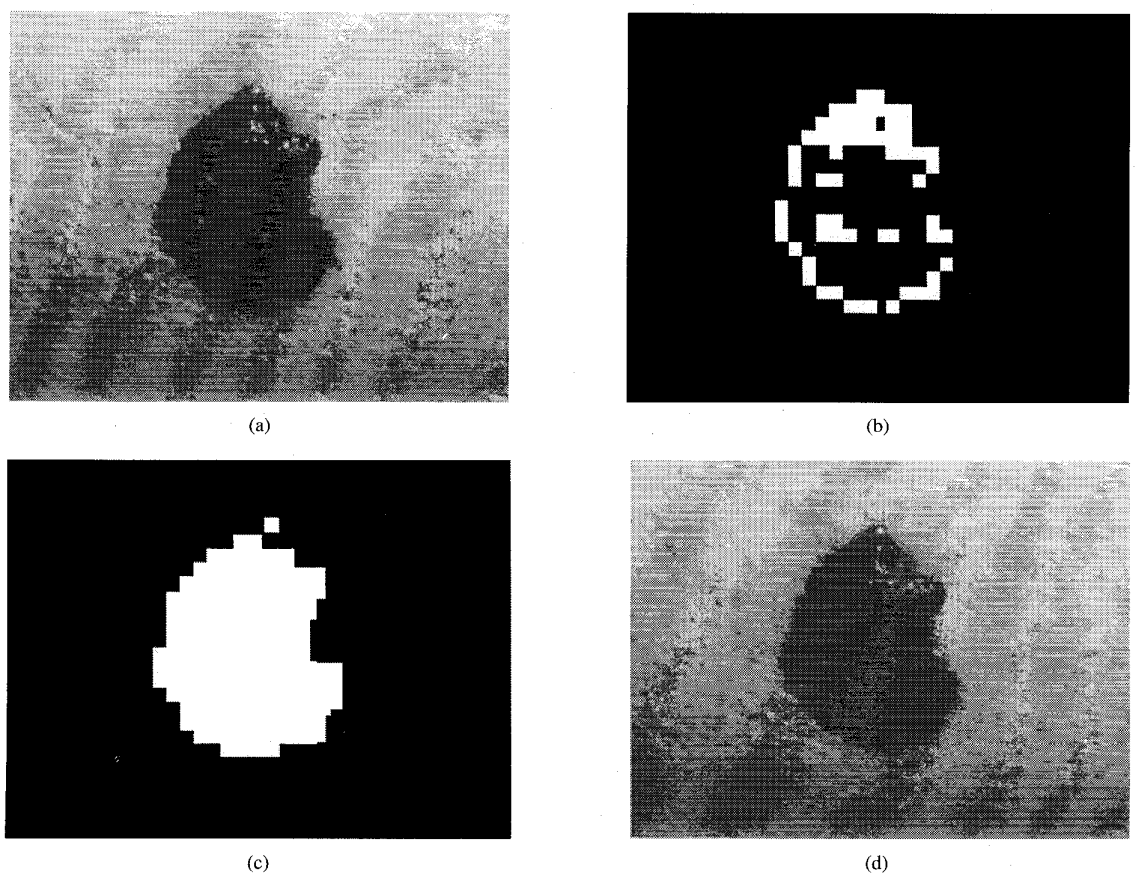


Fig. 7. (a) Original image, (b) neural network-based edge detection, (c) neural network-based classification map, and (d) reconstructed image.

the networks that will be required whenever the doctor or expert wishes to include a new case study in the already available training set. It would be desirable that the doctor or expert does not have to extensively run the training algorithm, adjusting parameters so as to achieve the best performance. PNN's are used as an appropriate scheme for the learning procedure. A hierarchical multiresolution model of such networks is presented in Section IV, which efficiently implements the desired classification task.

IV. Probabilistic Networks

Many neural network classifiers provide outputs which estimate Bayesian *a posteriori* probabilities. When estimation is accurate, network output values sum to one and can be treated as probabilities. Bayesian probabilities can be estimated by multilayer perceptrons, based on the minimization of a mean squared error or a cross entropy cost function [21]. The desired network outputs correspond to, say, M classes; one output is equal to unity and all others are equal to zero. Error feedback supervised learning algorithms using the Kullback Leibler (KL) criterion and the generalized sigmoidal function [21], [22] have been shown capable of producing Bayesian *a posteriori* probabilities or conditional likelihood estimates for classification purposes. The estimation accuracy generally depends on the network complexity, on the amount of training data, and on the degree to which training data reflect true likelihood distributions and *a priori* class probabilities.

Unlike perceptron type networks, which classify input vectors by learning multidimensional decision surfaces, PNN's [14] classify input vectors by forming nonparametric probability density functions (PDF's). The network structures are similar to those of multilayer perceptrons; the primary difference is that the sigmoid activation function is replaced by the exponential one. Key advantages of PNN's are that training requires only a single pass and that decision surfaces approach the Bayes-optimal decision boundaries as the number of training samples grows.

PNN's utilize the fact that, in the limit, any smooth and continuous PDF, say f_A, of a class of multidimensional data $\mathbf{X}$, can be estimated by a sum of multivariate Gaussian distributions centered at each training sample [23] as follows

$$f_A(\mathbf{X}) = \frac{1}{(2\pi)^{p/2}} \frac{1}{m} \sum_{i=1}^{m} \exp\left[-(\mathbf{X}-\mathbf{X}_{\alpha i})^T \frac{\mathbf{X}-\mathbf{X}_{\alpha i}}{\sigma^2}\right] \quad (1)$$

where i is the pattern number, m is the total number of training patterns in category A, $\mathbf{X}_{\alpha i}$ denotes the ith training

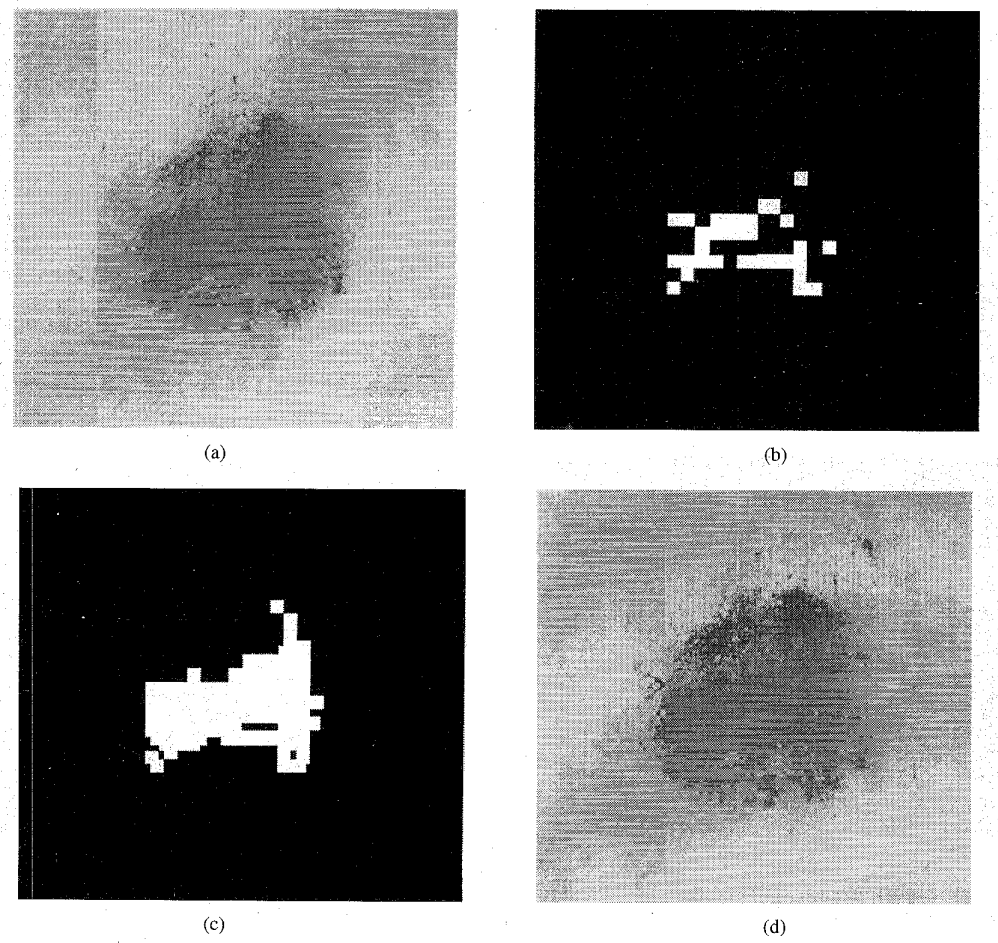


Fig. 8. (a) Original image, (b) neural network-based edge detection, (c) neural network-based classification map, and (d) reconstructed image.

pattern from category A, p is the dimensionality of input vector $\mathbf{X}$ and σ is a smoothing parameter.

Fig. 5 shows a neural network which can be used for estimating the PDF of classes of input patterns $\mathbf{X}$ and for subsequently classifying them into categories. A two category classification, into classes A and B, is shown in Fig. 5; extension to a larger number of classes is straightforward. The input units are merely distribution units that supply the p elements of each input data vector to all pattern units, the number of which is equal to the number m of training data. The ith pattern unit forms a dot product of the input data vector and a corresponding weight vector $\mathbf{W}_i$ and passes it through the exponential activation function, as follows:

$$\exp\left[\frac{\mathbf{W}_i\mathbf{X}^T - 1}{\sigma^2}\right]. \tag{2}$$

Assuming that both $\mathbf{X}$ and $\mathbf{W}_i$ are normalized to unit length, the output of the ith pattern unit is equivalent to a probability measure:

$$p_i = \exp\left[-(\mathbf{X} - \mathbf{W}_i)^T \frac{\mathbf{X} - \mathbf{W}_i}{2\sigma^2}\right]. \tag{3}$$

The network summation units compute through (3), each class PDF represented by (1). The output units simply select the class with the maximum PDF.

The network is trained by setting the $\mathbf{W}_i$ weight vector of the ith pattern unit equal to the corresponding $\mathbf{X}$ pattern in the training set, and by then connecting the unit output to the appropriate summation unit. The smoothing parameter σ controls the activation function; large (small) values of σ reduce (increase) the sensitivity of the exponential function.

The basic problem with probabilistic networks is the increase of pattern units proportionally to the number of training examples; this increase generates a requirement for large storage and computing power of the system implementing the algorithm. In the following, multiresolution analysis is used as a means for efficient use of probabilistic networks in compressing medical images.

A. Two-Dimensional Multiresolution Analysis

Multiresolution image representations have been known for a long time in applications such as correlation matching, edge detection, segmentation, and image analysis [24], [25]. These representations are generally used to reduce the

dimensionality of the problem as well as the associated computational load and to perform feature extraction at different resolution levels. Resolution reduction is most commonly performed by subsampling or local averaging of the image pixel values.

Multiresolution image analysis and processing based on the use of the wavelet/subband decomposition has recently attracted major interest, mainly for coding and scalability applications [26], [27]. In general, the transition from one level of resolution to a lower one is implemented by subsampling each dimension of the higher resolution data, usually by a factor of two, and by using a finite impulse response (FIR) filter with fixed taps; typical eight or 32 tap filters are given in [27]. As a consequence, four $(N/2 \times N/2)$ low-resolution images are produced from each color component of a $(N \times N)$ color high-resolution image; one of them, generally corresponding to the low-frequency content of the image, is the approximation image, while the rest are the detail images, generally including high-frequency content.

B. Hierarchical Probabilistic Networks

PNN's have serious advantages with respect to training time when compared to multilayer perceptrons, since they are single layer networks with a single-pass training procedure. As a consequence, a doctor can handle such networks more easily, especially if one has to frequently retrain them with new data sets obtained from different experiments. Probabilistic networks are therefore appropriate to be included in the two-level architecture proposed in Section III for adaptive ROI selection and compression of images.

The main drawback of probabilistic networks, however, is their requirement for large memory, since during training all training samples are stored in the weights of pattern units. It is, therefore, desirable to reduce the memory requirements, with negligible degradation of the network performance. Multiresolution analysis is an effective means for reducing the size of input images; use of the approximation image instead of the original one provides a significant reduction of the required number of weights of the pattern units in the corresponding probabilistic network; it is, however, required that this approximation image contains the maximum information quotient of the original image.

In the following, instead of using the subband/wavelet image decomposition, we generate the low resolution approximation and detail images by selecting different frequency bands of the original images, according to the progressive procedure included in the ROI-JPEG algorithm. The definition of low, medium, and high areas (see Fig. 1) can form the basis for this spectral partition. For example, let us assume that only the low-frequency area of Fig. 1, composed of 16 (4×4) out of 64 coefficients, is used to reconstruct the images, with no significant degradation of the quality of the original images. Then, for a sample training set consisting of a large number of image blocks m, the required number of weights connecting the pattern units to the inputs significantly reduces from $64m$ to only $16m$. Other choices of the frequency bands, which are defined following the well-known zigzag scanning of the DCT coefficients, are also possible and are considered when implementing the method, as described in Section V.

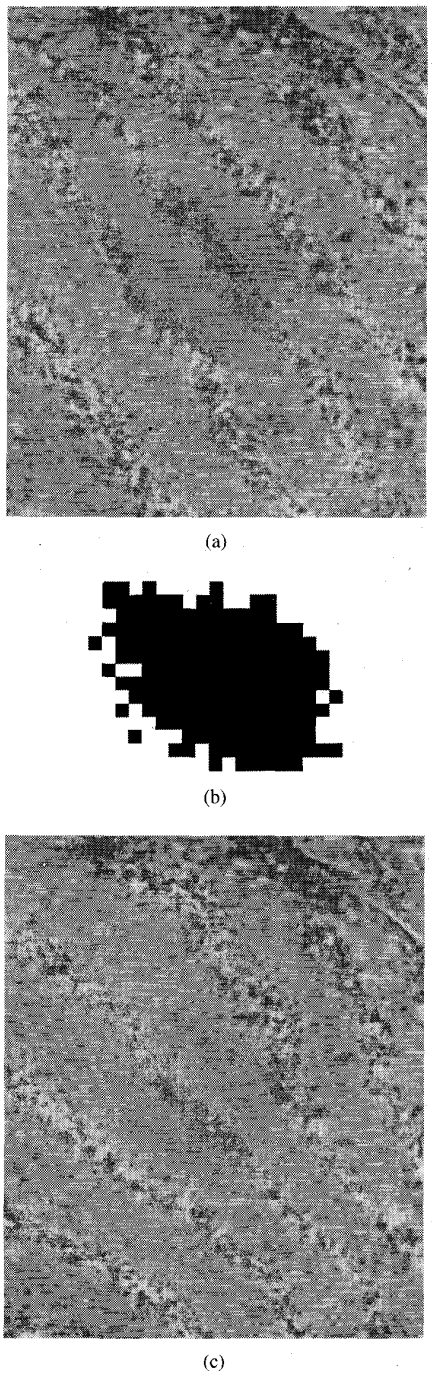
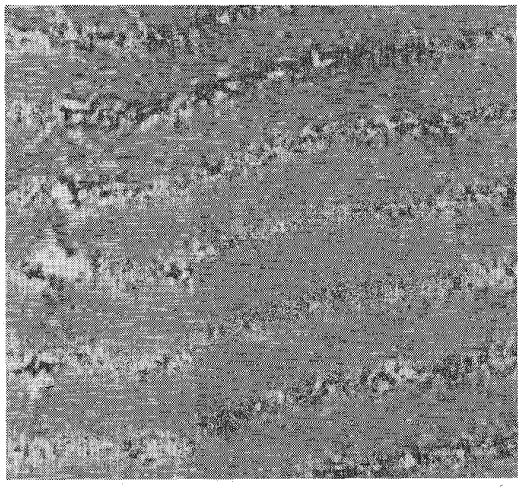
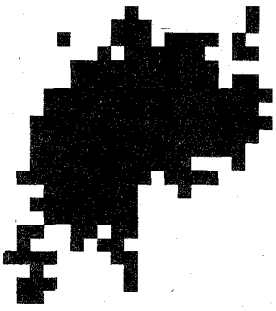


Fig. 9. (a) Original image, (b) neural network-based classification map, and (c) reconstructed image.

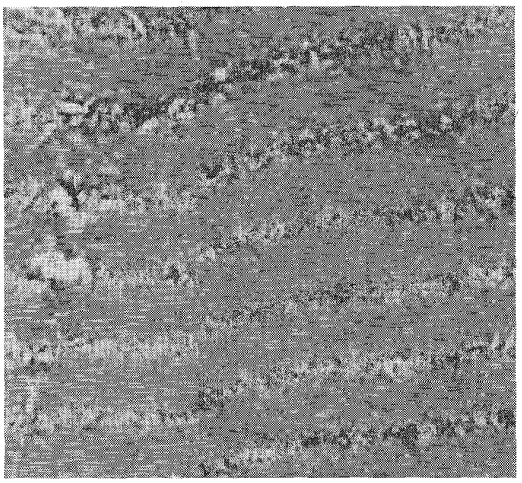
Let us assume that after training the network using a low resolution approximation image, the network generalization ability is tested using, for example, a validation data set and is not found satisfactory. In this case the method includes a projection of the computed weights toward the



(a)

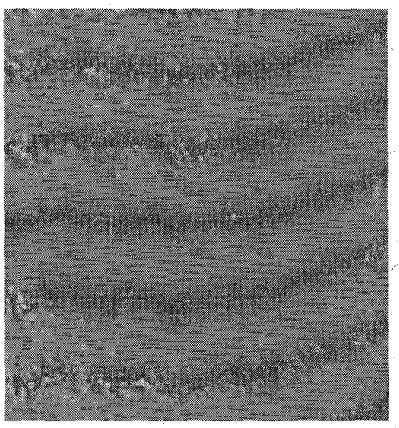


(b)

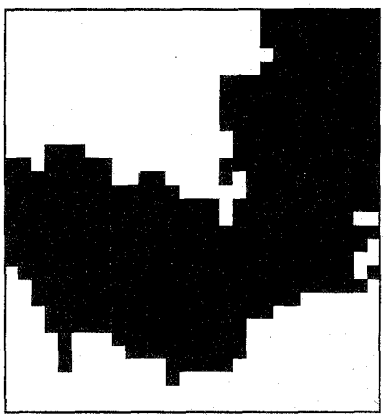


(c)

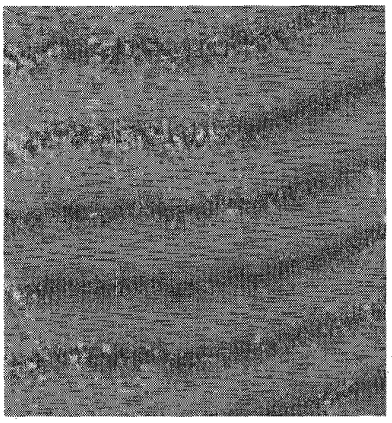
Fig. 10. (a) Original image, (b) neural network-based classification map, and (c) reconstructed image.



(a)



(b)



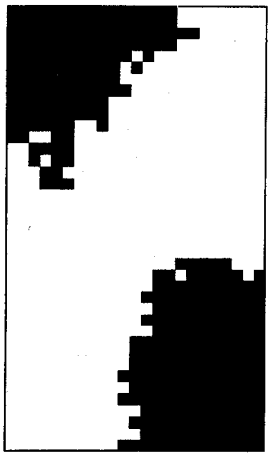
(c)

Fig. 11. (a) Original image, (b) neural network-based classification map, and (c) reconstructed image.

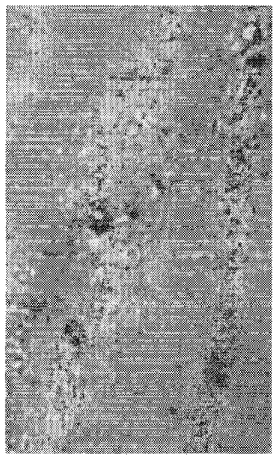
next resolution level, so that the already derived network knowledge, i.e., weight values, be included in the network architecture of the following level [28]–[30].

Let the image $\mathbf{X}$ at some resolution level $j+1$ be split into two parts $\mathbf{X} = [\mathbf{X}_a \mathbf{X}_b]$, the first of which represents the approximation image and the second the remaining detail subsampled replica of the original image. Let us consider a probabilistic network which is trained at resolution level j using only the approximation image. As was mentioned above, if the network performance is tested and not found acceptable, then training should proceed to the next higher resolution level $j+1$; it would, however, be desired that the training already performed at level j be used in this procedure, so that training at level $j+1$ does not start

(a)



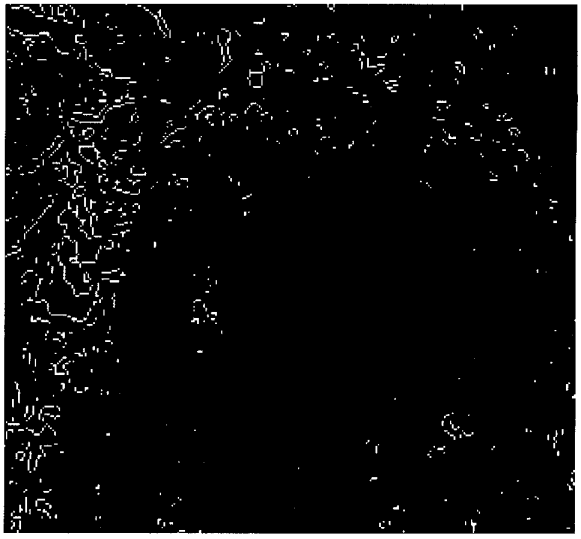
(b)



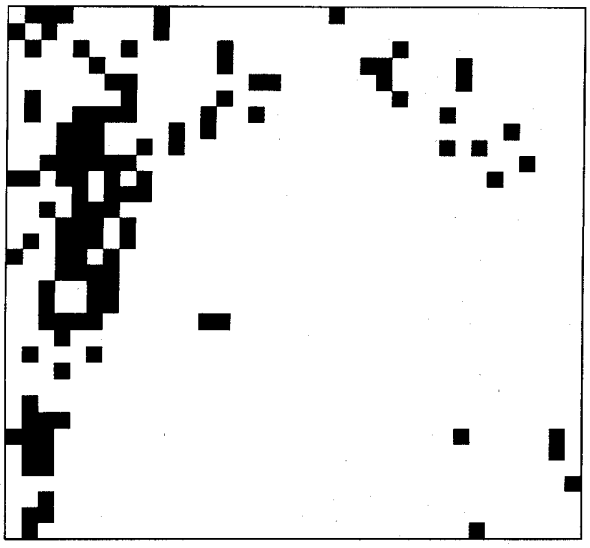
(c)

Fig. 12. (a) Original image, (b) neural network-based classification map, and (c) reconstructed image.

from zero initial conditions. To accomplish this, the detail image at resolution level j can be classified by a second network that is separately trained. It is shown next that the outputs of the two networks can be combined to construct the network at $j+1$.



(a)



(b)

Fig. 13. (a) Edge detection and (b) block classification.

Following the decomposition of the input vector $\mathbf{X}$, a corresponding decomposition of the network weights can be performed as follows:

$$\mathbf{W}_i = [\mathbf{W}_{\alpha i}\mathbf{W}_{bi}]. \tag{4}$$

The output of the pattern unit $p_i(\mathbf{X})$ can be consequently decomposed in the following form:

$$\begin{aligned} p_i(\mathbf{X}) &= \exp\left[\frac{\mathbf{W}_i\mathbf{X}^T - 1}{\sigma^2}\right] \\ &= \exp\left[\frac{(\mathbf{W}_{\alpha i}\mathbf{W}_{bi})(\mathbf{X}_{\alpha i}\mathbf{X}_{bi})^T - 1}{\sigma^2}\right] \\ &= k \cdot \exp\left[\frac{(\mathbf{W}_{\alpha i}\mathbf{X}_{\alpha i} - 1) + (\mathbf{W}_{bi}\mathbf{X}_{bi} - 1)}{\sigma^2}\right] \\ &= k \cdot p_{\alpha i}(\mathbf{X}) \cdot p_{bi}(\mathbf{X}) \end{aligned} \tag{5}$$

where k is a scaling factor equal to $\exp(1/\sigma^2)$.

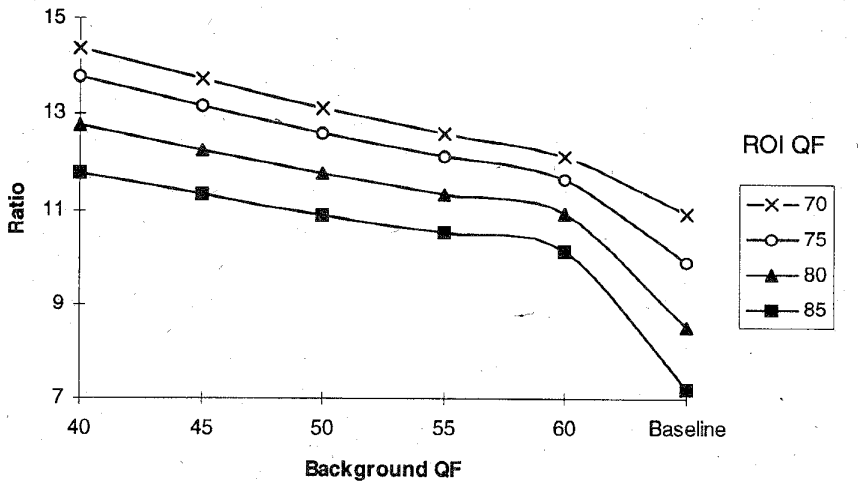


Fig. 14. ROI-JPEG compression ratios.

Thus in order to combine the probabilistic networks trained to classify the approximation and detail images at level j for classification at level $j+1$, we simply have to multiply the outputs of their respective pattern units and then feed them in the appropriate summation units, as depicted in Fig. 6.

V. Experimental Study

The proposed neural network architecture has been implemented and tested for adaptive lossy compression of images in two real life medical applications, namely transmission of images through a low bit rate telemedicine system and coding of pathology examinations in hospital environments.

The coder described in the previous sections has been implemented first to accommodate the requirements of a telemedicine system regarding dermatology examinations, using conventional telephone lines between remote health care centers and central hospitals. A set of still images has been obtained and, through a guided user interface, a medical expert selected the areas of the images (ROI's) that contained the most useful visual information (e.g., bruises, burns, and tumors). These areas have been used next to train the proposed two-level neural network architecture so as to automatically select ROI's in images containing similar visual information.

Figs. 7(a) and 8(a) show characteristic images of typical sizes 320×240 pixels which contain important ROI's (from a medical point of view), as well as areas of low importance. The left half of the image shown in Fig. 7(a), which includes ROI's (moles on human skin) and background areas (rest of the skin), was used to train the two-level neural network architecture described in Sections III and IV. As a preprocessing step, the ROI and background areas were separated into blocks of 8×8 pixels which were tagged both for edge/nonedge selection and for texture high/low-quality coding. The first level of the network architecture was trained to extract the edges and the second to classify background blocks to ROI's or non-ROI's, based on their frequency content. To accomplish this, the DCT transform of each block was computed, followed by zigzag scanning of the DCT coefficients. Various experiments have been performed to appropriately partition the frequency content of the blocks into two areas, that were used next to form the corresponding lower resolution approximation and detail image blocks for the proposed probabilistic network model. Our results indicated that the quality of the results was not affected, if at least the first 32 out of the 64 DCT coefficients of each (Y, C_r, C_b) component of the image block were selected to form the approximation image for the edge extraction task; in the second case, for texture ROI/non-ROI classification, the least required number of DCT coefficients was nine out of 64 for each image block component.

Following the above, the low resolution probabilistic network of the first level accepted at its input 96 (3×32) DCT coefficients. To compute the number of necessary DCT coefficients (which, in the selected form, are half of those computed for each block) we used the multiresolution analysis presented in Section IV-B (using a value of $\acute{o} = 0.0001$); we considered an initial set of the first 24 (3×8) DCT coefficients as the approximation image and sequentially added DCT coefficients, in multiples of eight coefficients, as detail images, until the obtained results were satisfactory enough. The training set was composed of 400 image blocks, 70 of which belonged to the ROI category; the network, therefore, included 400 pattern units. As a test set to examine the performance of the network we used the right half of the image shown in Fig. 7(a), as well as the whole image shown in Fig. 8(a). Finally, the computed outputs for all image blocks were collected over the whole images; the results are shown in Figs. 7(b) and 8(b), respectively. It can be easily verified that all significant edges have been detected and are included in these figures. In fact, the network was able to learn all edge information included in the DCT coefficients of the left half of the image; this knowledge was sufficient for successful generalization of the network performance in the

Table 1 Comparative Compression Ratio Table (ROI-JPEG versus JPEG)

ROI JPEG		WQF											
		70			75			80			85		
		max	min	avg	max	min	avg	max	min	avg	max	min	avg
QF	40	15.03	13.92	14.35	14.66	13.34	13.75	14.02	12.14	12.77	13.36	10.73	11.77
	45	14.18	13.12	13.71	13.85	12.79	13.16	13.28	11.83	12.25	12.68	10.5	11.32
	50	13.75	12.43	13.12	13.08	12.12	12.61	12.57	11.45	11.77	12.03	10.28	10.90
	55	13.43	11.8	12.63	12.64	11.53	12.15	12.01	11.09	11.35	11.52	10.09	10.53
	60	13.08	11.19	12.12	12.31	10.94	11.66	11.43	10.54	10.93	10.98	9.87	10.16
BASELINE JPEG		12.21	10.01	10.95	11.06	9.06	9.91	9.42	7.79	8.50	7.92	6.59	7.18

classification of the right half of the image in Fig. 7(a) and of the image in Fig. 8(a).

The second level of the proposed architecture was then used to select ROI's from the nonedge textured areas. In this case the network accepted 27 (3×9) DCT coefficients for each image block (which are only about 14% of the computed ones) and used the same training and test sets. The results of the network performance are shown in the classification maps of Figs. 7(c) and 8(c), which are constructed by combining the network outputs from both levels of the architecture. The success rate of the proposed architecture was very high; it was 96% in selecting ROI blocks and 98% in selecting background ones.

The proposed ROI-JPEG coding scheme was then applied to the original images, using the computed classification maps and corresponding quantization tables with WQF = 75 and QF = 50. The reconstructed images are shown in Figs. 7(d) and 8(d), respectively. It is easy to see that they are very good replicas of the original images.

The second application involves the coding for optimal storage and speedy retrieval of pathology images. These pictures are typically obtained from optical microscopes, with objectives ranging from 2.5× to 40× (typical eyepiece lens: 12.5×). The frames are photographed on slide format (35 mm) and subsequently scanned at resolutions ranging from 1200–2400 dots per inch (dpi). The images are initially stored in raw 24 bpp format, and are submitted in this form to the coding scheme already described. The frames appearing in this paper represent typical cases of pulmonary tuberculosis [31].

Figs. 9(a)–12(a) show four characteristic images that include regions of various shapes, which are of major importance for medical diagnosis. It can be seen, however, by testing the appearance of the images, that edge detection does not play a crucial role in the definition of ROI's, since the latter are basically textured; on the contrary, background seems to be rather noisy, including a lot of high frequencies. This example clearly indicates the fact that the proposed architecture is superior to any scheme which selects as ROI all blocks containing edges or contours, as well as the "interior" of the selected "edge" regions. The fact that extreme care should be taken for obtaining "closed" contours, which permit the definition of interior versus exterior areas, and that specific attention should be given to the choice of thresholding operations in edge extraction, leads to the conclusion that the use of edge detection for ROI selection is highly inadequate for the aforementioned image categories. Fig. 13(a) shows the edges extracted using standard Sobel and Prewitt operators on the image of Fig. 10; Fig. 13(b) shows the image blocks which have been selected as ROI, by testing whether each of them contained at least six pixels belonging to edges (approximately 9% of the block's pixels). It can be easily seen that this map greatly differs from the one shown in Fig. 10(b), which approximates the desired classification from a medical point of view. Consequently, we applied only the second level of the proposed architecture for classifying the image blocks to a ROI, or a non-ROI category, using only 27 (3×9) DCT coefficients (as in the previous experiment) as inputs to the network. The training set consisted of 30 images, resulting in 3000 training blocks, 750 of which belonged to the ROI category.

After training, the network performance was tested over the whole set of images. The results concerning the ef-

fectiveness of the network in classifying the images of Figs. 9(a)–12(a) are shown in the classification maps of Figs. 9(b)–12(b), which are constructed directly from the network output for each block of the images. The obtained success rate was very high, in this case as well; it was 97.13% in selecting ROI blocks and 97.90% in selecting background ones.

The ROI-JPEG coding scheme was then applied to the images of Figs. 9(a)–12(a), using the computed classification maps and corresponding quantization tables with WQF = 70 and QF = 50. The reconstructed images are shown in Figs. 9(c)–12(c), respectively. It can be easily observed, as was confirmed by visual assessment from the medical expert, that the reconstructed images constitute very good replicas of the original ones.

In the final set of our experiments we examined the compression ratios obtained by the proposed ROI-JPEG technique in all the above-described cases. Fig. 14 shows the average obtained compression ratios, for different values of the factors QF and ROI QF (WQF) factors which define the quantization tables used in the compression procedure. These ratios are also compared to the corresponding ratios that would be obtained, if the baseline JPEG compression algorithm were used instead of the ROI-JPEG one. The average ratios are also shown in Table 1, which also presents their minimum and maximum values, for each examined value of the quantization factors. It can be seen that an additional average compression of 20–25% was obtained in the above experiments (with WQF = 70–75 and QF = 50) compared to baseline JPEG. The percentage of the image blocks which have been classified as ROI varied from 17–55% with an average of 32% in the above images.

VI. Conclusions

An adaptive ROI-JPEG coding platform has been developed in this paper for effective lossy compression of medical images. The method uses a hierarchical neural network architecture, including probabilistic network models with multiresolution extensions, to adaptively select regions of interest in the images, following appropriate training with examples that an expert in the specific field provides and examines. The neural network architecture has been included as a separate subsystem in a coder based on the JPEG standard, controlling the quantization tables used for each block of the examined image.

The method is capable of achieving high compression ratios through the proposed local adaptivity of quantization and the corresponding source image coding. The adaptation of quantization is heavily psychophysically dependent, since training of the proposed neural network architecture has been based on decisions and knowledge of experts in the specific application domain. Since the algorithm uses the JPEG standard formulation, it can be easily implemented through a progressive or hierarchical procedure, where parts of the image are coded and then stored or transmitted sequentially. Moreover, since image edges, in the great majority of cases, belong to image ROI's, they are generally reconstructed without annoying ringing effects, particularly at low bit rates; edge enhancement and postprocessing can be also effectively applied only to the selected regions of interest.

Extension of the method to compress moving medical images, such as those used in ultrasonography, nuclear medicine, and angiography is possible by interweaving the proposed neural network architecture with an MPEG-like coding scheme [32]. Such experiments are now under investigation and constitute a topic of further research.

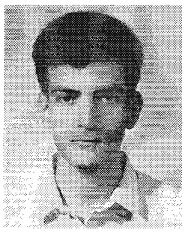
Acknowledgment

The authors wish to thank Dr. I. Panayiotides for providing medical data and for his assistance in examining them from a medical point of view, as well as for the visual evaluation of the obtained output images.

References

[1] S. Wong, L. Zaremba, D. Gooden, and H. K. Huang, "Radiologic image compression—A review," *Proc. IEEE*, vol. 83, pp. 194–219, Feb. 1995.
[2] R. D. Dony and S. Haykin, "Neural network approaches to image compression," *Proc. IEEE*, vol. 83, pp. 288–303, Feb. 1995.
[3] D. R. Hush and B. G. Horne, "Progress in supervised neural networks," *IEEE Signal Process. Mag.*, pp. 8–39, Jan. 1993.
[4] S. Haykin, *Neural Networks: A Comprehensive Foundation.* New York: Macmillan, 1994.
[5] E. Oja, "Principal components, minor components and linear neural networks," *Neur. Net.*, vol. 5, pp. 927–935, 1992.
[6] S. Y. Kung, K. I. Diamantaras, and J. S. Taur, "Adaptive principal component extraction (APEX) and applications," *IEEE Trans. Signal Process.*, vol. 42, pp. 1202–1217, May 1994.
[7] T. Kohonen, *Self-Organizing Maps.* Berlin: Springer-Verlag, 1995.
[8] I. Pitas, Ed., *Parallel Algorithms for Digital Image Processing, Computer Vision and Neural Networks.* New York: Wiley, 1993.
[9] Y. Pao, *Adaptive Pattern Recognition and Neural Networks.* Reading, MA: Addison Wesley, 1989.
[10] A. Ravichandran and B. Yegnanarayana, "Studies on object recognition from degraded images using neural networks," *Neur. Net.*, vol. 8, no. 3, pp. 481–488, 1995.
[11] D. Kontoravdis, S. Kollias, and A. Stafylopatis, "A two-phase connectionist approach to invariant picture interpretation," to be published in *Mathematics and Computers in Simulation: Special Issue on Neural Networks.*
[12] Z. Wang, C. Di Massimo, M. T. Tham, and A. J. Moris, "A procedure for determining the topology of multilayer feedforward neural networks," *Neur. Net.*, vol. 7, pp. 291–300, June 1994.
[13] J. A. S. Freeman and D. Saad, "Learning and generalization in radial basis function networks," *Neur. Computat.*, vol. 7, no. 5, pp. 1000–1020, Sept. 1995.
[14] D. F. Specht, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," *IEEE Trans. Neur. Net.*, vol. 1, Mar. 1990.
[15] T. Kohonen, "Improved versions of learning vector quantization," in *Proc. Int. Joint Conf. Neur. Net.*, 1990, vol. 1, pp. 545–550.
[16] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard.* New York: Van Nostrand Reinhold, 1993.
[17] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications.* New York: Academic, 1990.
[18] N. Panagiotidis and S. Kollias, "A progressive windowed JPEG coder for efficient image transmission," in *Proc. SPIE Europe. Symp. Adv. Net. and Services,* Amsterdam, 1995.

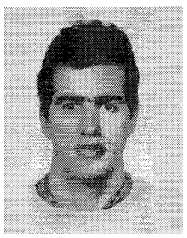
[19] S. Kollias and D. Anastassiou, "An adaptive least squares algorithm for the efficient training of multilayered networks," *IEEE Trans. Circ. and Syst.*, vol. 36, pp. 1092–1101, 1989.
[20] M. Morrison and Y. Attikiouzel, "A probabilistic neural network based image segmentation network for magnetic resonance images," in *Proc. IJCNN 92,* Portland, OR, 1992.
[21] R. P. Lippmann, "Neural network classifiers estimate Bayesian a posteriori probabilities," *Neur. Computat.,* vol. 3, pp. 461–483, 1992.
[22] J. Makhoul, "Pattern recognition properties of neural networks," in *Proc. IEEE-SP Workshop Neur. Net. Signal Process.,* Princeton, NJ, 1991, pp. 173–186.
[23] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math Stat.,* vol. 33, pp. 1063–1076, 1965.
[24] R. Fisher, *From Surfaces to Objects, Computer Vision and Three Dimensional Analysis.* New York: Wiley, 1989.
[25] S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," *Comput. Graph. and Image Process.,* vol. 4, pp. 104–109, 1975.
[26] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.
[27] I. Daubechies *et al.,* "Image coding using the wavelet transform," *IEEE Trans. Image Process.,* vol. 1, pp. 205–220, 1992.
[28] C. Hand *et al.,* "A neural-network feature detector using a multiresolution pyramid," in *Neural Networks for Vision, Speech and Natural Language,* R. Linngard, Ed. New York: Chapman and Hall, 1992.
[29] S. Kollias, "A multiresolution neural network approach to invariant image recognition," *Neurocomput.*, 1996.
[30] S. Kollias and D. Kalogeras, "A multiresolution probabilistic neural network for image segmentation," in *Proc. ICASSP-94,* Adelaide, Australia, 1994.
[31] J. Panayiotides, E. Protopappa, and G. Delides, "Nuclear morphometry as a prognostic factor in laryngeal squamous cell carcinomas," *Zentralblatt fur Pathologie,* vol. 139, pp. 221–224, 1993.
[32] A. Murat Tekalp, *Digital Video Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1995.



Nikos G. Panagiotidis (Student Member, IEEE) received the Diploma in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 1991. He is currently working on a Ph.D. dissertation at the NTUA.

He has participated in several research projects for the NTUA. His research interests include medical imaging systems, low bitrate video coding, applications of neural networks, and multimedia technologies. He has published six papers in refereed conferences and journals.

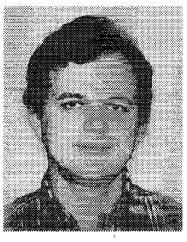
Mr. Panagiotidis is a member of the Technical Chamber of Greece.



Dimitris Kalogeras (Member, IEEE) received the Diploma and the Ph.D. degree in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, in 1990 and 1995, respectively.

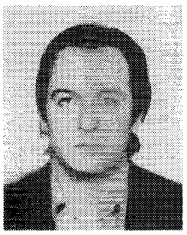
He is presently with the Computer Science Department at the NTUA. His research interests include high-speed data networks design and implementation. He has published eight papers in refereed conferences and journals. His research interests include low bitrate video coding, applications of neural networks, and applications of video transmission over various network architectures.

Dr. Kalogeras is a member of the Technical Chamber of Greece.



Stefanos D. Kollias (Member, IEEE) received the Diploma in electrical engineering from the National Technical University of Athens (NTUA), Greece, the M.Sc. degree in communication engineering from the University of Manchester Institute of Science and Technology, Manchester, U.K., and the Ph.D. degree in signal processing (computer science) from NTUA in 1979, 1980, and 1984, respectively.

He is currently an Associate Professor of Computer Science at NTUA. During 1987–1988 he was a Visiting Research Scientist in the Electrical Engineering Department of Columbia University, New York. His research interests include image processing and analysis, video coding, artificial neural networks, and multimedia applications. He has published more than 70 papers in international journals and conference proceedings.



Andreas Stafylopatis (Member, IEEE) received the Diploma in electrical and electronics engineering from the National Technical University of Athens (NTUA), Greece, and the Ph.D. degree in computer science from the University of Paris-Sud, Orsay, France, in 1979 and 1982, respectively.

He has been with the Department of Electrical Engineering at the NTUA since 1984, where he is now an Associate Professor. His research interests include parallel processing and computational intelligence.

Dr. Stafylopatis is a member of the IEEE's Computer and the Systems, Man and Cybernetics Societies. He is also a member of the Association for Computing Machinery, the European Neural Network Society, and the International Neural Network Society.