Optimization of Wavelet Decomposition for Image Compression and Feature Preservation

Shih-Chung B. Lo*, Huai Li, and Matthew T. Freedman

***Abstract*—A neural-network-based framework has been developed to search for an optimal wavelet kernel that can be used for a specific image processing task. In this paper, a linear convolution neural network was employed to seek a wavelet that minimizes errors and maximizes compression efficiency for an image or a defined image pattern such as microcalcifications in mammograms and bone in computed tomography (CT) head images. We have used this method to evaluate the performance of tap-4 wavelets on mammograms, CTs, magnetic resonance images, and Lena images. We found that the Daubechies wavelet or those wavelets with similar filtering characteristics can produce the highest compression efficiency with the smallest mean-square-error for many image patterns including general image textures as well as microcalcifications in digital mammograms. However, the Haar wavelet produces the best results on sharp edges and low-noise smooth areas. We also found that a special wavelet (whose low-pass filter coefficients are 0.32252136, 0.85258927, 0.38458542, and −0.14548269) produces the best preservation outcomes in all tested microcalcification features including the peak signal-to-noise ratio, the contrast and the figure of merit in the wavelet lossy compression scheme. Having analyzed the spectrum of the wavelet filters, we can find the compression outcomes and feature preservation characteristics as a function of wavelets. This newly developed optimization approach can be generalized to other image analysis applications where a wavelet decomposition is employed.**

***Index Terms*—Image compression, image feature restoration, neural network, optimization of wavelet, wavelet decomposition.**

I. Introduction

In the field of transform coding, discrete cosine transform (DCT)-based decomposition methods were developed extensively in the 1970s and 1980s. Most of these techniques are associated with block DCT [1]–[4]. However, several investigators have indicated that the use of full-frame DCT [5]–[7] can produce high compression efficiency with high data fidelity without blocky artifacts. This method is particularly appropriate for high-resolution large-sized images. Recently, the subband and discrete wavelet transform (DWT) have been widely used in image compression research [8]–[10]. The JPEG-2000 compression standard [11] is a result of the wavelet research activities in the past decade. Unlike the DCT, wavelet transform coefficients are partially localized in both spatial and frequency domains and form a multiscale representation of the image. In addition, wavelet transform coefficients possess orientation specificity. Since the wavelet transform has these attractive features, many efforts have been made to effectively encode transformed coefficients [12]–[14]. As a result, it has been shown that the wavelet transform, in many situations, obtained significantly greater compression results than those obtained from block DCT techniques in a course of image compression.

The wavelet decomposition is determined by one mother wavelet function and its dilation and shift versions. Since enormous numbers of wavelet filters are available, investigators often have a difficulty in selecting an optimal wavelet for a specific image procession application. Many technical issues relating to this research area remain unsolved. The early work on selecting wavelet bases can be found in [15]–[18]. The choice of optimal wavelet depends on different criteria in various applications. In [15], Tewfik *et al.* proposed a method for selecting a wavelet for signal representation based on minimizing an upper bound of the L^2 norm error in approximating the signal up to a desired scale. In [16], [17], Coifman and Wickerhauser developed an entropy-based algorithm for choosing the best wavelet basis that can achieve higher compression ratios with a generalization of wavelet packets, at the expense of increased processing time. Villasenor, *et al.* [18] derived a wavelet filter evaluation metric according to the filter impulse response and step response in addition to regularity. Based on this metric, several filters belong to the biorthogonal wavelet filter bank were selected for the decomposition process in a course of image compression.

In the field of image compression, one major criticism in the evaluation of reconstructed images is that investigators usually present a global measure of a large image rather than a local quantification of a specific image pattern. The former only provides overall performance of a compression technique on an image. In many applications, particularly in medical imaging, a local image pattern can be of major concern. In such a case, the performance at the region of interest (ROI) should be weighted much higher than that of other places. Our research goal is to

Manuscript received December 31, 2000; revised April 4, 2003. This work was supported in part by an the National Institutes of Health (NIH)/National Cancer Institute (NCI) under Grant RO1CA79139 and in part by the U.S. Army under Grant DAMD17-01-1-0267. The content of this paper does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. W. Vannier. *Asterisk indicates corresponding author*.

*S.-C. B. Lo is with the Center of Imaging Science and Information Systems, Radiology Department, Georgetown University Medical Center, 2115 Wisconsin Avenue, N.W., Suite 603, Washington, D.C. 20007 USA (e-mail: lo@isis.imac.georgetown.edu).

T. Freedman is with the Center of Imaging Science and Information Systems, Radiology Department, Georgetown University Medical Center, Washington, D.C. 20007 USA.

H. Li was with the ISIS Center, Radiology Department, Georgetown University Medical Center, Washington, D.C. 20007 USA. He is now with the High Performance Computing and Information Office, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892 USA.

Digital Object Identifier 10.1109/TMI.2003.816953

investigate which wavelet filter produces the best compression result for a given image pattern. In our experiment, we isolated various types of ROIs on several medical images for the evaluation of data fidelity and compression ratios using wavelet decomposition techniques. Although clustered calcifications and masses are primary cancer signs in mammography, no major attention is paid to masses if calcifications are well preserved. This is because calcifications will be first affected before any degradation of masses can be observed while increasing the compression ratio in a typical compression experiment. Lucier *et al.* [19] reported some compression results using the Haar wavelet in mammography, however, several investigators [20]–[22] chose to use the Daubechies wavelets for compression of digital mammography. The comparison between different wavelets is one of the research goals in this study.

In this paper, we employed a linear convolution neural network system that possesses an embedded wavelet operation to serve as a research framework. Through a controlled back-propagation (BP) algorithm, the neural network is capable of searching for an optimal wavelet that minimizes quantization errors and at the same time produces the highest compression efficiency. This newly developed convolution neural network method can also be extended to evaluate various wavelets in preserving defined image features.

The remainder of this paper is organized as follows. In Section II, DWT is briefly reviewed and a wavelet-based convolution neural network is described. In addition, migration from a wavelet kernel to another, which is embedded in the searching method of the neural network system, is also presented. Section III describes the experimental method used to evaluate the study objects. The results are given in Section IV. Section V discusses the results of wavelet search with the image patterns and characteristics of optimized wavelets. Section VI summarizes the technical achievements and their implications in the field of medical image compression.

II. Algorithm Development

A. Two-Dimensional (2-D) Wavelet Decomposition

Following Mallat's 2-D wavelet analysis [9], the 2-D scaling function is composed of two one-dimensional (1-D) scaling functions in both directions if they are separable

$$\phi(x,y) = \phi(x)\phi(y) \tag{1}$$

where $\phi(x)$ is a scaling function. The associated 2-D wavelets are defined as

$$\psi^H(x,y) = \phi(x)\psi(y) \tag{2}$$

$$\psi^V(x,y) = \psi(x)\phi(y) \tag{3}$$

$$\psi^D(x,y) = \psi(x)\psi(y) \tag{4}$$

where $\psi(x)$ is the 1-D wavelet corresponding to the 1-D scaling function. Using the subband coding algorithm, the 2-D wavelet transform (2-D DWT) of a matrix has four compartments

$$W_{\mathrm{LL}}(f(x,y)) = \sum_{u,v}[(f(x,y)h(u-2x,0))h(0,v-2y)] = \sum_{u,v}[f(x,y)h_{\mathrm{LL}}(u-2x,v-2y)] \tag{5}$$

$$W_{\mathrm{LH}}(f(x,y)) = \sum_{u,v}[(f(x,y)h(u-2x,0))g(0,v-2y)] = \sum_{u,v}[f(x,y)h_{\mathrm{LH}}(u-2x,v-2y)] \tag{6}$$

$$W_{\mathrm{HL}}(f(x,y)) = \sum_{u,v}[(f(x,y)g(u-2x,0))h(0,v-2y)] = \sum_{u,v}[f(x,y)h_{\mathrm{HL}}(u-2x,v-2y)] \tag{7}$$

$$W_{\mathrm{HH}}(f(x,y)) = \sum_{u,v}[(f(x,y)g(u-2x,0))g(0,v-2y)] = \sum_{u,v}[f(x,y)h_{\mathrm{HH}}(u-2x,v-2y)] \tag{8}$$

where h and g functions are the low-pass and high-pass filters of the subband decomposition with condition

$$g(u) = (-1)^u h(1-u). \tag{9}$$

The low-pass filter, h, also must satisfy two criteria to construct the orthogonal basis of compactly supported wavelets [8], [9]. For simplicity, we also use g_u and h_u to replace $g(u)$ and $h(u)$, respectively, in this paper.

1)
$$\left[\sum_u h_{2u}\right] - \frac{\sqrt{2}}{2} = \left[\sum_u h_{2u+1}\right] - \frac{\sqrt{2}}{2} = 0; \tag{10}$$

2) should be orthogonal; this means that
$$\left[\sum_u h_u \times h_{u+2n}\right] - \delta_{u,u+2n} = 0 \tag{11}$$
where $\delta_{i,j}$ is Dirac delta function and n is an integer. However, high degree of regularity and high degree of vanish moments were not imposed in this study. From a compression point of view, these factors are not as significant as criteria 1) and 2) [18]. Typically, those filters performing perfect reconstruction are eligible in a data compression scheme. With a single low-pass filter system, the criterion of perfect reconstruction is the same as that in (11) because the synthesis filter is the reflection function of the analysis filter. For simplicity, we will use a single low-pass filter to perform wavelet transform in the following algorithm development. The 2-D filters at the second forms of (5)–(8) are the vector products of h and/or g filters. The relationship between high-pass and low-pass filters makes the unification of the four sets of decomposition possible as shown in Section II-D.

According to wavelet theory, it is known that given a set of h, one can calculate the Fourier transform of the scaling and wavelet functions as follows:

$$\Phi(w) = H_0\left(e^{iw/2}\right)\Phi\left(\frac{w}{2}\right) \tag{12}$$

$$\Psi(w) = H_1\left(e^{iw/2}\right)\Phi\left(\frac{w}{2}\right) \tag{13}$$

where H_0 and H_1 are Fourier transforms of h and g filters, respectively. Hence, both the scaling and wavelet functions can be obtained through infinite recursion by using (12) and (13), respectively.

B. Construction of a Convolution Neural Network Using Wavelet Decomposition

The artificial neural network described in this paper is based on the linear convolution process which is used in subband and wavelet decomposition techniques. Each wavelet processing in the neural network performs exactly the same as the conventional wavelet transform given in (5)–(8). Our approach is to employ the convolution neural network [23] that is capable of searching for the most suitable wavelet kernel through an image compression scheme. A part of the searching algorithm would minimize error and simultaneously achieve the highest compression efficiency during the course of compression and decompression processes. In order to match the subband decomposition, several characteristics of the neural network must be established: 1) no hidden but one output layer is used; 2) local connection through the convolution process rather than fully connected nets is employed; 3) the convolution process must be reversible; and 4) convolution kernels are wavelet based. In order to study the data fidelity, we add a quantization to the compression process; therefore, the image cannot be fully reconstructed after the decompression process. The differences between original and reconstructed images do not result from the inverse transformation but rather because of the inaccuracy of the quantized transform coefficients. Fig. 1 shows the structure of the neural network using quantized transform coefficients as the targets.

Minimization of quantization errors was not the only issue in our technical consideration. The method of minimizing the entropy must also be taken into account for the optimization in a course of data compression. We combined both issues by multiplying the mean square error (mse) function with an imposed entropy reduction function. The cost (error) function for searching the optimal wavelet kernel in the neural network becomes

$$Ef(i,j) = Z(QT(i,j)) \times \frac{[T(i,j) - QT(i,j) \times q]^2}{2} \quad (14)$$

where $T(i,j)$ is the wavelet transform of $I(x,y)$, $QT(i,j)$ is the quantized transform coefficient at pixel (i,j), and $Z(QT(i,j))$ is the entropy reduction function for a set of quantization coefficients. The quantization was performed as follows:

$$QT(i,j) = \text{NINT}\left[\frac{T(i,j)}{q}\right] \quad (15)$$

q is a quantization factor and NINT represents the nearest integer operation. The inverse quantization is simply the multiplication of the quantization factor and the quantized coefficients (i.e., $QT(i,j) \times q$). The entropy reduction function is given as

$$Z(QT(i,j)) = \begin{cases} 0 & \text{for } QT(i,j) = 0, \\ 1, & \text{for } |QT(i,j)| = 1 \\ F(n,q), & \text{for } |QT(i,j)| = n \end{cases} . \quad (16)$$

$F(n,q)$, which is a ramp function and is somewhat inversely proportional to the quantized integer n. The value of the ramp function should always be smaller than one.

The reason to design the entropy reduction function for a fixed quantizer, q, using (16) is threefold: 1) since most low value coefficients ($-0.5q < T(i,j) < 0.5q$) are associated with noise when q is not a very large value, there is no need to backpropagate errors from the output node possessing quantized value zero in the neural net; 2) the more number of small quantized values there are, the lower the ensemble entropy; and 3) the probability of turning a large quantized value into a small quantized value is very low, therefore, errors backpropagated from high quantized value should be less emphasized as compared with low quantized value (e.g., one, two, or so). When q is very small, the quantization error is in the range of global image noise. In this case, the neural network would rely on the guidance of Z function to search for a wavelet filter that produces more number of small transform values. The success of this cost (i.e., error) function design is depicted in our experiment shown in Section IV.

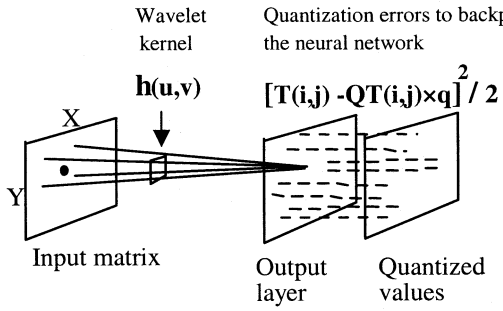

Fig. 1. A wavelet-based neural network system that was used to seek an optimal wavelet for minimization of the quantization errors. $T(i,j)$ and $QT(i,j)$ denote transform and quantized coefficients in the high-frequency domain, respectively.

Based on the neural network shown in Fig. 1, we can seek an optimal convolution kernel. The specific searching algorithm is given in Section II-C. Section II-D shows a method to conduct orthogonal wavelet decomposition without using the high-pass filter. Hence, the low-pass filter is the only kernel to process the image data into the four compartments in each level of the wavelet decomposition. In practice, the kernel directly suggested by the neural network in each epoch may not be a wavelet kernel. Section II-E provides algorithms that modify the kernel to fulfill the requirements of wavelet kernel. Through this process, we can find a wavelet that produces the lowest quantization errors with the lowest entropy of the quantized transform coefficients.

C. Signal Propagation Through Convolution Process and Searching Methods in the Neural Network

The signal propagation from input layer to output layer involving convolution computation is given as [24]

$$T_c(i,j) = K_c(i,j) \otimes S(i,j) \quad (17)$$

where $\otimes$ represents convolution, $S(i,j)$ is the original image, subscript c denotes the wavelet compartment number, and $K_c(i,j)$ is the convolution kernel for compartment c. For the wavelet decomposition, the relationship between $K_c(i,j)$ and the wavelet filters (i.e., h and g filters) will be given in Section II-C and Section II-D

Since we treat the wavelet transform as a locally connected neural network, the well-known BP method can be used as a searching process by altering the 2-D convolution kernels in each epoch [25]. For each composed signal, a linear function instead of a typical sigmoid function in the neural network system

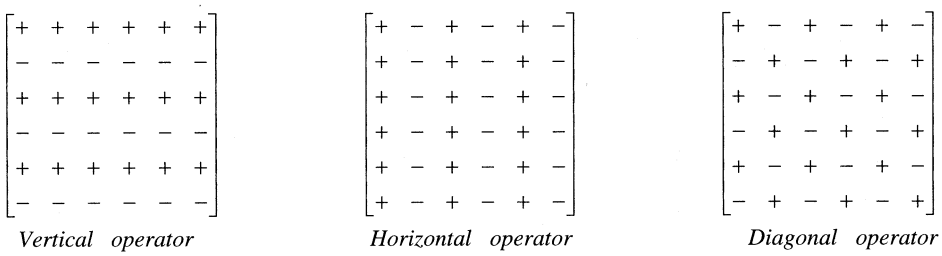

Fig. 2. Three matrices used for the cross product precalculation.

is used in this process. The updated kernel suggested by BP in the neural network is given by

$$K_c(u,v)[t+1] = K_c(u,v)[t] + \eta \sum_{i,j} \mu(i,j) S(u-i, v-j) + \alpha \Delta K_c(u,v)[t] \quad (18)$$

where t is the iteration number during the searching, α is the gain for the momentum term received in the previous learning loop, η is the gain for the current weight changes, and μ is the weight-update function which is given by

$$\mu(i,j) = \frac{\partial Ef}{\partial K_c(u,v)}. \quad (19)$$

Since a 2-D DWT is composed of four wavelet decompositions, there are four associated convolution kernels to be updated simultaneously. The association between low-pass and high-pass filters, as shown in (9), is a necessary constraint of compact support in orthogonal wavelets. In order to preserve this property, we rearrange the decomposition method; hence, only a single kernel is needed to perform the 2-D DWT as demonstrated in Section II-D.

D. Unification of the Four Decompositions in 2-D DWT

Using (6) as an example to rewrite the decomposition equation by replacing the g with the h filter, we have

$$W_{\mathrm{LH}}(f(x,y)) = \sum_{u,v} \Big[(f(x,y)h(u-2x,0))(-1)^v \cdot h(0, 2y+1-v) \Big] \quad (20)$$

or

$$\begin{aligned} W_{\mathrm{LH}}(f(x,y)) &= \sum_{u,v} [(((-1)^v f(x,-y))h(u-2x,0)) \cdot h(0, v-2y)] \\ &= \sum_{u,v} [(((-1)^v f(x,-y))h_{\mathrm{LL}}(u-2x, v-2y)] \\ &= \sum_{u,v} [f_{\mathrm{LH}}(x,y) h_{\mathrm{LL}}(u-2x, v-2y)]. \end{aligned} \quad (21)$$

Converting (7) and (8) to use the 2-D low-pass filter as the kernel is a matter of changing the orientation from y to x direction. These conversions also indicate that one can use a single 2-D filter to compute the four quadrants of the 2-D wavelet transform by flipping the matrix position in x and/or y directions and alternating the sign of the flipped matrix corresponding to the directions.

The alternated sign of the source vector makes the convolution operation unconventional. A precalculation method, that involves a cross product of two vectors, can be employed: flipping the data sequence of an image is the first vector and the second vector is fixed and composed of $+1$ and -1. An example of 1-D precalculation steps for tap-6 kernel prior to the convolution operation is given as

$$\begin{aligned} &\text{original data sequence :} a_1, a_2, a_3, a_4, a_5, a_6 \\ &\text{flipped data sequence :} a_6, a_5, a_4, a_3, a_2, a_1 \\ &\text{resultant data sequence :} a_6, -a_5, a_4, -a_3, a_2, -a_1 \end{aligned}$$

In the case of 2-D, three matrices associated with horizontal, vertical and diagonal decomposition for the second matrix in the precalculation process are shown in Fig. 2. With this precalculation (or cross product of two matrices), only the low-pass filter $h_u h_v$ (h_u in 1-D) is needed for the final wavelet transform operation.

Nevertheless, the resultant matrix of this precalculation or the cross product of two matrices must be held in the computer memory to facilitate the computation for forward convolution and the corresponding BP. After precalculation, the size of the intermediate images is $(k/2 \times k/2)$ times the original image size. The factor of $1/2 \times 1/2$ is due to the 1/2 down sampling two-dimensionally in a conventional forward wavelet transform. Fig. 3 shows the decomposition scheme that can search for an optimal wavelet kernel using the convolution neural network. The largest three blocks shown in Fig. 3 are the intermediate images $S_o(xk/2, yk/2)$.

Based on each precalculated image $S_o(xk/2, yk/2)$ described earlier, (18) can be rewritten for updating the 2-D wavelet kernel

$$K(u,v)[t+1] = K(u,v)[t] + \eta \sum_{i,j} \mu(i,j) S_0\left(u - \frac{xk}{2}, v - \frac{yk}{2}\right) + \alpha \Delta K(u,v)[t] \quad (22)$$

where index $i = 0, 1, \ldots (k-1)^2$ corresponds to the subimage of S_o matched to the kernel size. The perfect reconstruction criterion of the new filter may not be self-sustained with each updated version. Equation (22) represents the updated kernel suggested by the backprogapation, these values require a conversion to become a new wavelet kernel $h'_u h'_v$. Assuming the wavelet filter is a 2-D vector (i.e., $h_u h_v = h_v h_u = h_{\mathrm{LL}}$, where $u \& v = 0, 1, 2, \ldots k-1$), there are only k number of free parameters that must be updated for each wavelet process. A solution to satisfy the wavelet constraints and to make $h'_u h'_v$ approximately equal to $K(u,v)[t+1]$ is given in Section II-E.

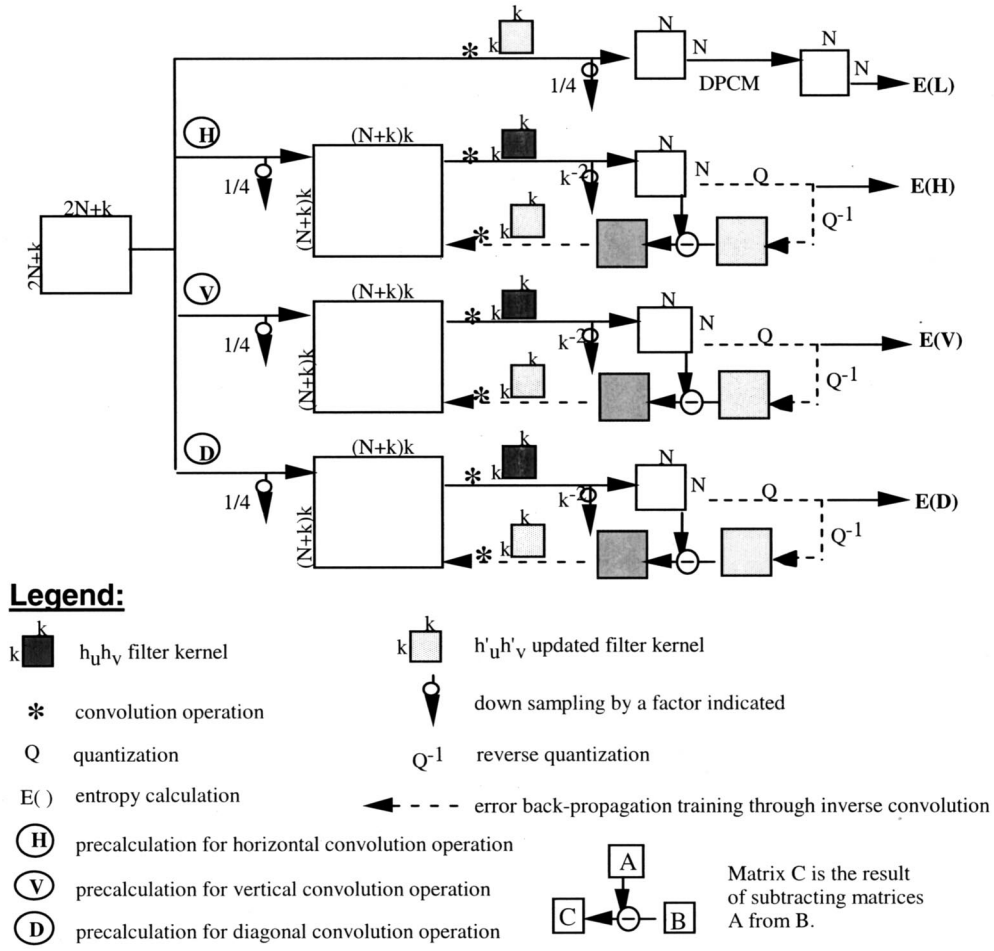

Fig. 3. The wavelet-based decomposition scheme coupled with the quantization process. The whole process is conducted in the convolution neural network and is designed to search for an optimal wavelet kernel.

E. Converting the Kernel Suggested by the Neural Network to Fulfill Requirements of a Wavelet Filter

Since (22) only provides an approximation that the changes of weights can produce a lower value for the defined error function, Ef, we ought to further convert the changes into a new wavelet kernel. Let us assume that there exists a set of h'_u so that the updated 2-D version of the wavelet filter is very close to $K(u, v[t+1])$. A function based on the square difference is used in the derivation

$$f(h'_u) = \sum_{u,v} \left(h'_u h'_v - K(u,v)[t+1]\right)^2. \tag{23}$$

Here, we intend to minimize the function, f, subject to the constraint equations. The Lagrangian multiplier method can be employed to solve this problem by combining f and constraint equations

$$df(h'_u) + \sum_p \lambda_p dC_p(h'_u) = 0 \tag{24}$$

where d represents the differentiation operation of a function and λ_p is the Lagrangian multiplier for the corresponding constraint equation, $C_p(h'_u) = 0$, referred to (10) and (11). Using this approach, we can obtain a set of h'_u while f is also minimized.

III. Materials and Experimental Methods

A. Mammographic Calcifications as Test Subjects

A database consisting of 45 mammograms was used in the study. Of these mammograms, 38 contain biopsy proven clustered microcalcifications. A total of 220 microcalcifications were embedded in 41 clusters. All 45 mammograms were digitized by a LumyScan (model 150) film digitizer (Lumysis Sunnyvale, CA) with spot size of 0.1 mm. Each patch of 32×32 pixels (i.e., an area of 3.2 mm $\times$ 3.2 mm) with its center at the peak value was isolated for the study of quantization impact on microcalcifications. Note that typical sizes of microcalcifications range from 0.2 to 1.0 mm. It is important to isolate this specific image pattern, otherwise the neural network search could be out of focus and turn into a failure search. Fig. 4 shows several examples of the clustered microcalcifications extracted from the mammograms.

The processes of searching optimal wavelet kernels for original mammograms and microcalcification patches were conducted in separate studies. Each image patch was decomposed by a three-level wavelet transform. Quantization values were q,

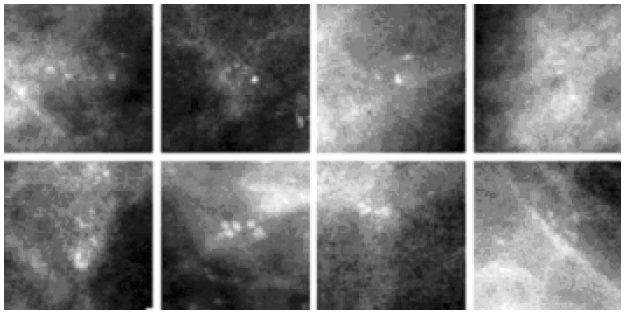

Fig. 4. Samples of clustered microcalcifications extracted from the mammograms.

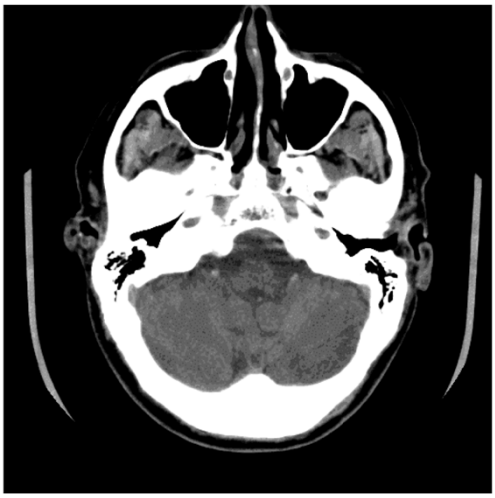

Fig. 5. A CT head image that was one of the six CT samples.

$q/2$ and $q/4$ for decomposition of high frequency coefficients on levels one- two, and three, respectively. For each searching epoch, the mse between the original and decompressed images as well as %zeros (i.e., number of zeros/total number of pixels) were computed. Since the value of %zeros is a dominated factor to gain a high compression ratio, it was used as a primary index for the evaluation of compression efficiency in every searching epoch.

In order to demonstrate each wavelet performance, we sorted the first coefficient h_0 of the low-pass filter associated with the mother scale function as the horizontal scale because the searching epoch could not represent the wavelet being used as shown in Fig. 6. All h_0 values are greater than $-0.146\,446\,609\,4$ and smaller than 0.852 555 339 05. The corresponding h_1 values are greater than 0.353 553 39 and smaller than 0.852 555 339 05. Those h_1 values, which are greater than $-0.146\,446\,609\,4$ and smaller than 0.353 553 39, have corresponding conjugate values in the former set and can be ignored.

The isolated 220 microcalcification patches were also used for the feature preservation study. The 2-D profiles of microcalcifications and their nearby areas (i.e., the areas that are not included in the microcalcification profiles but within the isolated patch 32×32 pixels) were evaluated separately with the neural network searched wavelets. In addition, features of the microcalcifications were computed to evaluate their changes. The features of microcalcifications used in the evaluation study are as follows:

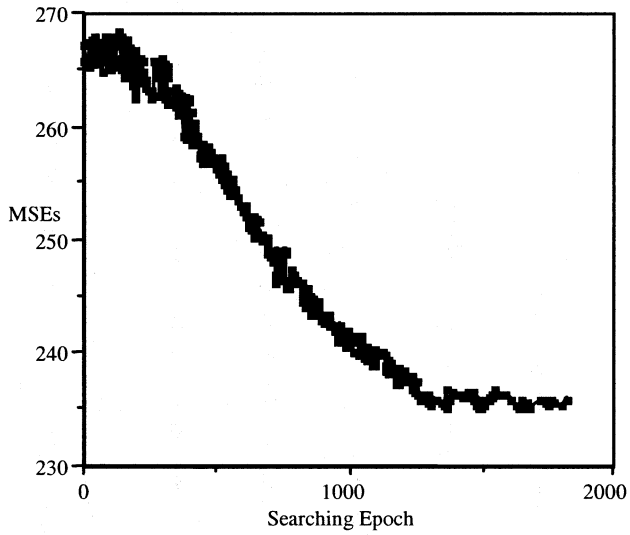

Fig. 6. MSEs were decreased during the neural network search on 220 microcalcifications.

1) peak value, P;
2) contrast, $\mathrm{C} = \mathrm{P} - m_b$; where m_b is the average background value which is the immediate boundary of the microcalcification profile;
3) signal-to-noise-ratio (SNR), $\mathrm{PSNR} = \mathrm{C}/\sigma_b$; where σ_b stands for the standard deviation of the background;
4) area, A, occupied by the 2-D microcalcification profile.

B. Other Test Subjects

We also use the searched wavelet spectrum obtained in the microcalcification study to test on the Lena image and the selected 45 mammograms. The two types of images were selected to represent general images and X-ray medical images, respectively. We also randomly selected six computed tomography (CT) head images for the study. We chose the CT head images to study the compression efficiency in skulls, bones, and their edges. The wavelet processing and compression operation were only applied to the skulls, bones and their edges, about 0.2 cm inside the soft tissue area. Fig. 5 was one of the selected CT head images.

IV. Results

A. Study Results in Mammographic Calcifications

In the neural network searching process, the mse was not the only factor to be considered; the entropy reduction function was another factor that drove the neural network to perform a search. In the first neural network experiment, we found that both the %zero and mse changed very little with a low quantization factor ($q = 16$). This situation was improved when $q = 64$. We found that %zeros did not change much until h_0 was greater than 0.6 in the microcalcification study. Fig. 6 shows the original searching steps versus mse's. The figure indicates that mse's moved toward smaller values using the convolution neural network searching mechanism. It was

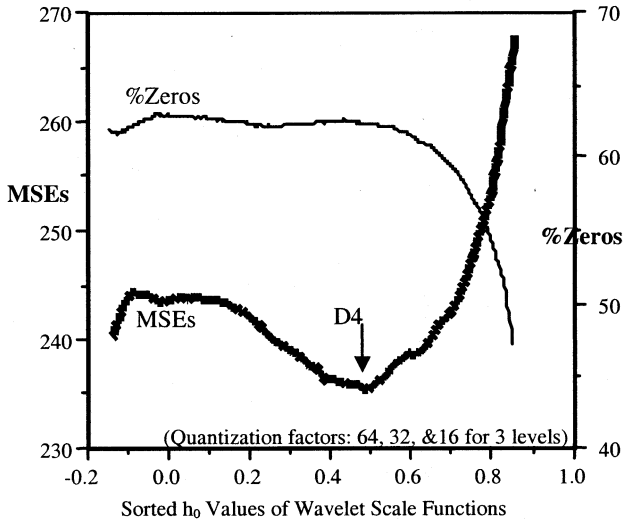

Fig. 7. Decomposition performance of wavelets on the 220 microcalcification patches.

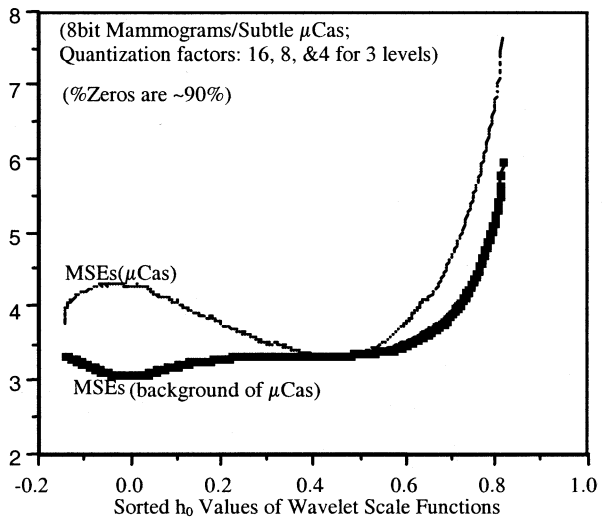

Fig. 9. Decomposition performance of wavelets on 220 microcalcification profiles and background (8-bit image).

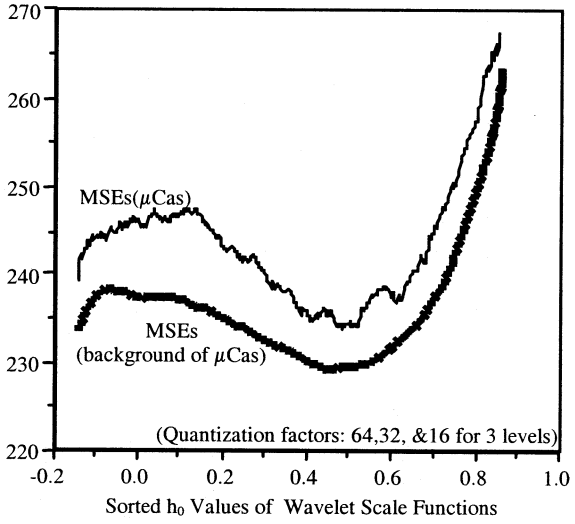

Fig. 8. Decomposition performance of wavelets on 220 microcalcification profiles and background.

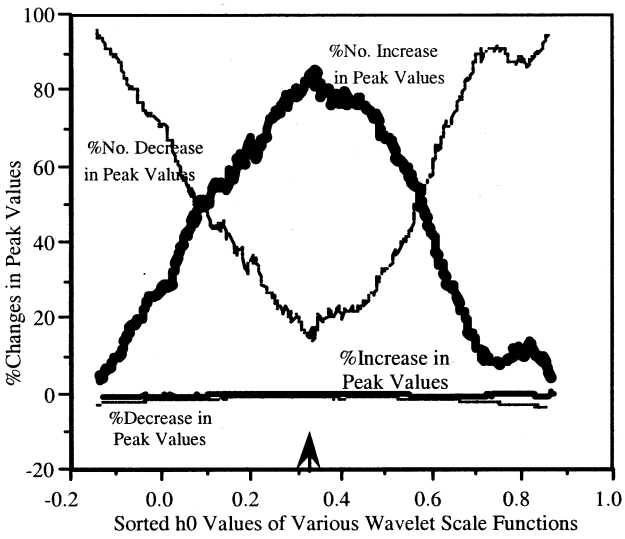

Fig. 10. Peak value changes due to quantization effects on the wavelet domain for microcalcifications.

difficult to analyze what was going on in each searching step.microcalcification profiles suffered higher mse's than their background areas as indicated in Fig. 8. We then decided to sort the wavelets as a filter bank or a filter group based on the ascending order of their first coefficients. Fig. 7 replots the mse's as well as %zeros versus the h_0 values in ascending order representing every searching epoch. It is interesting to see that the Daubechies wavelet ($h_0 = 0.482\,962\,91$) produces the lowest mse in Fig. 7. In the same study, we separated each patch into 1) microcalcification profile and 2) background. We found that

In one study, we preprocessed all the digital values in microcalcification patches such that each datum on the image was rounded from 12-bit to 8-bit prior to the quantization study. This part of the study was intended to simulate when mammograms were digitized at 8-bit per pixel [26]. The quantization factor would have an effect 16 times greater (i.e., 2^4) on an 8-bit image than that on a 12-bit image. We used $q = 16$ for 8-bit mammograms and found that the curves of mse's and %zeros were significantly changed as compared with the study using 12-bit mammograms. Although the largest quantization factor of 16 was applied to 8-bit mammograms, its effect was equivalent to $q = 256$ applied to 12-bit mammograms. Fig. 9 shows that the Haar wavelet ($h_0 = 0.0$) performs the highest and the lowest mse's for 2-D microcalcification profiles and their backgrounds, respectively. However, the Daubechies wavelet performs in an opposite way. This is because the Haar wavelet can produce a lower entropy in low-noise smooth areas.

The results of the microcalcification evaluation study based on quantized wavelet coefficients are shown in Figs. 10–13. In fact, the evaluation was performed with an identical experimental condition as that in Fig. 9. However, microcalcification features were measured instead of mse's and %zeros. Note that the percent numbers decrease in peak values, contrast and SNR were shown in negative values. In other words, the lower the percent number decrease value is, the more microcalcifications

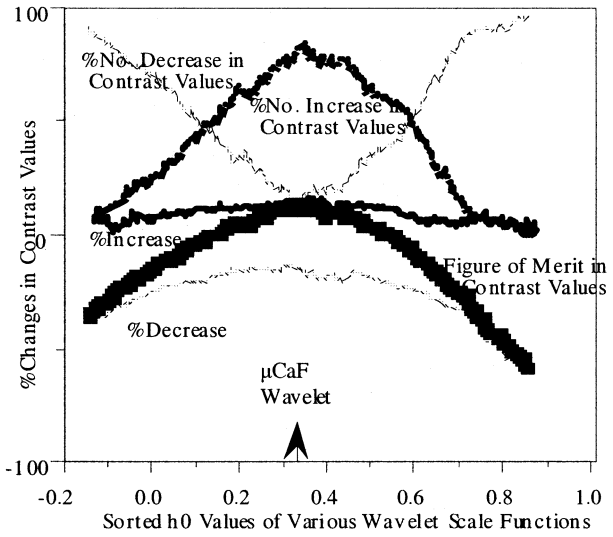

Fig. 11. Contrast changes due to quantization effects on the wavelet domain for microcalcifications.

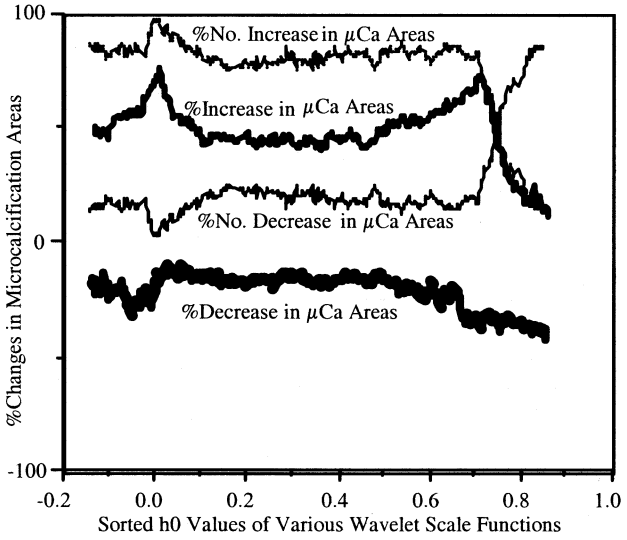

Fig. 13. Percent changes in microcalcification profiles due to quantization effects on the wavelet domain.

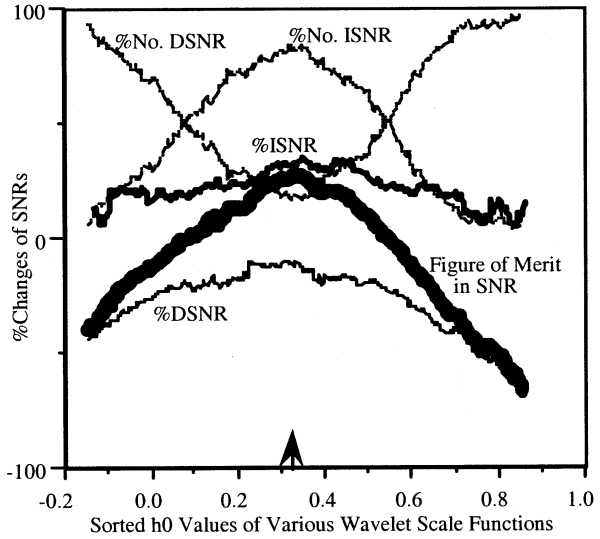

Fig. 12. PSNR changes due to quantization effects on the wavelet domain for microcalcifications.

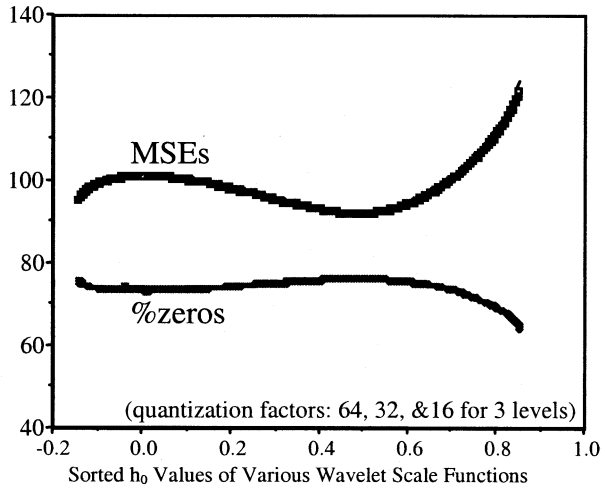

Fig. 14. Decomposition performance of wavelets on mammograms ($q = 64$).

involving negative changes. The figure of merit (FOM) for each measure was a composed value given by

$$\text{FOM} = \big(\%\text{no. decrease} \times \%\text{decrease} + \%\text{ no. increase} \times \%\text{increase}\big) \times 100. \quad (25)$$

As indicated in Fig. 10, the peak values had very little changes. However, the percent in number increasing in peak values, contrast values and SNRs of microcalcifications had approximately the same distribution in Figs. 11–13. The highest FOMs in all three measures occurred at the wavelet with the low-pass filter coefficients: (0.322 521 36, 0.852 589 27, 0.384 585 42, $-0.145\,482\,69$) which is marked with an arrow sign on the sort h_0 axis of the figures. We call this wavelet a microcalcification friendly wavelet or μCaF wavelet for short. Fig. 13 shows minor percent area changes of microcalcification profiles from 0.2 to 0.6 of h_0 values. These effects were not observed when a low quantization factor was used.

B. Study Results in the Mammograms and Lena Image

In the mammogram study, the neural network process in searching for the best wavelet kernel was random and no minimum of mse could be found. We then used the same set of the wavelet spectrum obtained in the microcalfications study to test on the mammogram images. The same method was also used to investigate the performance of the wavelets on the Lena image. Figs. 14 and 15 show the curves of mse's and %zeros against the sorted h_0 values in the selected 45 mammograms and the Lena image, respectively. In the both figures, the Daubechies wavelet ($h_0 = 0.482\,962\,91$) and its nearby wavelets produce slightly higher %zeros than the rest of wavelet spectrum in the mammograms and the Lena image. However, the lowest mse's were located at the Daubechies wavelet.

C. Study Results in Bones and Edges of the CT Head Images

We also applied the same set of the wavelet spectrum to test their compression performance on isolated CT bone regions. Only skulls, bones as well as edges between bones and soft

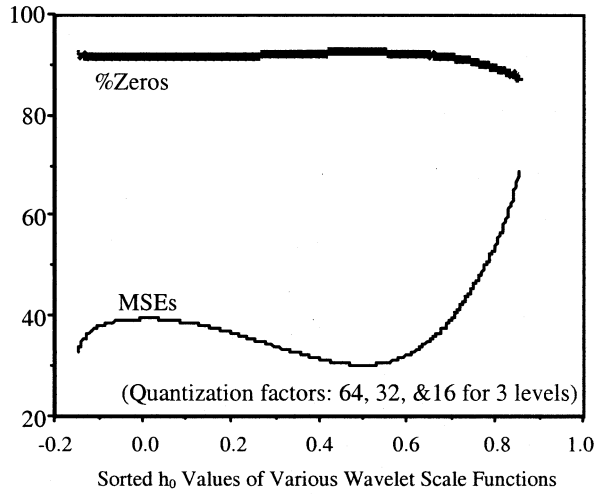

Fig. 15. Decomposition performance of wavelets on the Lena image ($q = 64$).

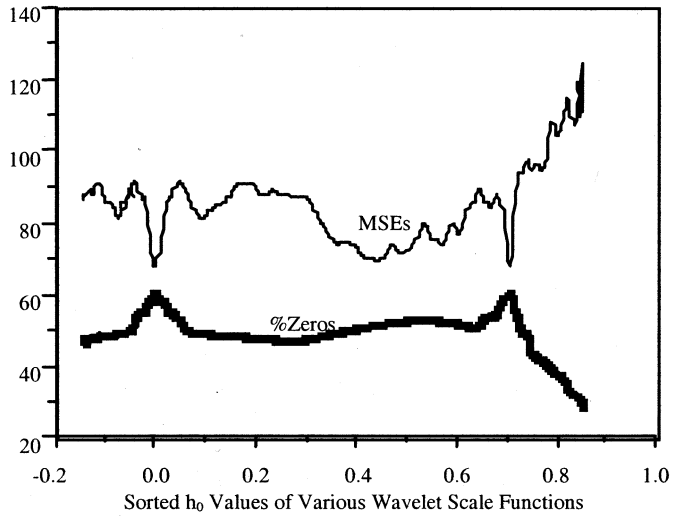

Fig. 16. Decomposition performance of wavelets on bones and edges ($q = 64$) using the six CT head images.

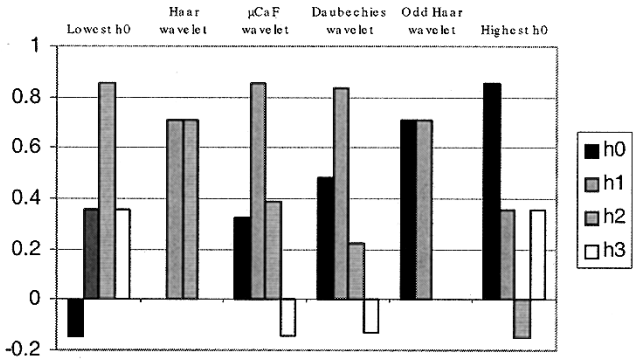

Fig. 17. Low-pass filters of several interesting wavelets.

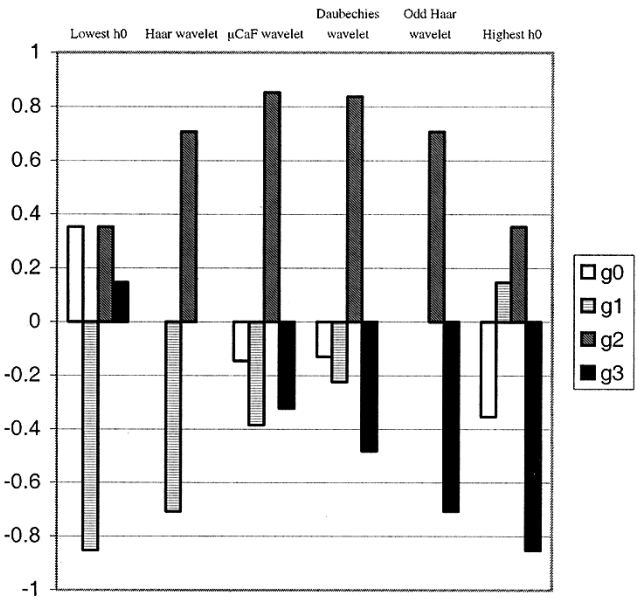

Fig. 18. High-pass filters of the same wavelets.

tissues on the six CT images were the subjects in the evaluation. Comparing the kernel size (4×4 pixels) to the six CT head bone regions, the latter contains much larger data. The mse and %zero versus the sorted h_0 values were shown in Fig. 16 which came out very differently from the other sessions of this research. In this particular test, the Haar wavelet greatly outperforms other wavelets. The odd Haar is a conjugate version of the Haar wavelet in the spectrum of Fig. 16. In terms of data decomposition, the two versions of the Haar wavelet would produce identical results.

V. Discussion

Having observed the above results, it is technically interesting to review the spectrum of wavelets mentioned. The low-pass h and the high-pass g filters of the wavelets are shown in Figs. 17 and 18, respectively. Note that the μCaF wavelet is the same wavelet marked on the horizontal axis in Figs. 10–12. Attention should be paid to the g filters since they are responsible for decomposing high frequency coefficients regardless the quantization. Essentially, the g filter performs calculation involving the positive weight multiplied by the center pixel value plus the adjacent pixel values on the two sides multiplied by the negative weights of the g filter. The Daubechies wavelet has quite balanced negative terms at the two sides of the positive weight and the sum of negative weights is negatively equal to the positive weight. The latter is a constraint in all wavelet filters anyway. In addition, the absolute value of $g_1(= -h_2)$ or $g_2(= h_1)$ should be reasonably large, which would maintain the low-pass and the high-pass characteristics for h and g filters, respectively. In fact, those wavelets near the Daubechies wavelet including the μCaF wavelet possess this property. From the signal processing point of view, these balanced weights in a filter are very important characteristics to create low entropy values for general image textures. We suspect that this property may have something to do with so called "high regularity" in the wavelet theory.

In short, we found that the main reason that a wavelet filter can produce a low entropy for a set of data is because the weight sum of the g filter is zero. For a general data sequence, the g filter can perform even better when

a) the absolute value of $g_1(= -h_2)$ or $g_2(= h_1)$ is much larger than that of other weights;

b) the opposite signed weights are evenly distributed at the two sides of g_1 or g_2.

For low-noise smooth areas, the Haar wavelet may slightly outperform the others. For sharp edges, the Haar wavelet would greatly outperform the others, as depicted in Fig. 16.

It was interesting to find out that the $\mu\mathrm{CaF}$ wavelet with

$$(h_u, u = 0, 1, 2, 3) = (0.322\,521\,36, 0.852\,589\,27, 0.384\,585\,42, -0.145\,482\,69)$$

results in the highest fidelity of the features. Fig. 8 provides evidence showing mse's of 2-D microcalcification profiles and background gradually merge from the Haar to the Daubechies wavelets. Since the contrast and PSNR values are computed using the peak and background values of the microcalcifications, the optimization of these measures should occur somewhere between the Haar and the Daubechies wavelets. It happens to be at the $\mu\mathrm{CaF}$ wavelet in the spectrum.

In the field of compression, it is known that the higher the compression ratio is, the higher the amount of error is generated in the decompressed image. However, through these studies we discovered a new phenomenon associated with these two main quantitative measures in compression. We found that higher compression coincided with less error in all the studies (see Figs. 6–8 and 18) using a fixed quantizer. This may be because high compression is associated with low entropy, which means that the data contains more low values and less variation between the originally transformed and quantized coefficients. This phenomenon happens only when the quantization factor is fixed. We would like to call to the reader's attention regarding the link between this phenomenon and the designed error function that comprises mse and entropy reduction terms for searching an optional wavelet in the convolution neural network. With this concurrent trend (i.e., less error is associated with low entropy using a fixed quantizer), the neural network seems to be effectively operated in the defined searching task. Otherwise they would have functioned as competing factors and would have made the task difficult during the neural network search.

Although we have shown the general framework of a wavelet filter search using a neural network algorithm, only tap-4 wavelets were employed in our experiment. It seems that the above findings can be generalized for high order wavelets because the g filter is the key operator for the wavelet decomposition. The distribution of weights for high order wavelets should be maintained as discussed above in order to obtain a low entropy. We will continue to investigate the performance of biorthonormal wavelets where two sets coefficients are used in the convolution process. We predict that high performance wavelets in compression and data accuracy should possess a short kernel and a Daubechies-like weight distribution in the g filter of wavelets [10], [18].

In our previous papers [27], we indicated that wavelet decomposition might be appropriate for low-resolution small images such as the Lena image, CTs, and magnetic resonance (MR) images. For high-resolution large images such as digitized chest radiographs and mammograms, we found that the full-frame DCT produced the highest compression efficiency. This is because the DCT can pack highly correlated image information in a small frequency area. The DWT, however, requires many levels in decomposition to achieve a high compression ratio. The data inaccuracy would propagate from high-level wavelet domains to low level and to the reconstructed image.

VI. Conclusion

A convolution neural-network-based framework has been developed to search for optimal wavelet kernels which can produce the most favorable set of transform coefficients to preserve data accuracy and defined image features during the compression. In this paper, our technical achievements are: 1) development of a unified method to facilitate the multichannel decomposition in wavelet filtering; 2) designing a cost error function consisting of mse and imposed entropy reduction function to seek an optional wavelet kernel in the convolution neural network; and 3) converting a neural network suggested kernel into a filter constrained by the wavelet requirements.

In all medical image modalities, we have tested so far (including mammography, CT, and MR imaging), Daubechies wavelet or its nearby wavelets generally performs slightly greater compression results than those of other wavelets based on the measure of mse. With a large quantization factor, the Haar wavelet produces the lowest and highest mse's for the background and microcalcification profile areas, respectively. However, the Daubechies wavelet produces an opposite result. In addition, we found that the $\mu\mathrm{CaF}$ wavelet (i.e., 0.322 521 36, 0.852 589 27, 0.384 585 42, and $-0.145\,482\,69$) possesses the highest feature preservation capability in microcalcification peak, contrast and PSNR. Through this study, we also found that the Haar wavelet produces dramatically greater result in high contrast edge regions. In addition, optimization usually occurs on a band of wavelets not at a single wavelet.

We, therefore, conclude that the Daubechies wavelet (and its nearby wavelets) is generally applicable for image compression. However, the Haar wavelet is suitable for low-noise smooth areas and sharp edges. For a specific image pattern, such as microcalcifications on mammograms, one might find that one wavelet filter outperform others in the preservation of the features. Finally, we would to comment that optimal wavelets for general image texture have some things in common: they possess balanced negative terms at the two sides of the positive weight and the absolute value of g_1 or g_2 is much larger than that of the other weights in the wavelet coefficients.

References

[1] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
[2] A. K. Jain, "Image data compression: A review," *Proc. IEEE*, vol. 69, pp. 349–389, 1981.
[3] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*: Academic Press, 1982.
[4] "Initial Draft for Adaptive Discrete Cosine Transform Technique for Still Picture Data Compression Standard,", ISO/IEC JTC1/SC2/WG8 N800, JPEG Tech. Spec., 1990.
[5] S. C. Lo and H. K. Huang, "Compression of radiological images with matrix sizes 512, 1024, and 2048," *Radiology*, pp. 519–525, 1986.
[6] H. K. Huang, S. C. Lo, B. K. Ho, and S. L. Lou, "Radiological image compression using error-free an irreversible compression and reconstruction in 2-dimensional direct cosine transform coding techniques," *J. Opt. Soc. Amer.*, pp. 984–992, 1987.

[7] S. C. Lo, B. H. Krasner, S. K. Mun, and S. C. Horii, "The full-frame entropy encoding for radiological image compression," in *SPIE Proc. Medical Imaging V*, vol. 1444, 1991, pp. 265–277.
[8] I. Daubechies, "Orthonormal basis of compactly supported wavelets ," *Comm. Pure Appl. Math.*, vol. XLI, pp. 909–996, 1988.
[9] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Feb. 1992.
[11] "JPEG 2000 Part I: Final Draft International Standard (ISO/iec Fdis15444-1),", ISO/IEC JTC1/SC29/WG1 N1855, 2000.
[12] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Trans. Image Processing*, vol. 1, pp. 526–533, Oct. 1992.
[13] J. Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients," in *Proc. IEEE Data Compression Conf. '93*, Mar. 1993, pp. 214–223.
[14] Z. Xiong, K. Ramchandran, M. T. Orchard, and K. Asai, "Wavelet packets-based image coding using joint space-frequency quantization," in *Proc. IEEE Int. Conf. on Image Proc.*, Austin, TX, 1994, pp. 324–328.
[15] A. H. Tewfik, D. Sinha, and P. Jorgensen, "On the optimal choice of a wavelet for signal representation," *IEEE Trans. Inform, Theory*, vol. 38, pp. 747–765, Mar. 1992.
[16] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.
[17] M. V. Wickerhauser, *Adapted Wavelet Analysis From Theory to Software*. Wellesley, MA: A.K. Peter Ltd, 1994.
[18] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, pp. 1053–1060, Aug. 1995.
[19] B. J. Lucier, M. Kallergi, W. Qui, R. A. DeVore, R. A. Clark, E. B. Staff, and L. P. Clarke, "Wavelet compression and segmentation of digital mammograms," *J. Digital Imag.*, vol. 7, no. 1, pp. 27–38, 1944.
[20] S. C. Lo, E. Makariou, H. P. Chan, M. T. Freedman, D. D. Dorfman, and K. Berbaum, "Integer wavelet compression guided by a computer-aided detection system in mammography," *Med. Imag.*, vol. 4322, pp. 643–648, 2001.
[21] D. Shin, H.-H. Wu, and J.-C. Liu, "Region-of-interest-based wavelet compression scheme for medical images," in *SPIE Proc., Med. Imag.*, vol. 3031, 1997, pp. 790–798.
[22] N. S. Moon, J. K. Song, M. Kwon, J. H. Kim, and C. W. Lee, "Block-Based wavelet transform coding of mammograms with region adaptive quantization," in *SPIE Proc., Med. Imag.*, vol. 3335, 1998, pp. 556–562.
[23] S. C. Lo, H. Li, J. S. Lin, A. Hasegawa, Y. C. Wu, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network with wavelet kernel for disease pattern recognition," *Neural Networks*, vol. 8, no. 7/8, pp. 1201–1214, 1995.
[24] Y. LeCun, I. Guyon, L. D. Jackel, D. Henderson, B. Boser, R. E. Howard, J. S. Denker, W. Hubbad, and H. P. Graf, "Handwritten digital recognition: Applications of neural network chips and automatic learning," *IEEE Commun. Mag.*, pp. 41–46, Nov. 1989.
[25] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, vol. 1, pp. 318–362.
[26] S. C. Lo, B. H. Krasner, and S. K. Mun, "Noise impact on error-free image compression," *IEEE Trans. Med. Imag.*, vol. 9, pp. 202–206, Apr. 1990.
[27] S. C. Lo, H. Li, B. H. Krasner, M. T. Freedman, and S. K. Mun, "Large-frame compression using DCT and wavelet transform techniques," in *SPIE Proc., Med. Imag.*, vol. 2431, 1995, pp. 195–202.