International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:1, No:12, 2007

Intelligent System for Breast Cancer Prognosis using Multiwavelet Packets and Neural Network

Sepehr M.H.Jamarani, M.H.Moradi, H.Behnam, and G.A.Rezai Rad

***Abstract*—**This paper presents an approach for early breast cancer diagnostic by employing combination of artificial neural networks (ANN) and multiwaveletpacket based subband image decomposition. The microcalcifications correspond to high-frequency components of the image spectrum, detection of microcalcifications is achieved by decomposing the mammograms into different frequency subbands,, reconstructing the mammograms from the subbands containing only high frequencies. For this approach we employed different types of multiwaveletpacket. We used the result as an input of neural network for classification. The proposed methodology is tested using the Nijmegen and the Mammographic Image Analysis Society (MIAS) mammographic databases and images collected from local hospitals. Results are presented as the receiver operating characteristic (ROC) performance and are quantified by the area under the ROC curve.

***Keywords*—**Breast cancer, neural networks, diagnosis, multiwavelet packet, microcalcification.

I. Introduction

BREAST cancer is the type of cancer with highest incidence rates in women. It is also the most common cause of cancer death in women in many countries, only exceeded by lung cancer in Asian countries and recently in the United States [1]. The early detection of breast cancer is vital to improve its prognosis. Moreover, it is well known that screening mammography is the best tool available for detecting cancerous lesions before clinical symptoms appear. Since about half of cancers detected by mammography correspond to clustered microcalcifications, these lesions are one of the mammographic hallmarks of early breast cancer [1].Usually, the shape and arrangement of microcalci-fications help the radiologist to judge the likelihood of cancer being present. Malignant calcifications are typically very numerous, clustered, small, dot-like or elongated, variable in size, shape and density. Benign calcifications are generally larger, more rounded, smaller in number, more diffusely distributed, and more homogeneous in size and shape. However, because of the small size of microcalcifications, the characterization of benign and malignant lesions represents a very complex problem even for an experienced radiologist. Moreover, there are many cases in which the structure of malignant microcalcifications is not very different from that of benign ones. These perceptual problems result in screening errors that lead either to missed malignant cases or more often to unnecessary biopsies. It has been reported that only 17% of calcifications requiring biopsy are cancerous [1]. Computer-aided diagnosis can help to reduce the number of false positives and therefore the number of unnecessary biopsies.

In the literature, several techniques have been proposed to detect the presence of microcalcifications using various methodologies. Concerning image segmentation and specification of regions of interest (ROIs), several methods have been proposed such as classical image filtering and local thresholding [6]. Stochastic fractal models [9], wavelet analysis [10,11]. Furthermore, various classification methodologies have been reported for the characterization of ROI such as, fuzzy logic systems [4]. Nevertheless, the most work reported in the literature employs neural networks for cluster characterization [3,5,7].

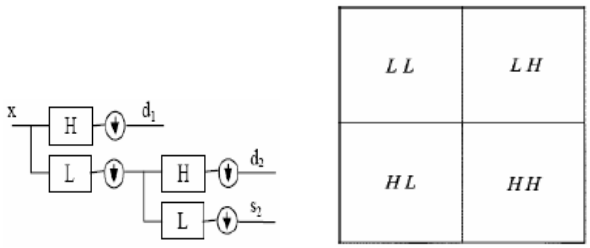
In this study, we present system, aiding radiologist for breast cancer diagnosis and identification of microcalci-fication clusters in digitized mammographic images. As the microcalcifications correspond to high-frequency components of the image spectrum, detection of microcalcifications is achieved by decomposing the mammograms into different frequency subbands, suppressing the low-frequency subband, and finally, reconstructing the mammogram from the subbands containing only high frequencies. The wavelet transform often fails to accurately capture high-frequency information, especially at low bit rates where such information is lost in quantization noise. Coifman *et al.* developed a technique called wavelet packets that is better able to represent high-frequency information [12].To achieve the best result we employed different types of multiwavelet packets such as GHM,CL. Orthogonal and biorthogonal, all are from the class of SA multifilters. The orthogonal SA multi filters used are“SA4” and “ORT4” [13], [14]; for biorthogonal SA multi filters we used “BSA7/5” and “BSA9/7” [15]. We use these results as inputs of neural network for classification. The neural network contains one input, two hidden and one output layers. A neural network is a set of connected input/output units where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class of the input samples. The back propagation algorithm performs learning on a multi layer feed-forward neural network.

Sepehr M.H. Jamarani is with Department of Biomedical Eng., Science and Research Branch, Islamic Azad University, Tehran, Iran. (e-mail: jamarani@shemroon.com).
M.H. Moradi is with the Department of Biomedical Eng., Amir Kabir University of Technology, Tehran, Iran.
H.Behnam and G.A.Rezai Rad is with the Department of Electrical Eng., Iran University of Science and Technology, Tehran, Iran.

II. Background

A. Wavelet

Multi-scale representation has proven to be useful in many Image processing applications. Wavelet analysis is one way to generate such representation. Mallat's pyramidal algorithm has been successfully used to decompose an input digital image into a hierarchy of independent detail subimages in subsequent (fine scales and an approximation sub-image in the coarsest scale [10, 11].Wavelet transform has been used for mammographic image compression, image enhancement, microcalcification detection, and feature extraction. Fig. 2 shows the structure of wavelet.



(a) (b)

Fig. 1 (a) Decimated analysis wavelet filter bank (b) wavelet output

The single level expansion results in 3 "details" images: d_{HH}, d_{HL}, and d_{LH}, (shorter: HH, HL, LH) covering Independent bands in the frequency domain. The "approximation" a_{LL} (or LL) is a low-pass component, which is passed to the next level of decomposition.(Figure1).

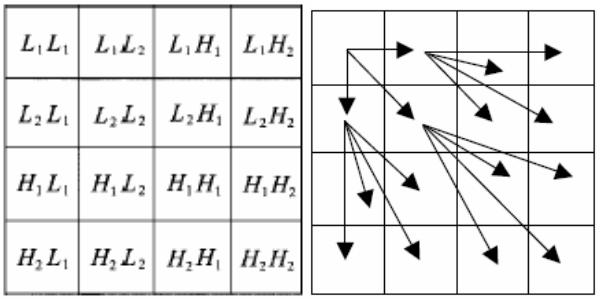
B. Multiwavelet

A newer alternative to the wavelet transform is the multiwavelet transform. Multiwavelets are very similar to wavelets but have some important differences. In particular, whereas wavelets have an associated scaling function $\phi(t)$ and wavelet function $\psi(t)$, multiwavelets have two or more scaling and wavelet functions. For notational convenience, the set of scaling functions can be written using the vector notation $\Phi(t) \equiv [\phi_1(t)\ \phi_2(t)\ \cdots\ \phi_r(t)]^T$, where is called the multiscaling function. Likewise, the multiwavelet function is defined from the set of wavelet functions as $\Psi(t) \equiv [\psi_1(t)\ \psi_2(t)\ \cdots\ \psi_r(t)]^T$. When $r = 1$, $\Psi(t)$, is called a *scalar* wavelet, or simply wavelet. While in principle can be arbitrarily large, the multiwavelets studied to date are primarily for $r = 2$.The multiwavelet two-scale equations resemble those for scalar wavelets.

$$\Phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} H_k \Phi(2t-k), \tag{1}$$

$$\Psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} G_k \Phi(2t-k). \tag{2}$$

Note, however, that $\{H_k\}$ and $\{G_k\}$ are *matrix* filters, i.e., H_k and G_k are $r \times r$ matrices for each integer k. The matrix elements in these filters provide more degrees of freedom than a traditional scalar wavelet. These extra degrees of freedom can be used to incorporate useful properties into the multiwavelet filters, such as orthogonality, symmetry, and high order of approximation. The key, then, is to figure out how to make the best use of these extra degrees of freedom. Multifilter construction methods are already being developed to exploit them [14], [15]. However, the multi-channel nature of multiwavelets also means that the subband structure resulting from passing a signal through a multi filter bank is different. Sufficiently different, in fact, so that established quantization methods do not perform as well with multiwavelets as they do with wavelets.



(a) (b)

Fig. 2 (a) Image subbands after single level decomposition (b)Parent-children relationship among subbnads

C. Wavelet Packet

Multiwavelets provide one alternative to the wavelet transform. Another alternative is the wavelet packet transform. Despite its general success, the wavelet transform often fails to accurately capture high-frequency information, especially at low bit rates where such information is lost in quantization noise. Coifman *et al.* developed a technique called wavelet packets that is better able to represent high-frequency information.

A multilevel wavelet filter bank involves iterating the lowpass–highpass filtering and downsampling procedure only on the output of the lowpass branch of the previous stage. Coifman *et al.* formulated an extension of the octave-band wavelet decomposition to full tree decomposition by allowing the lowpass–highpass filtering and downsampling procedure to be iterated also on highpass (bandpass) branches in the tree. They defined the new basis functions, called wavelet packets, as follows. Let $\phi(t)$ and $\psi(t)$ be the scaling and wavelet functions, respectively, which obey the two-scale equations

$$\phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} h_k \phi(2t-k), \tag{3}$$

$$\psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} g_k \phi(2t-k). \tag{4}$$

Note that the sequences $\{h_k\}$ and $\{g_k\}$ are the scaling and wavelet filter coefficients. Now let $u_0(t) \equiv \phi(t)$ and $u_1(t) \equiv \psi(t)$, and define

$$u_{2n}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} h_k u_n(2t-k), \quad (5)$$

$$u_{2n+1}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} g_k u_n(2t-k). \quad (6)$$

Taking dyadic rescaling and translations of these functions yields a library of functions $\{2^{-j/2}u_n(2^{-j}t-k)\}$. This library is over complete, but a proper complete basis can be found by selecting a subset of the library with the right set of parameters $\{n, j, k\}$.

This selection of a basis can be viewed in terms of a tree structure, in which the set of elements of each basis corresponds in a one-to-one fashion to a particular set of terminal nodes of a binary tree. Wavelet packets impose increased computational complexity due to the basis selection process. Selection of a "best" basis for any particular image may be performed in a number of ways. Coifman *et al.* suggested the use of an additive cost function that is applied to each set of parent and child nodes in the pruning process. If the sum of the costs of the children is greater than the parent's cost, the children are pruned; otherwise the children are kept. The performance of this method depends entirely on the choice of cost functions. Some cost functions that have been proposed include: Shannon entropy, the number of coefficients in the node that are significant compared to (i.e.,greater than) some threshold [16], and the number of bits required to represent all the coefficients in the node (introduced in this paper).Newer methods for selecting a basis approach the problem from a rate-distortion perspective. Ramchandran and Vetterli proposed a method that attempts to select the set of terminal nodes that are optimal in a rate-distortion sense. Their approach involves the minimization at each branch of a Lagrangian "cost function," $J(\lambda) = D + \lambda R$, where D is the average distortion and R is the target average bit rate. The value of that minimizes $J(\lambda)$ determines whether to prune and also gives the best quantizer for that node (which is then used for uniform quantization of the coefficients of that node). More recently, Xiong *et al.*have taken this idea and merged the basis optimization with their space-frequency quantization (SFQ) approach, yielding impressive results [17].

D. Multiwavelet Packet

Just as with scalar wavelets, the multiwavelet filter bank procedure involves iterating the filtering operation on the lowpass channel of the filter bank. And, just as with scalar wavelets, new basis functions can be produced by iterating on the highpass channels of multiwavelet filter banks too. This new approach combines wavelet packet decomposition with multiwavelet filters; hence, we call it multiwavelet packet decomposition. We define multiwavelet packets in a manner analogous to the definition of wavelet packets.

Let $U_0(t) \equiv \Phi(t)$ and $U_1(t) \equiv \Psi(t)$, and define

$$U_{2n}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} H_k U_n(2t-k), \quad (7)$$

$$U_{2n+1}(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} G_k U_n(2t-k). \quad (8)$$

Note the similarity between these mulitwavelet packet equations (7) and (8) and the corresponding wavelet packet equations, (5) and (6).

The basis selection algorithms and cost functions used to prune the resulting tree structure are identical to those of the scalar wavelet packet case with one exception: each branching in the multiwavelet packet tree structure creates four new channels (assuming $r = 2$) instead of just two. Since the multiwavelet packet tree then has four children for each parent, the computational complexity for multiwavelet packets may be higher than for wavelet packets. Cost function based methods will be essentially unaffected because they just operate on all the pixels corresponding to each node; with multiwavelet packets there are four nodes instead of two, but each node represents half as much data. However, methods that perform some form of rate-distortion optimization will require more computation due to the increased number of nodes. To achieve the best result we employed different types of multiwavelet packets such as GHM, CL. Both orthogonal and biorthogonal multiwavelets were tested, and all are from the class of SA multi filters. The orthogonal SA multi filters used are"SA4" and "ORT4" [13], [14]; for biorthogonal SA multifilters weused "BSA7/5" and "BSA9/7" [15]

III. ROC Curves

The ROC *receiver operating characteristic* curve is a plot of the sensitivity against one minus the specificity for different values of the threshold. Comparison is usually in terms of the area under the curve, which gives a summary of performance over the whole range of values and is independent of the prevalence of the condition unlike the accuracy, which weights the sensitivity and specificity in proportion to their prevalence. It measures the probability that for any pair of patients, one of whom with an event and one without, the patient for whom the event has occurred will have a higher predicted probability of the event than the other.

True Positive (TP): lesions called cancer and prove to be cancer. False Positive (FP): lesions called cancer that prove to be benign. False Negative (FN): lesions that are called negative or benign and prove to be cancer. True Negative (TN): lesions that are called negative and prove to be negative. False positive fraction is FPF (9) and true positive fraction is TPF(10)

$$TPF = \frac{TP}{TP+FN} \quad (9)$$

$$FPF = \frac{FP}{FP+TN} \quad (10)$$

Alternatively, the free receiver operating characteristic (FROC) curve may be used which considers the number of false positive clusters per image instead of the specificity value.

IV. Database

We used a mammogram database developed by N. Karssemeijer in the University Hospital Nijmegen, the Netherlands and Mammographic Image Analysis Society (MIAS)[2] and some other clinical images that collected form local hospitals mentioning that biopsy has done on the patients so we already know the results of benign or malignancy . The mammograms were recorded with a Kodak MIN-R/SO 177 screen/ (lm combination. An Eikonix 1412 CCD camera was used to digitize the images. The spatial resolution was set to be 2048 by 2048 pixels per image with a gray-level resolution of 12 bits per pixel. There are 66 mammograms in the database taken from 32 different patients. There are a total of 196 microcalcifications clusters, 63 benign and 133 malignant cases, in the database with varying sizes and visibility. Containing microcalcification clusters with different visibility, this database is a good representative of clinical cases.

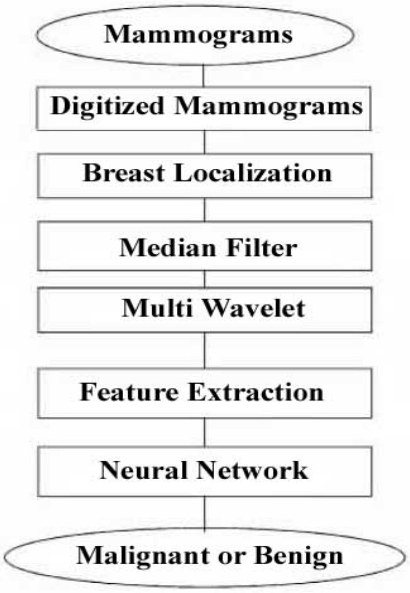


Fig. 3 Microcalcification detection system

V. Classification Module

In order to specify the features that will be used as inputs to the classification system, at first 32 features are identified and computed characterizing either an individual micro-calcification (object) or a group of them in a specific ROI. Those features fall into three categories related with the intensity, shape and texture properties of each object. The selection of the five largest microcalcifications is made since a very small microcalcification does not have enough pixels for reliable feature value computation. (Fig. 3) shows how the classification module works. Since the number of the computed features is quite large and their discriminative power varies, a feature validation together with feature selection procedure is applied. The receiver operating characteristic (ROC) curve is plotted for each feature and the area Az under the ROC curve is computed. Features with the highest Az are selected, resulting in a set of 18 features.

TABLE I
Features related to the Shape and Appearance of Microcalcifications

Number of microcalcifications in the cluster
Maximum size of calcifications in cluster
Standard deviation of the size of calcifications in cluster
Number of calcifications with size of 1 pixel
Sum of the area of the calcifications in each cluster
Maximum value of compactness in cluster
Average compactness in cluster
Radius of the circle that best fits the cluster
Scattering of the microcalcifications
Average gray level of the microcalcifications in cluster
Standard deviation of the mean of the microcalcification gray levels in the cluster
Maximum standard deviation of the gray levels in each calcification
Average standard deviation of the gray levels in each calcification in cluster
Area of the cluster convex hull
The length of the cluster convex hull
Neighbouring with a larger cluster
Average microcalcifications intensity
Average local microcalcification background

It must be noted that most of the selected features correspond to the mammographic characteristics that radiologists examine during a diagnostic procedure such as shape, density, size, distribution of the examined group or individual objects. In the next step of the classification module the selected features are fed into a neural network Classification system. The neural network that is used for characterization is a feed forward neural network with sigmoid hidden nodes.

In order to select appropriate architecture (number of hidden layers and hidden nodes per layer) several networks were tested with one or two hidden layers and different number of hidden nodes. In order to reduce the dimensionality of the input vector, a PCA (principal component analysis) was applied to eliminate the features that contribute less than 3% to the total variation of the data set. The PCA procedure transforms each 18-dimensional feature vector into a 7-dimensional feature vector that will constitute the input to the neural network (Fig. 4). The neural network contains one input, two hidden and one output layers. For learning phase we used back propagation algorithm which the input signals are computed and passed through the neural network layer by layer. Then the neurons in output layer products the output signals then error signals can be generated by comparing the output response with the desired response. This work can be used to minimize the distortion of the MLP. This learning algorithm is iteratively executed for the training set and then products the synaptic weight vectors. We used 60% of our images for training and the remaining was used for test process. Training is terminated either when the training error is less than a very small given value (10^{-5}) or when 2000 iterations have been performed.

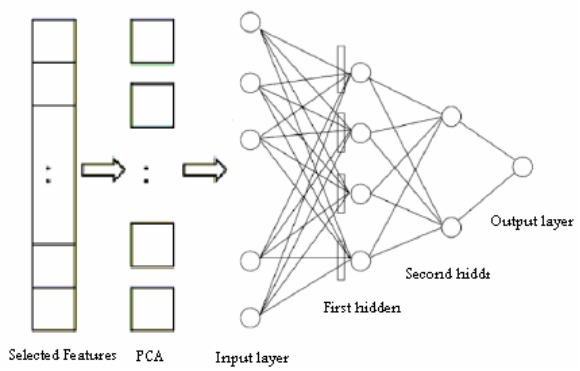


Fig. 4 PCA and Neural network architecture

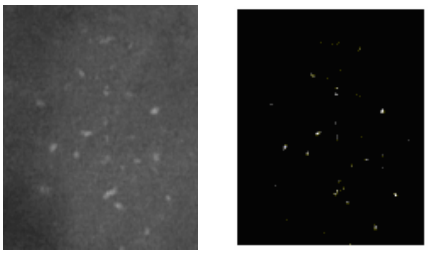
The network performance is measured using the area (Az) under a ROC curve generated by plotting the true positive fraction (sensitivity) against the false positive fraction (1-specificity) of the cases for various threshold values.

VI. Experimental Results

As mentioned before we applied different types of multiwavelet packets such as GHM, CL. Orthogonal and biorthogonal multiwavelets, all are from the class of SA multifilters. The orthogonal SA multi filters used are"SA4" and "ORT4" [13], [14]; for biorthogonal SA multi filters we used "BSA7/5" and "BSA9/7" [15]. The ROC curves, which presents the performance of the microcalcification system detection has shown in figure6.

VII. Conclusion

Intelligent system has been developed for the identification of microcalcification clusters in digitized mammograms, aiding the radiologists for breast cancer prognosis.



(a) (b)

Fig. 5 (a) Original mammogram (b) Result after applying wavelet

The method employs multiwavelet packets and neural network. We tested our system in the Nijmegen and the MIAS and images collected from local hospitals with satisfactory results. As it shown, comparing the multiwavelet packets, best performance was achieved by ORT4 multiwavelet packet with areas under ROC curve ranging around 0.98. The system successfully combines intelligent methods and image processing techniques which contribute to the enhancement of mammographic diagnosis sensitivity and reduction of negative biopsies. The proposed methodology could be an essential part of an integrated CAD (computer aided diagnosis) technique, which could assist radiologists in mammogram analysis and diagnostic decision making.

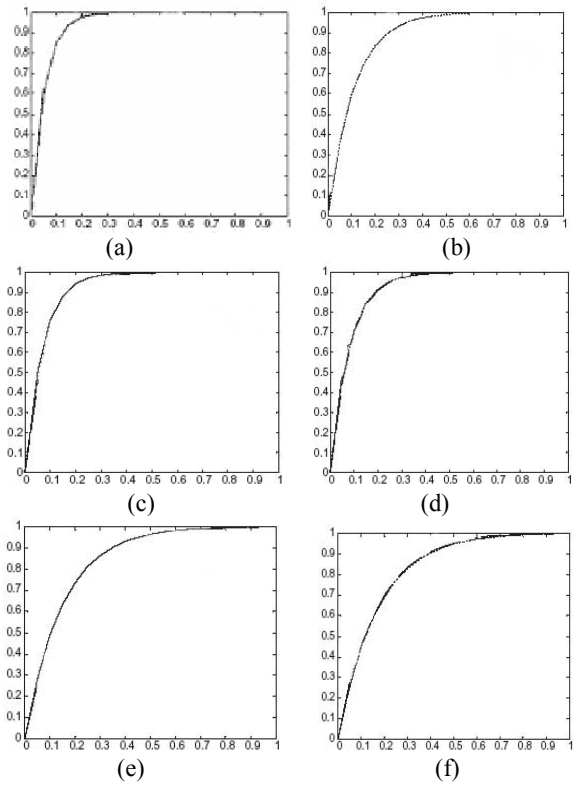


Fig. 6 ROC curves obtained from system corresponding to Mutiwavelet packets (a) ORT4 (b) SA4 (c) CL (d) GHM (e) BSA7/5 (f) BSA9/7

References

[1] CancerNet, A service of the National Cancer Institute http://cancernet.nci.nih.gov
[2] http://www.waiu.man.ac.uk/services/MIAS
[3] P. Sajda and C. Spence. Learning Contextual Relationships in Mammo-grams using a Hierarchical Pyramid Neural Network *IEEE Transactions on MedicalImaging* 21 (3) (2002)
[4] B. Verma and J. Zakos, \A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," Information Technology in biomedicine IEEE 5, pp. 46{54,March 2001
[5] Zheng L, Chan A. An artificial intelligent system for tumor detection in screening mammogram. IEEE Trans Med Im2001;20(7):559-67
[6] Moti Melloul, Leo Joskowicz, Segmentation of microcalci fication in X-ray mammograms using entropy thresholding technical Report, May 2002, Hebrew University, Leibniz Center.
[7] Z. R. Yang and R. G. Harrison, \Detecting false benign in Breast cancer diagnosis," Neural Networks, IEEE 3, pp. 655{658, July 2000.
[8] Dengler J, Behrens S, Desage JF. Segmentation of microcalcifications in mammograms. IEEE Trans Med Image 1993; 12:634-42.
[9] Li H, Liu KJR, Lo SCB. Fractal modelling and segmentation for the enhancement of microcalcifications in digital mammograms. IEEE Trans Med Imag 1997; 16(6):785-98.
[10] Yoshida H, Doi K, Nishikawa RM, Giger ML,Schmidt RA.An Improved CAD scheme using wavelet transform for detect- Ion of clustered microcalcifications in digital mammograms. Acad Radiol 1996;3:621-7.

[11] Lado MJ, Tahoces PG, Mendez AJ, Souto M, Vidal JJ. A wavelet-based algorithm for detecting clustered microcalcifications in digital mammograms. Med Phys 1999; 26(7):1294-305.
[12] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, wavelet Analysis and signal processing," in *Wavelets and Their A pplications*.Boston,MA: Jones and Bartlett, 1992, pp. 153–178.
[13] J. Y. Tham, L.-X. Shen, S. L. Lee, and H. H. Tan, "A general approach for analysis and application of discrete multiwavelet transforms," *IEEE Trans. Signal Processing*, vol. 48, pp. 457–464 Feb. 2000.
[14] T. Xia and Q. Jiang, "Optimal multifilter banks: Design, related symmetric extension transform and application to image compression," *IEEE Trans. Signal Processing*, vol. 47, pp. 1878– 1889, July 1999.
[15] S. S. Goh, Q. Jiang, and T. Xia, Construction of biorthogonal multiwaveletsusing the lifting scheme, preprint, 1998.
[16] G. Meyer,A. Z.Averbuch, and J. O. Strömberg, "Fast adaptivewavelet packet image compression," *IEEE Trans. Image Processing*, vol. 9, pp 792–800, May 2000.
[17] Xiong, K. Ramchandran, and M. T. Orchard, "Wavelet packet image coding using space-frequency quantization," *IEEE Trans. Image Processing*,vol. 7, pp. 892–898, June 1998.