[10] V.P. Kumar and T. Poggio, "Learning-Based Approach to Real Time Tracking and Analysis of Faces," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 96-101, 2000.
[11] G.R. Bradski, "Real Time Face and Object Tracking as a Component of a Perceptual User Interface," *Proc. IEEE Workshop Applications of Computer Vision*, pp. 214-219, 1998.
[12] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[13] B. Schölkopf, K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Pogio, and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," *IEEE Trans. Signal Processing*, vol. 45, no. 11, pp. 2758-2765, 1997.
[14] S. Haykin, *Neural Network—A Comprehensive Foundation*. second ed. Prentice Hall, 1999.
[15] T.M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Electronic Computers*, vol. 14, pp. 326-334, 1965.
[16] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 1-47, 1998.
[17] K.K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, Jan. 1998.
[18] T. Joachims, "Making Large-Scale SVM Learning Practical," *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, eds., pp. 169-184, Cambridge, Mass.: MIT Press, 1999.
[19] H. Li and D. Doermann, "A Video Text Detection System Based on Automated Training," *Proc. Int'l Conf. Pattern Recognition*, pp. 223-226, 2000.
[20] K.Y. Jeong, K. Jung, E.Y. Kim, and H.J. Kim, "Neural Network-Based Text Location for News Video Indexing," *Proc. Int'l Conf. Image Processing*, pp. 319-323, 1999.
[21] E.Y. Kim, K.I. Kim, K. Jung, and H.J. Kim, "A Video Indexing System Using Character Recognition," *Proc. IEEE Int'l Conf. Consumer Electronics*, pp. 358-359, 2000.
[22] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790-799, Aug. 1995.
[23] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 142-149, 2000.
[24] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1197-1203, 1999.
[25] Praktische Informatik IV, MOCA Project, http://www.informatik.uni-mannheim.de/informatik/pi4/projects/MoCA, 2003.
[26] Y. Zhong, H. Zhang, and A.K. Jain, "Automatic Caption Localization in Compressed Video," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 4, pp. 385-392, Apr. 2000.
[27] M.A. Smith and T. Kanade, "Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 775-781, 1997.
[28] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley-Interscience, 1973.
[29] L. Agnihotri and N. Dimitrova, "Text Detection for Video Analysis," *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries*, pp. 109-113, 1999.
[30] B. Schölkopf, "Support Vector Learning," PhD thesis, Munich: Oldenbourg Verlag, 1997.
[31] F. Idris and S. Panchanathan, "Review of Image and Video Indexing Techniques," *J. Visual Comm. and Image Representation*, vol. 8, no. 2, pp. 146-166, 1997.
[32] A.K. Jain, *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.
[33] E. Osuna, R. Freund, and F. Girosi, "An Improved Training Algorithm for Support Vector Machines," *Proc. IEEE Workshop Neural Network for Signal Processing*, pp. 276-285, 1997.
[34] Language and Media Processing (LAMP) Laboratory, media group, Univ. of Maryland, College Park, http://lamp.cfar.umd.edu, 2003.

Extraction of Features Using M-Band Wavelet Packet Frame and Their Neuro-Fuzzy Evaluation for Multitexture Segmentation

Mausumi Acharyya, Rajat K. De, *Member*, *IEEE*, and Malay K. Kundu, *Senior Member*, *IEEE*

Abstract—In this paper, we propose a scheme for segmentation of multitexture images. The methodology involves extraction of texture features using an overcomplete wavelet decomposition scheme called discrete M-band wavelet packet frame (DMbWPF). This is followed by the selection of important features using a neuro-fuzzy algorithm under unsupervised learning. A computationally efficient search procedure is developed for finding the optimal basis based on some maximum criterion of textural measures derived from the statistical parameters for each of the subbands. The superior discriminating capability of the extracted features for segmentation of various texture images over those obtained by several existing methods is established.

Index Terms—Texture segmentation, M-band wavelet packet frames, feature selection, fuzzy feature evaluation index, neural networks.

1 Introduction

Segmentation of multitexture image is an important problem in image analysis [1], [2]. In this regard, some applications of octave band wavelet decomposition scheme for texture analysis have been attempted [3], [4]. The main difficulty in octave band wavelet decomposition is that it can provide only a logarithmic frequency resolution which is not suitable for the analysis of high-frequency signals with relatively narrow bandwidth. The investigations of Chang and Kuo [5] and Laine and Fan [6] indicate that the texture features are more prevalent in the intermediate frequency band and showed promising results using wavelet packet frames [7], [8]. Therefore, the main motivation of this work is to utilize the decomposition scheme based on M-band ($M > 2$) wavelets, which, unlike the standard wavelet, provides a mixture of logarithmic and linear frequency resolution [9], [10]. The use of M-band wavelet decomposition gives rise to a large number of features, most of which are redundant for accurate and efficient recognition process. Therefore, proper selection of the appropriate features using some feature selection algorithms is necessary. Some recent attempts have been made for multiscale basis and feature selection in the framework of artificial neural networks (ANN) [11], [12]. In this paper, we propose a method of multitexture segmentation scheme based on discrete M-band wavelet packet frame analysis ($DMbWPF$) in order to extract the most significant information of a texture which often appears in the middle frequency channels. The other motivation of *frame* analysis is to achieve the translational invariance in texture analysis.

The proposed methodology has two parts: In the first part, we develop a computationally efficient and adaptive technique for finding out an appropriate tree-structured (less than the complete tree) M-band wavelet packet basis, to avoid a full decomposition. This selection of basis is based on some maximal criterion of textural measures to locate dominant information in each frequency channel (subband) and to decide whether information from a particular subband is needed or not. With this transform,

• *The authors are with the Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata - 700 108, India.*
E-mail: mau_ach@yahoo.co.in, {rajat, malay}@isical.ac.in.

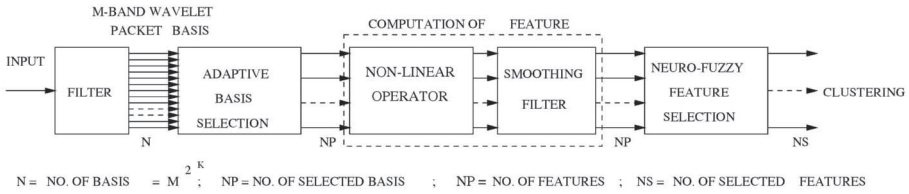

Fig. 1. Experimental setup.

we are able to zoom into any desired frequency channels for further decomposition. The textural measure derived from the statistical parameters, e.g., energy, is then extracted from each of the subbands. In order to reduce the number of bases (features) further, we have used in the second part, a neuro-fuzzy approach where a fuzzy feature evaluation index is defined and is used to minimize the total number of features to have same quality of output in a connectionist framework.

The article is organized as follows: In Section 2, we formulate the methodology for extraction of features. Section 3 provides the neuro-fuzzy feature selection algorithm. Section 4 analyzes experimental results and the conclusion is in Section 5.

2 Extraction of Multiscale Wavelet Features

In this section, we design a methodology of extracting multiscale wavelet features of a texture image. The entire methodology is depicted in Fig. 1.

2.1 M-band Wavelets

M-band ($M > 2$) wavelet decomposition is a direct generalization of the classical ($M = 2$-band) wavelet [10], [9]. An M-band wavelet system consists of a scaling function $\psi^1(x)$ which is given by, $\psi^1_{j,k}(x) = \sum_k M^{j/2}\psi^1(M^j x - k)$. Additionally, there are $M-1$ wavelet functions given by, $\psi^r_{j,k}(x) = \sum_k M^{j/2}\psi^r(M^j x - k), r = 2, \ldots, M$. Here, j and k are the scaling and translation parameters, respectively, and r gives the index of the wavelet functions in the wavelet system.

Although the M-band wavelet decomposition results in a combination of linear and logarithmic frequency (scale) resolution, we conjecture that a further recursive decomposition of the high-frequency regions would characterize textures better. This results in a tree structured multiband extension of the M-band wavelet transform which is the discrete M-band wavelet packet transform (DMbWPT). Thus, a finer and adjustable resolution at high frequencies is allowed as compared to the case of 2-band wavelet packet transform. In this work, we use a DMbWPF which is similar to DMbWPT, except that no downsampling occurs between scales (levels of decomposition) to achieve translational invariance.

In the filtering stage, we make use of $M(=4)$-band, orthogonal and linear phase wavelet filter bank following [10]. The one-dimensional (1D), $M(=4)$-band scaling (lowpass) and wavelet (bandpass) bases ψ^r ($r = 1, 2, 3, 4$) are given in [10]. Here, ψ^1 is the scaling function and ψ^r, with $r = 2, 3, 4$, are the wavelet functions.

The filter responses in the frequency domain $H_{j,r}(\omega)$ (for $r = 1, \ldots, 4$) at level j are generated as, $H_{j,r} = H_{0,r}(M^j\omega|_{r=1,\ldots,4})$. Scale $j = 0$ corresponds to the highest resolution of the signal, i.e., the original signal $I(x)$ before decomposition. Let, $\hat{I}^j_{b,r}(\omega)$ be the Fourier transform of the output of the wavelet packet b at decomposition level j obtained from the corresponding signal at the $j-1$th level with the rth band of the M-band filters. Then, for $0 \le b \le 4^j - 1$ and $r = 1, 2, 3, 4$, we have $\hat{I}^{j+1}_{4b+(r-1),q}(\omega) = H_{j,r}(\omega)\hat{I}^j_{b,q}(\omega)$.

From the filter bank theoretic point of view [13], this corresponds to a filter bank with channel filters $\{filt^j_{b,q}(\omega)|q = 1, \ldots, 4.\}$. $filt^j_{b,q}(\omega)$ are given by the recursive relation, $filt^0_{0,q}(\omega) = H_{0,r}(\omega)$ and $filt^{j+1}_{4b+(q-1),l}(\omega) = H_{j,r}(\omega)filt^j_{b,q}(\omega) = H_{0,r}(M^j\omega)filt^j_{b,q}(\omega)$. Fig. 2 shows a general tree structure of 1D discrete M-band wavelet packet frame decomposition. Module-A in Fig. 2 comprises of all the filters $H_{j,r}$ with $r = 1, 2, 3, 4$.

For images, we simply use tensor product extension for which the channel filters are written as $filt^j_{b,r[x\times y]}(\omega_x, \omega_y) = filt^j_{b,r[x]}(\omega_x)filt^j_{b,q[y]}(\omega_y)$. At scale $j = 0$, the image is first decomposed into $M \times M$ channels using all the filters $H_{0,r}$ and $H_{0,q}$ with $r, q = 1, 2, 3, 4$, and without downsampling. The process is repeated for each of the subbands in subsequent scales (j).

2.2 Adaptive Basis Selection

An M-band wavelet packet decomposition gives rise to M^{2^j} number of bases, for a decomposition depth J. It is quite evident that an exhaustive search technique to determine the optimal basis from this large set is computationally expensive. In order to find out a suitable basis without going for a full decomposition, we propose an adaptive decomposition algorithm using a maximal criterion of textural measures extracted from each of the subbands. Then, the significant subbands are identified and it is decided whether further decomposition of a particular channel would generate more information or not. This computationally efficient search enables one to zoom into any desired frequency channel for further decomposition [14].

For this purpose, the image is first decomposed into $M \times M$ channels using the 2D M-band wavelet transform without downsampling (oversampled). Energy for each subband is then computed. Among various subbands, those for which energy values exceed ϵ_1 percent of the energy of the parent band, are considered and decomposed further. We further decompose a subband if its energy value is more than some ϵ_2 percent of the total energy of all the subbands at the current scale. The analysis is performed upto the second level of decomposition and this results in a set of wavelet packet bases. These bases corresponding to different resolutions are assumed to capture and characterize effectively different scales of texture of the input image. Empirically, we have seen that a value of $\epsilon_1 = 2 - 5$ percent and $\epsilon_2 = 50$ percent are good choices for the images we have considered here. This simple top-down splitting technique performs well for most images.

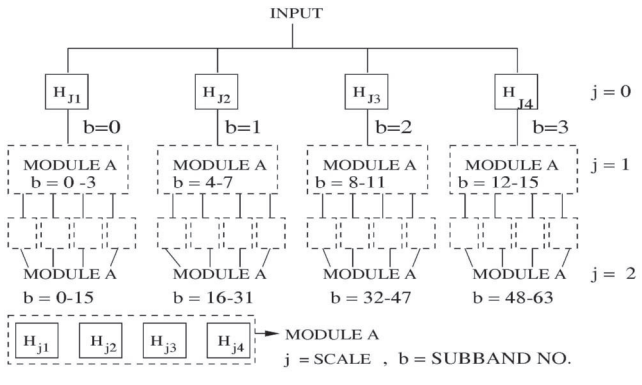

Fig. 2. Tree structure of 1D DMbWPF transform and related indices.

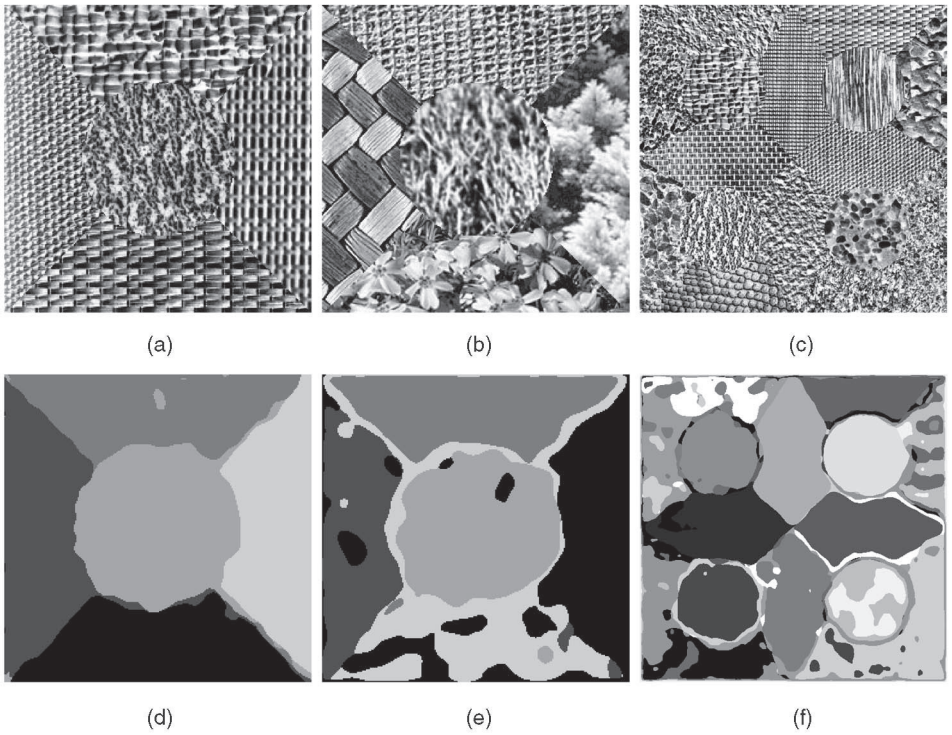

Fig. 3. Test images (a) $Nat5b$, (b) $Nat5v$, and (c) $Nat16c$. The corresponding segmented outputs (d), (e), and (f), respectively, after neuro-fuzzy feature evaluation.

2.3 Local Estimator of Textural Measure

Raw wavelet coefficients only are insufficient as complete texture cues. They are helpful in splitting the textured information into different frequency channels but without local information (statistics) around a pixel. A nonlinearity is needed in order to discriminate texture pairs with identical mean brightness and second-order statistics. To calculate local features of an image, we slide a fixed size window on the wavelet coefficients of an image and compute the local statistics of window in each individual position, associate these values as feature values of the central pixel in each of these windows. There exists a wide variety of textural measures but, in the current study, energy measure that represents textural uniformity, i.e., pixel pairs repetitions is considered as feature. We have used modulus operator as the nonlinearity and *Average Absolute Deviation (AAD)* from the mean is used as a generalized definition of energy to get separation of features for different patterns.

For a subband image $F_b(x,y)$ where $0 \leq x \leq M-1, 0 \leq y \leq N-1$ and subband number b, the local energy $Eng_b(x,y)$ around the (x,y)th pixel is expressed as

$$Eng_b(x,y) = \frac{1}{R}\sum_{m=1}^{w}\sum_{n=1}^{w} \mid (F_b(m,n) - \overline{F}_b(x,y)) \mid, \qquad (1)$$

where w is the window size and area $R = w \times w$. The term $\overline{F}_b(x,y)$ is the mean around the (x,y)th pixel and $F_0(x,y) = I(x,y)$, the original image.

This step is followed by a smoothing stage using Gaussian low pass filter $h_G(x,y)$ to get a feature image $Feat_b(x,y)$ a function of subband image $F_b(x,y)$ and is given by

$$Feat_b(x,y) = \sum_{(a,b)\in G_{xy}} \Gamma(F_b(a,b)h_G(x-a,y-b)),$$

where $\Gamma(\cdot)$ gives the energy measure and G_{xy} is a $G \times G$ window centered around a pixel with coordinates (x,y). Use of a Gaussian (weighting) window results in less sparse points (i.e., denser feature distributions) as compared to when uniform weighting window is used. The local AAD values from the mean (as shown in (1)) of a Gaussian window is found to provide robust quality of features in the feature space for all of test images used. Another issue in this regard is the size of local window. From a number of experiments, it is found that the choice of size of the local window is very crucial for extracting proper features. In an image with patterns of different texel sizes, the selection of window size suitable for various texel is a difficult task. For larger texels, it is better to choose larger local window size but this may introduce more uncertainty in detection of the boundary regions. This problem can be solved if the effective window size changes with the level of resolution. The window of varying sizes will be able to capture textures with different texel sizes. Likewise, the size G of the Gaussian averaging window is also an important parameter. Reliable measurement of texture feature demands larger window size but, on the other hand, more accurate localization of region boundaries requires smaller window. After extracting a set of feature images $Feat_b(x,y)$, a set of feature vectors are derived from them. The oversampled wavelet transforms introduce redundancy in filtered images that may be useful getting for a reliable result in a recognition problem.

3 Selection of Wavelet Features Using a Neuro-Fuzzy Method and Segmentation

3.1 Fuzzy Feature Evaluation Index and Membership Function

The wavelet features extracted as mentioned in the previous section are evaluated and few of them are selected using a neuro-fuzzy feature selection criterion under unsupervised learning. The process begins with the clustering of the entire feature space using $k-means$ clustering algorithm for grouping the data points into different clusters with their centers cen_q's, i.e., two sets of samples, namely, $S = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p, \cdots, \boldsymbol{x}_{N^2}\}$ and $S_c = \{\boldsymbol{cen}_1, \boldsymbol{cen}_2, \cdots, \boldsymbol{cen}_c\}$ are formed. Based on this first hand knowledge about the cluster centers, the neuro-fuzzy feature selection algorithm is developed.

TABLE 1
Performance with Different Test Images

Test figure	*Steps*				
	With feature evaluation		*Without feature evaluation*		*With Post-processing*
	Classification	Number of features	Classification	Number of features	Classification
Nat5b	97.9%	5	94.6%	21	98.7%
Nat5v	84.9%	5	84.4%	32	86.2%
Nat10a	79.5%	7	71.7%	19	84.0%
Nat16c	79.0%	11	67.3%	29	80.4%
Patch5	89.9%	5	88.1%	12	92.2%

This method involves the formulation of a fuzzy feature evaluation index followed by its minimization in connectionist framework. The feature evaluation index for a set of transformed features is defined as

$$E = \frac{2}{s(s-1)} \sum_p \sum_q \frac{1}{2} \left[\mu_{pq}^T \left(1 - \mu_{pq}^O\right) + \mu_{pq}^O \left(1 - \mu_{pq}^T\right) \right], \qquad (2)$$

where s is the number of samples in which the fuzzy feature evaluation index is computed. $\mu_{pq}^O \in [0,1]$ and $\mu_{pq}^T \in [0,1]$ are the degree of similarity between the pth pattern and qth center in the n-dimensional original feature space, and in the n'-dimensional ($n' \leq n$) transformed feature space, respectively (μ_{pq} is the membership value of a pair of patterns belonging to a fuzzy set "Similar"). E decreases as the degree of similarity between x_p and cen_q in the transformed feature space tends to either 0 (when $\mu^O < 0.5$) or 1 (when $\mu^O > 0.5$). Therefore, our objective is to select those features for which the evaluation index becomes minimum; thereby optimizing the degree of the similarity of a pair of patterns μ_{pq} in a feature space, satisfying the characteristics of E in (2). This may be defined as [15]

$$\begin{aligned} \mu_{pq} &= 1 - \tfrac{d_{pq}}{D} \quad if\ d_{pq} \leq D, \\ &= 0, \qquad otherwise. \end{aligned} \qquad (3)$$

d_{pq} is the distance between the pth pattern and qth cluster center in the feature space and is defined as

$$\begin{aligned} d_{pq} &= \left[\textstyle\sum_i w_i^2 (x_{pi} - cen_{qi})^2 \right]^{\frac{1}{2}}, \\ &= \left[\textstyle\sum_i w_i^2 \chi_i^2 \right]^{\frac{1}{2}}, \chi_i = (x_{pi} - cen_{qi}), \end{aligned} \qquad (4)$$

where $w_i \in [0,1]$ represents weighting coefficient corresponding to ith feature. The terms x_{pi} and cen_{qi} are values of ith feature of pth pattern and qth cluster center, respectively.

The term D in (3) is a parameter which indicates the minimum separation between a pair of dissimilar patterns. When $d_{pq} = 0$ and $d_{pq} = D$, we have $\mu_{pq} = 1$ and 0, respectively. In our investigation, we have chosen $D = \alpha d_{max}$, where d_{max} is the maximum separation between a pair of patterns in the entire feature space, and $0 < \alpha \leq 1$ is a user defined constant. α determines the degree of flattening of the similarity function in (3). The higher the value of α, more will be the degree, and *vice versa*. d_{max} is defined as

$$d_{max} = \left[\sum_i (x_{maxi} - x_{mini})^2 \right]^{\frac{1}{2}}, \qquad (5)$$

where x_{maxi} and x_{mini} are the maximum and minimum values of the ith feature in the corresponding feature space.

The weight w_i in (4) indicates the relative importance of the feature x_i in measuring the similarity of a pair of patterns. The higher the value of w_i, the more is the importance of x_i in characterizing a cluster.

The computation of μ^T requires (3), (4), and (5), while μ^O needs these equations with $w_i = 1, \forall i$. Therefore, the evaluation index E in (2) is a function of w, i.e., $E(w)$, if we consider ranking of n features in a set. The problem of feature selection/ranking thus reduces to finding a set of w_is for which E becomes minimum; w_is indicating the relative importance of x_is. For the details concerning the operation of the network, one may refer to [15].

3.2 Connectionist Model

The network consists of an input, a hidden, and an output layer [15]. The input layer consists of a pair of nodes corresponding to each feature, i.e., $2n$ nodes, for n-dimensional (original) feature space. The hidden layer consists of n number of nodes which compute the part χ_i^2 in (4) for each pair of patterns. The output layer consists of two nodes. One of them computes μ^O and the other μ^T. The feature evaluation index $E(w)$ in (2) is computed from these μ-values of the network. During learning, each pair of patterns are presented at the input layer and the evaluation index is computed. The connection weights $\mathbf{W}(= w_j)$s are updated in order to minimize the index $E(\mathbf{W})$. The task of minimization of $E(\mathbf{W})$ in (2) with respect to $\mathbf{W}$ is performed using gradient-descent technique. $E(\mathbf{W})$, after convergence, attains a local minimum and then the weights ($W_j = w_j^2$) of

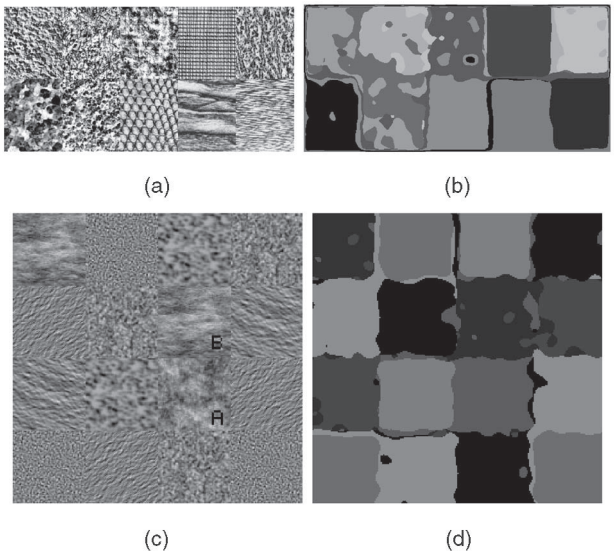

Fig. 4. Test images (a) $Nat10a$ and (c) $patch5$. The corresponding segmented outputs (b) and (d), respectively, after neuro-fuzzy feature evaluation.

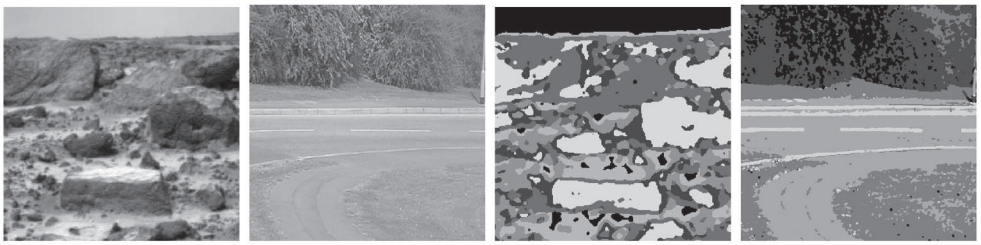

Fig. 5. Test images (a) $ROCK$, (b) $MB433$, and their segmented outputs (c) and (d), respectively, after neuro-fuzzy feature evaluation.

the links connecting hidden nodes and the output node computing μ^T-values, indicate the order of importance of the features.

As indicated in [15], for selecting an optimal set of features out of the total $s = N^2$ (image of size $N \times N$), the number of patterns to be presented to the connectionist system in one epoch, during its training, is $\frac{s(s-1)}{2} = \frac{N^2(N^2-1)}{2}$, which is a fairly large quantity. This requires a very high computational cost. On the other hand, in the proposed modified technique, the similarity between the patterns and cluster centers are computed, instead of computing it for every pair of patterns. These cluster centers may be considered as prototypes for all points belonging to the respective clusters. Thus, the number of patterns to be presented to the network in one epoch becomes $\frac{s(s_c-1)}{2} = \frac{N^2(c-1)}{2}$, where $s = |S|$ and $s_c = |S_c| << s$.

The features thus selected are used for segmenting a texture image. For this purpose, we have used a *k-means* clustering algorithm.

4 Experimental Results

In this section, we demonstrate the effectiveness of the proposed methodology on various multitexture and real life images for segmentation. The superiority of this methodology over other existing methods studied in [16], is also established.

Here, we have considered two 5-texture images $Nat5b$ and $Nat5v$ (each of 256×256), a 10-texture image $Nat10a$ (256×512), and a 16-texture image $Nat16c$ (512×512) [16]. We have also worked on another test image $Patch5$ (256×256) comprised of seven texture classes. Moreover, we have considered two natural scenes $ROCK$ and $MB433$ (each of 512×512). $Rock$, a complicated image consists of six classes, viz., dark rocks, light rocks, pebbles, dust, sky, and horizon. This is an image taken from Sol-3 of the Mars Pathfinder Mission [17]. $MB433$, consists of five classes, e.g., leaves1, leaves2, grass, road, and miscellaneous [18]. Throughout the study, the parameters are chosen to be, $\epsilon_1 = 2 - 5$ percent, $\epsilon_2 = 50$ percent, local window sizes ($w \times w$) are 9×9, and 17×17 for the first and second level of resolutions. The averaging window sizes are also chosen in commensuration with the local windows.

4.1 Performance Evaluation of the Proposed Methodology

To evaluate the performance of the proposed methodology, we have experimented with energy as textural measure. The percentage of correctly classified pixels is used as the segmentation quality measure. In order to demonstrate the importance of neuro-fuzzy feature evaluation, we present here the segmented outputs for all the test images with and without feature evaluation. The results show that the feature dimensionality is greatly reduced after feature evaluation. Simple median filtering is applied to the class maps as a postprocessing step to improve the segmentation results.

Fig. 3d shows the segmented image of $Nat5b$ (Fig. 3a). The five texture classes can easily be identified here. But, for $Nat5v$ (Fig. 3b), the classification error is more in the segmented output in Fig. 3e as compared to Fig. 3d, due to wider within class variation. The percentages of correctly classified pixels are found to be 97.9 percent and 84.9 percent for $Nat5b$ and $Nat5v$, respectively. Moreover, the number of features have been greatly reduced from 21 to 5 for $Nat5b$ and 32 to 5 for $Nat5v$, by using neuro-fuzzy feature evaluation (Table 1).

The texture mosaic $Nat16c$ (Fig. 3c) which is comprised of as many as 16 Brodazt textures, also shows very complex boundaries between various constituent textures. The segmented image (Fig. 3f) shows 15 different classes, where two of the classes have been merged which are also visually unidentifiable from the original image. For texture mosaic $Nat10a$ (Fig. 4a), the segmented output is presented in Fig. 4b. Note that, although the image contains 10 different Brodazt textures, some of them are not distinctly identifiable visually. Interestingly, the proposed methodology is able to identify more or less all the classes. In the case of $Patch5$ (Fig. 4c), there are seven different texture classes. Here, a particular class is not only confined to a specific region, but also mixed up with other classes in various regions of the image. Moreover, the boundaries between various classes, like other test images, are not easily discriminable. Even then, the different classes in the segmented image (Fig. 4d) are identified satisfactorily. Table 1 summarizes the performance of segmentation for all test images of Figs. 3 and 4.

The real-life natural scene images contain natural objects like roads, brick walls, etc., having uniform texture, and snow, leaves, trees, etc., more of a fractal nature. Since ground truths about these images are not always available, we have used a quantitative performance measure β [19]. It is defined as the ratio of the total variation and within class variation. The higher the value of β, the better the segmentation is. In the natural scene, $Rock$ (Fig. 5a), there are six different texture classes. Here, the boundaries between various classes, like other images, are not easily discernable. Even then the classes in the segmented image has been identified satisfactorily. Similar observations are noted for the image $MB433$ (Fig. 5b). The segmented images with the neuro-fuzzy feature evaluation are illustrated in Figs. 5c and 5d. Table 2 summarizes the β values with and without the feature evaluation step and also the number of features required for segmentation.

As a part of the investigation, an extensive comparison has been made to show the superiority of the proposed methodology over a number of existing related algorithms studied by Randen and Husøy in [16]. The results have been presented here for $Nat5b$,

TABLE 2
Performance with Different Scenes

Test figure	*Steps*			
	With feature evaluation		*Without feature evaluation*	
	β	Number of features	β	Number of features
Rock	2.931870	3	3.559690	5
$MB433$	4.272661	1	6.2240	5

TABLE 3
Comparative Study of the Performance on $Nat5b$, $Nat5v$, $Nat16c$, and $Nat10a$

Methods/filters		*Test image*							
		$Nat5b$		$Nat5v$		$Nat16c$		$Nat10a$	
		% classfn.	No. of features	% classfn.	No. of features	% classfn.	No. of features	% classfn.	No. of features
Proposed method M-band wavelet	Without feature evaluation	94.6	21	84.4	32	67.3	29	71.7	19
	With feature evaluation	97.9	5	84.9	5	79.0	11	79.5	7
Gabor filter bank (d) [20]		91.8	40	81.1	40	63.6	40	67.7	20
Gabor filter [21]		92.8	10	70.6	10	65.3	16	64.1	15
$f8a$ (d) [20]		92.8	40	81.1	40	63.6	40	60.3	40
$f8a$ (d) [20]		90.6	40	79.5	40	60.9	40	57.7	13

$Nat5v$, $Nat10a$, and $Nat16c$ only. The comparison is made with respect to the number of features used for texture segmentation and the percentage of correctly classified pixels. Table 3 summarizes the comparative results and establishes the superiority of the proposed method compared to the methods tabulated.

5 Conclusion

In this paper, we have described a feature extraction method based on M-band wavelet packet frames followed by neuro-fuzzy evaluation of the extracted features for texture segmentation. The use of M-band wavelet decomposition of the texture image provides an efficient representation of the image in terms of frequencies in different directions and orientations at different resolutions. This representation thus facilitates an improved segmentation of the different texture regions. The neuro-fuzzy feature evaluation method helps in finding out important features efficiently from a texture image where the various texture classes are overlapping in nature. The methodology does not require a priori knowledge about the spatial relationship of different classes in the test images.

The features obtained by the feature extraction method have been able to segment various synthetic texture images and as well as real life natural images satisfactorily with a greatly reduced number of extracted features and with improved quality of segmentation results.

References

[1] *Soft Computing for Image Processing*, S.K. Pal, A. Ghosh, and M.K. Kundu, eds., Heidelberg: Physica Verlag, 2000.

[2] M. Acharyya and M.K. Kundu, "An Adaptive Approach to Unsupervised Texture Segmentation Using M-Band Wavelet," *Signal Processing*, vol. 81, no. 7, pp. 1337-1356, 2001.

[3] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, July 1989.

[4] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames," *IEEE Trans. Image Processing*, vol. 4, no. 11, pp. 1549-1560, 1995.

[5] T. Chang and C.C.J. Kuo, "Texture Analysis and Classification with Tree Structured Wavelet Transform," *IEEE Trans. Image Processing*, vol. 2, no. 4, pp. 42-44, 1993.

[6] A. Laine and J. Fan, "Frame Representation for Texture Segmentation," *IEEE Trans. Image Processing*, vol. 5, no. 5, pp. 771-779, 1996.

[7] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186-1190, Nov. 1993.

[8] K. Etemad and R. Chellappa, "Separability Based Tree Structured Basis Selection for Texture Classification," *Proc. First Int'l Conf. Image Processing*, pp. 441-445, Nov. 1994.

[9] P. Steffen, P.N. Heller, R.A. Gopinath, and C.S. Burrus, "Theory of Regular M-Band Wavelet Bases," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3497-3510, 1993.

[10] O. Alkin and H. Caglar, "Design of Efficient M-Band Coders with Linear Phase and Perfect Reconstruction Properties," *IEEE Trans. Signal Processing*, vol. 43, no. 7, pp. 1579-1590, 1995.

[11] N. Saito and R.R. Coifman, "Local Discriminant Basis," *Proc. SPIE 2303, Math. Imaging: Wavelet Applications in Signal and Image Processing*, A.F. Laine and M.A. Unser, eds., pp. 2-14, July 1994.

[12] K. Etemad and R. Chellapa, "Separability-Based Multiscale Basis Selection and Feature Extraction for Signal and Image Classification," *IEEE Trans. Image Processing*, vol. 7, no. 10, pp. 1453-1465, Oct. 1998.

[13] M. Vetterelli and C. Herley, "Wavelets and Filter Banks: Theory and Design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207-2232, 1992.

[14] M. Acharyya and M.K. Kundu, "Adaptive Basis Selection for Multitexture Segmentation by M-Band Wavelet Packet Frame," *Proc. 2001 Int'l Conf. Image Processing*, pp. 622-625, Oct. 2001.

[15] J. Basak, R.K. De, and S.K. Pal, "Unsupervised Feature Selection Using Neuro-Fuzzy Approach," *Pattern Recognition Letters*, vol. 19, pp. 997-1006, 1998.

[16] T. Randen and J.H. Husøy, "Filtering for Texture Classification: A Comparative Study," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291-310, Apr. 1999.

[17] M.A. Ruzon, V.C. Gulick, R.L. Morris, and T.L. Roush, "Autonomous Scene Analysis of Digital Images for Mars Sample Return and Beyond," *Proc. 30th Lunar and Planetary Science Conf.*, 1999.

[18] MINERVA, http://www.project-minerva.ex.ac.uk. 2003.

[19] S.K. Pal, A. Ghosh, and B. Uma Shankar, "Segmentation with Remotely Sensed Images with Fuzzy Thresholding, and Quantitative Evaluation," *Int'l J. Remote Sensing*, vol. 21, no. 11, pp. 2269-2300, 2000.

[20] T. Randen and J.H. Husøy, "Multichannel Filtering for Image Texture Segmentation," *Optical Eng.*, vol. 33, pp. 2617-2625, 1994.

[21] A.C. Bovik, "Analysis of Multichannel Narrow-Band Filters for Image Texture Segmentation," *IEEE Trans. Signal Processing*, vol. 39, pp. 2025-2043, Sept. 1991.