International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 1, No. 1 (2003) 115–149

World Scientific
www.worldscientific.com

M-BAND WAVELETS: APPLICATION TO TEXTURE SEGMENTATION FOR REAL LIFE IMAGE ANALYSIS

MALAY K. KUNDU* and MAUSUMI ACHARYYA†

Machine Intelligence Unit, Indian Statistial Institute,
203 B. T. Road, Kolkata 700108, India
**malay@isical.ac.in*
†*res9522@isical.ac.in*

Received 1 November 2002

This paper describes two examples of real-life applications of texture segmentation using M-band wavelets. In the first part of the paper, an efficient and computationally fast method for segmenting text and graphics part of a document image based on textural cues is presented. It is logical to assume that the graphics part has different textural properties than the non-graphics (text) part. So, this is basically a two-class texture segmentation problem. The second part of the paper describes a segmentation scheme for another real-life data such as remotely sensed image. Different quasi-homogeneous regions in the image can be treated to have different texture properties. Based on this assumption the multi-class texture segmentation scheme is applied for this purpose.

Keywords: Texture segmentation; M-band wavelets; document image; remotely sensed image.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

In this paper we discuss the problems of segmentation of two-class and multi-class texture images by using a generalization of the wavelet transforms to the M-band case.[1,2] M-band wavelet is a tool for viewing signals at different scales and directions and decomposes a signal by projecting it onto a family of functions generated from a single wavelet basis via its dilations and translations.[3,4]

The standard 2-band wavelets are not suitable for the analysis of high frequency signals with relatively narrow bandwidth. So the main motivation of the present paper is to use the decomposition scheme based on M-band ($M > 2$) wavelets, which may improve the segmentation accuracy to a good degree. The motivation for a larger M ($M > 2$) comes from the fact that, unlike the standard wavelet

*Corresponding author

decomposition which results in a logarithmic frequency resolution, the M-band decomposition generates a mixture of logarithmic and linear frequency resolution. That is, generates a more flexible tiling of the scale-space plane than that resulting from 2-band wavelet. Furthermore, M-band wavelet decompositions yield a large number of subbands which are required for good quality segmentation.

Texture analysis using M-band wavelet have been investigated by Greiner *et al.*[5] They have worked with a 3-band extension of 2-band bi-orthogonal wavelets. Relevant texture features are extracted by a 3-band bi-orthogonal wavelet, and are subsequently used in classification of textures. Recently Chitre and Dhawan[6] have used M-band wavelets for texture classification.

In the first part of the paper, we describe a scheme for document image segmentation. Here we present an application of the two-class texture segmentation scheme.[1,7] The image is first decomposed into $M \times M$ subbands by applying the M-band wavelet in a separable manner without down sampling, which gives an oversampled (redundant) representation of the image. Then various combinations of these subbands are taken that gives multidirectional and multi-stage representation of the image. Texture features are extracted from these subbands in the next step. This is achieved by estimating in each subband, the local energy around each pixel over a small neighborhood. This requires a nonlinear operation followed by smoothing (averaging) with carefully chosen window size. In our work this is determined adaptively based on the spectral frequency content of the images. Superior discriminating capability of the extracted features over those obtained by several existing methods is shown.

In the second part of the paper, we discuss about a segmentation scheme for remotely sensed images. Here we present a multi-class texture segmentation technique using M-band wavelet packet frames.[2,9] 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 the textures better. M-band wavelet packet transform recursively decomposes both the high frequency and low frequency bands. Also most of the significant information of a texture often appears in the middle frequency channels. Considering the above facts, the present study investigates the potentiality of an *M-band wavelet packet* transform for this purpose. Besides this, *frame* analysis is employed to accomplish translation invariance. This transform corresponds to a general tree-structured filter bank. Features are derived from these filter outputs. The decomposition scheme employing M-band wavelet packet leads to a large number of independent bases. It is usually unnecessary to decompose all subbands in each scale to achieve the full decomposition tree. Hence, an appropriate means of selecting the significant bases is required. The selection of basis is based on some criterion of a maximum measure of textural cue, in order to locate only those frequency channels (subbands) which convey dominant information, and to decide whether further sub-division of a particular subband is needed or not. With this transform, we are able to zoom into any desired frequency channels for

further decomposition. The textural measure based on the statistical parameters, viz., energy, is then extracted from each of the subbands. The energy measures textural uniformity, i.e. extent of repetitions of pixel pairs.

Even if decomposition of a full tree is avoided by selection of appropriate bases, the number of features generated are still large, which incurs redundancy, so it is necessary to reduce further the number of redundant features. Several algorithms for multi-scale basis and feature selection are reported in the literature.[10−13] The idea of designing local basis using class separability criteria has been studied by Saito and Coifman.[14] Other techniques for best basis selection are also reported.[15,16] Some recent attempts have been reported for feature selection in the framework of artificial neural networks (ANN).[17−20] Recently, Etemad and Chelappa[21] have employed a class separability measure for selection of a suitable set of basis functions, in connectionist framework, from a tree structured wavelet packet basis set. We focus here on a neuro-fuzzy approach under unsupervised learning in our work, for selecting an optimal set of features, from those obtained after the selection of the basis. For this purpose, we have developed a neuro-fuzzy feature selection framework, which involves formulation of a layered network for minimization of a fuzzy feature evaluation index. This is a modification of the methodology of Pal *et al.*[22] It is to be noted that neuro-fuzzy computing[23,24] which integrates the merits of fuzzy set theory and ANN, enables the feature selection process artificially more intelligent.

The paper is organized as follows. In the following section we briefly discuss about M-band wavelets. In Sec. 3 a survey of some existing document and remotely sensed image segmentation techniques is presented. Section 4 presents the M-band wavelet filtering technique and extraction of features, and further discusses about the proposed feature selection algorithms for segmenting document and remotely sensed images. Finally we conclude our study in Sec. 5.

2. M-Band Wavelet Transform

The *wavelet transform* maps a function $f(x) \in L^2(R)$ onto a scale-space plane. The wavelets are obtained from a single prototype function $\psi(x)$ by scaling parameters a and shift parameters b.[25−27] The continuous wavelet transform of a function $f(x)$ is given as,

$$Wf_a(b) = \int f(x)\psi^*_{a,b}(x)dx \tag{2.1}$$

M-band wavelet decomposition is a direct generalization of the above two-band case.[3,4] Let $\phi(x)$ be the scaling function satisfying,

$$\phi(x) = \sum_k h(k)\sqrt{M}\phi(Mx - k)\,. \tag{2.2}$$

In addition there are $M - 1$ wavelets which also satisfy,

$$\psi^{(j)}(x) = \sum_k \sqrt{M}h^{(j)}(k)\psi(Mx - k)\,. \tag{2.3}$$

In discrete form these equations can be written as,

$$\phi_{ik}(x) = \sum_k M^{-i/2}\phi(M^{-i}x - k) \tag{2.4}$$

and

$$\psi_{ik}^{(j)}(x) = \sum_k M^{-i/2}\psi^{(j)}(M^{-i}x - k)\,, \quad j = 1, \ldots, M-1\,. \tag{2.5}$$

Let the subspaces spanned by the functions $\phi_{ik}(x)$ and $\psi_{ik}^{(j)}(x)$ be respectively defined as,

$$V_i = \overline{\text{span}\phi_{ik};\ \forall\, k \in Z}\,, \tag{2.6}$$

$$W_i^{(j)} = \overline{\text{span}\psi_{ik}^{(j)};\ \forall\, k \in Z}\,. \tag{2.7}$$

It follows from (2.2) that the V_i subspaces have a nested property. If the scaling and the wavelet functions satisfy the orthonormality condition, then the subspaces $\{W_i^{(j)}\}$ form an orthogonal decomposition of the $L^2(R)$ function space and are related to the V_i nested subspaces by,

$$V_i = V_{i+1} \oplus [\oplus_{j=1}^{M-1} W_{i+1}^{(j)}]\,. \tag{2.8}$$

A function $f(x) \in V_0 \subset L^2(R)$ can be constructed from a discrete sequence $a(k) \in l^2(R)$ in the form,

$$f(x) = \sum_k a(k)\phi(x - k) \tag{2.9}$$

$f(x)$ can also be expressed in terms of the sum of projections onto subspaces V_i and $W_i^{(j)}$ as,

$$f(x) = \sum_k c(k)\phi_{i,k}(x) + \sum_{j=1}^{M-1}\sum_k d_j(k)\psi_{i,k}^{(j)}(x)\,. \tag{2.10}$$

The expansion coefficients can be expressed as,

$$\begin{aligned} c(k) &= \langle f, \phi_{i,k}\rangle, \\ d_j(k) &= \langle f, \psi_{i,k}^{(j)}\rangle, \quad j = 1, \ldots, M-1\,. \end{aligned} \tag{2.11}$$

Using (2.2) and (2.3) in (2.11), it can be shown that,

$$c(k) = \frac{1}{\sqrt{M}}\sum_l a(l)h(Mk - l)\,, \tag{2.12}$$

$$d^{(j)}(k) = \sum_l a(l)h^{(j)}(Mk - l)\,, \tag{2.13}$$

which is equivalent to processing the sequence $a(k)$ with a set of linear time-invariant filters of impulse responses $p_j = \frac{1}{\sqrt{M}}h^{(j)}(k)$ and down sampling filter outputs by M. The M-band wavelet system has been in the focus of several recent

investigations. Noteworthy advantages of M-band wavelet systems over two-band wavelet systems are their richer parameter space which leads to a greater variety of compactly supported wavelets. These are practically implementable and have their ability to achieve more rapidly a given frequency resolution as a function of decomposition scale. This facts provide greater freedom and flexibility in choosing time frequency tilling.

3. A Brief Survey of Literature

3.1. *Document segmentation*

With the advent of modern publishing technologies, the layout of today's documents has been a very complex affair. Most of them contain not only text and background regions, but also graphics, tables and pictures. Therefore scanned documents must often be segmented before other document processing techniques, such as compression or rendering, can be applied. Documents can be more effectively represented by separating out the text and the graphics/image part and storing the text as ASCII (character) set and the graphics/image part as bitmaps. So research in the area of document image analysis has become an important issue.

3.1.1. *Previous work*

Several useful techniques for text-graphics segmentation,[28] the most popular amongst these being the *top-down* and *bottom-up* approaches. The most common *top-down* techniques are *run length smoothing*[29,30] and the *projection profiles*.[31,32] *Top-down* approaches first split the document into blocks which are then identified and subdivided appropriately, in terms of columns first and then split them into paragraphs, text lines and may be into words. Wong *et al.*[30] have proposed a technique called the run length smoothing algorithm (RLSA) to partition a binary document image into blocks. Each block is then classified as text or picture according to some statistical features. A similar algorithm has also been investigated by Wang *et al.*[31] and Chauvet *et al.*[29] A more detailed survey of these approaches can be found in a paper by Haralick.[33] Recent block-based segmentation algorithms are developed mostly for gray-scale or color document images. Some of these algorithms, use features extracted from the discrete cosine transform (DCT) coefficients to separate text blocks from picture blocks. For example Konstantinides and Tretter[34] used DCT block activity measure. Some assumed these blocks to be only rectangular.[35] The *top-down*/block-based methods are not suitable for skewed texts as these methods are restricted only to rectangular blocks. Whereas the *bottom-up* methods are typically variants of the *connected components*,[36] which iteratively group together components of the same type. It starts from the pixel level and form the higher level descriptions of the printed regions of the document (words, text lines, paragraphs etc.).[37] The major drawbacks of the connected

component method are that it is sensitive to character size, scanning resolution, inter-line and inter-character spacings.

Several other approaches use the contours of the white space to delineate text and non-text regions.[38] These methods can only be applied to low noise document images which are highly structured, that is all objects are separated by white background and objects do not touch each other.

Each of the above methods relies to a great extent on *a priori* knowledge about the rectangularity of major blocks, consistency in horizontal and vertical spacings, independence of text, graphic and image blocks and/or assumptions about textual and graphical attributes like font size, text line orientation etc. So these methods cannot work in a generic environment. It is desirable to have segmentation techniques which do not require any *a priori* knowledge about the content and attributes of the document image, as any such knowledge might not be available, in some applications. In addition these methods operate on thresholded images. In case, if the input is a degraded image due to poor image capturing conditions, the extraction of the corresponding appropriate thresholded image becomes a difficult task.

Jain and Bhattacharjee's[39] method has been able to overcome these restrictions and does not require an *a priori* knowledge of the document to be processed. The document segmentation is achieved by a texture segmentation scheme using Gabor filter as the feature extractor. One major drawback of this approach is that, the use of Gabor filter makes it computationally very expensive. Randen and Husøy[40] have proposed a method using critically sampled infinite impulse response (IIR) QMF banks for extracting features. Both the aforementioned methods do not take into consideration the possibility of overlapped/mixed classes. Etemad *et al.*[41] have developed an algorithm for document segmentation using multi-scale wavelet packet feature vectors and fuzzy local decision information.

More recent research on document segmentation include works by Choi and Baraniuk[42] which is based on wavelet-domain hidden Markov tree (HMT). Tang *et al.*[43] have introduced a mixture-state document segmentation method based on wavelet and the hidden Markov tree (HMT) models. Li and Gray[44] have developed an algorithm for segmenting document images into four classes: background, photograph, text, and graph. Features used for classification are based on the distribution patterns of wavelet coefficients in high frequency bands. Harit *et al.*[45] have presented a new model-based document image segmentation scheme that uses XML-DTDs (eXtensible Markup Language Document Type Definitions).

3.2. *Segmentation of remotely sensed image*

The segmentation of different land cover regions of a remotely sensed image has been recognized as a complex problem. These images usually have poor illumination and are highly dependent on the environmental conditions. Spatial resolution of these images are also very low, e.g., of the order of 20 m $\times$ 20 m. A natural scene generally

contains many objects (regions), e.g., vegetation, water bodies, habitation, concrete structures, open spaces etc., and these regions are not very well defined because of spatial ambiguities. Moreover, the gray value assigned to a pixel is the average reflectance of different types of land covers present in the corresponding pixel area. Assigning unique class levels with certainty is thus a problem of remotely sensed images.

Remotely sensed images contain information on a large range of scales and the frequency structure changes throughout the signal (i.e. non-periodic signal). In remote sensing perspective, the different resolution of the imagery may present different signatures of the objects of interest. So it is important to understand how information changes over different scales of imagery. This problem leads naturally to multi-resolution type analysis which could be done most effectively using wavelets. Moreover, wavelet theory is well suited in the area of study where signals are complex and non-periodic. Furthermore, wavelets are particularly good at describing a scene in terms of the scale of the textures in it.

In remotely sensed image analysis, texture is considered to be the visual impression of coarseness or smoothness caused by the variability or uniformity of image tone. These textural properties of a remotely sensed image provide valuable information for image segmentation, where different object regions are treated as different texture classes. So the first step of image analysis task basically boils down to a multi-texture segmentation problem. Note that segmentation of these images is necessary in order to identify regions of vegetation, habitation, water bodies, city area etc. Here we incorporate wavelet based texture segmentation algorithm for identifying the various object regions in a remotely sensed image.

3.2.1. *Previous work*

Haralick *et al.*[46] used gray level co-occurrence features to analyze remotely sensed images. They computed gray level co-occurrence matrices in four directions (0°, 45°, 90° and 135°). For a seven-class problem, they are able to achieve 80% classification accuracy. Rignot and Kwok[47] analyzed SAR (Synthetic Aperture Radar) images using texture features computed from gray level co-occurrence matrices. However, they supplemented these features with knowledge about the properties of SAR images. The use of various texture features had been studied for analyzing SAR images by Du.[48] He used Gabor filters for extracting texture features and successfully segmented SAR images into categories of water, new forming ice, older ice, and multi-year ice.

The work of Mecocci *et al.*[49] presented a wavelet-based algorithm combined with a *fuzzy c-means* classifier. Lindsay *et al.*[50] used the 1D discrete wavelet transform (DWT) based on Daubechies wavelet filter. A wavelet-based texture feature set was derived and used in a work by Fukuda *et al.*[51] The set consists of the energy of sub-images obtained by the overcomplete wavelet decomposition of local regions a in SAR image.

Simard *et al.*[52] studied the use of a decision tree classifier and multi-scale texture measures for the extraction of thematic information on the tropical vegetation cover, from the Global Rain Forest Mapping (GRFM) JERS-1 SAR mosaics. The aim of the work by Niedermeier[53] was to show, how coastline can be derived from SAR images by using wavelet and active contour methods. In the first step, edge detection method suggested by Mallat *et al.*[54] was applied to SAR images to detect all edges above a certain threshold. A block-tracing algorithm (BA) then determines the boundary area between land and water.

Several other wavelet-based segmentation for geoscience and remote sensing applications have also been reported in the literature.[55,56]

Other approaches to segmentation of remotely sensed images include various fuzzy thresholding techniques.[57] Genetic algorithm as a classifier has been investigated. In the domain of satellite imagery. This was used for partitioning different land cover regions from satellite images, having complex/overlapping class boundaries in Ref. 58 Muchoney and Williamson[59] had shown that neural network classifiers can provide improved supervised classification results, which are significantly better than that of traditional classification algorithms such as the Bayesian (maximum likelihood [ML]) classifier.

4. Proposed Methodologies

4.1. *Application of two-class texture segmentation: Document image segmentation*

Some of the common difficulties that occur in document segmentation are,

- Differences in font size, column layout, orientation and other textual attributes.
- Document skewed with text regions having different orientations.
- Combinations of varying text and background gray level.
- Text regions touching or overlapping with non-text regions.
- Irregular layout structures with non-convex or overlapping object boundaries.

We develop a texture based document image segmentation scheme, which takes care of all the above observations. The present work is based on the assumption that the text part in the document image comprise one texture class and the non-text part as the other. Basically this is a two-texture segmentation problem. It is already well known that textures can be characterized by their energies. A composite texture can be discriminated if it is possible to obtain information about the texture signal energies. The basic idea is to decompose the composite image into different frequency bands at different scales. The objective is to transform the edges between textures into detectable discontinuities. This creates the feature maps which give a measure of local energy around each pixel over small windows. In the present work we apply a two-texture segmentation scheme.

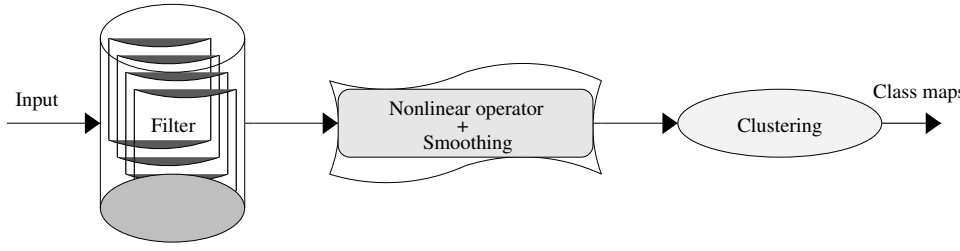


Fig. 1. Fast iterative implementation of the algorithm used for extracting texture features.

4.1.1. *Computing texture features*

An important aspect of texture analysis is to develop a set of texture measures (features) that can successfully discriminate arbitrary textures. In texture segmentation a local neighborhood is investigated. The goal is to extract some implicit local information that lets similar and dissimilar areas to be recognized. The feature extraction scheme consists of three main operating modules as shown in Fig. 1. The first module is the filtering stage, which is followed by a nonlinear operation stage and a smoothing filter (the latter two constitute the local energy estimator). The final module is the clustering stage. The objective of the filtering is to transform the edges between textures into detectable discontinuities (i.e. to extract local frequencies of the textures). While the purpose of the local energy estimator, is to estimate the energy of the filter output in a local region, the clustering step gives the final class maps.

M-band wavelet filters:

The filter bank in essence is a set of bandpass filters with frequency and orientation selective properties. The filtering stage consists of orthogonal and linear phase M-band wavelet filters.[4] These filters have perfect reconstruction with quadrature mirror filter (PR-QMF) structure and are symmetrical. Intuitively, for good edge boundary localization, it is desirable to have a filter with compact spatial domain representation, while for reliable discrimination of different texture frequency contents the filter should have a good frequency response localization and high stop band attenuation. Also symmetry of the filter responses is an important factor. A non-symmetric filter response consistently leads to edge detection error and consequently higher classification error. The filters we have chosen satisfy all of the above requirements since, QMF filters have significantly compact frequency response.

The texture image is decomposed into $M \times M$-channels, corresponding to different directions and resolutions. The $M(= 4)$-band wavelet system consists of the scaling filter ϕ and the wavelet filters ψ_m for $m = 1, 2, 3$. The impulse responses of these filters are denoted by $h(n)$ and $g_m(n)$ respectively, and are tabulated in Table 1.

Table 1. Filter coefficients for eight-tap four-band wavelet transform.

No. of taps (n)	$h(n)$	$g_1(n)$	$g_2(n)$	$g_3(n)$
0	−0.067371764	−0.094195111	−0.094195111	−0.067371764
1	0.094195111	0.067371764	−0.067371764	−0.094195111
2	0.40580489	0.56737176	0.56737176	0.40580489
3	0.56737176	0.40580489	−0.40580489	−0.56737176
4	0.56737176	−0.40580489	−0.40580489	0.56737176
5	0.40580489	−0.56737176	0.56737176	−0.40580489
6	0.094195111	−0.067371764	−0.067371764	0.094195111
7	−0.067371764	0.094195111	−0.094195111	0.067371764

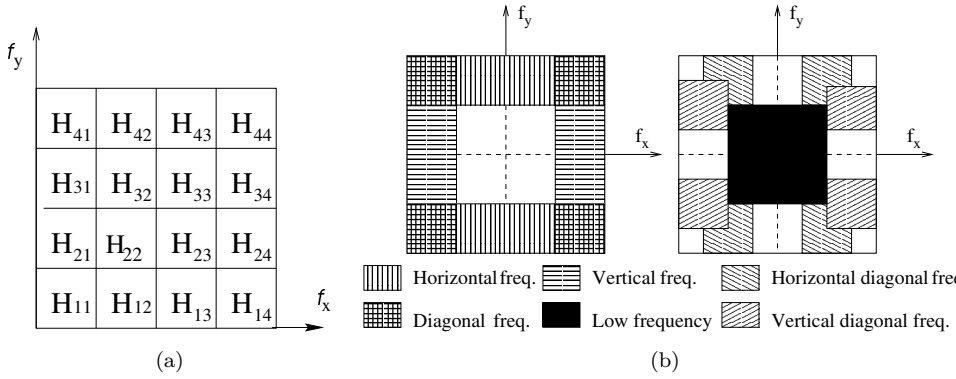


Fig. 2. (a) Frequency bands corresponding to decomposition filters, (b) frequency sector representations for filtering in horizontal (H), vertical (V) and diagonal (D) directions.

In this work we have developed for M^2-channel, the 2D separable transform by the tensor product of M-band 1D wavelet filters but without any subsampling. The (x, y)th resolution cell is obtained via the filtering steps

$$H_{11} = H_1(\omega_x)H_1(\omega_y) \tag{4.14}$$

and

$$H_{m'_x,m'_y} = H_{m'_x}(\omega_x)H_{m\prime_y}(\omega_y) \quad \text{for } m'_x = m'_y = 2, 3, 4\,, \tag{4.15}$$

where H denotes the transfer function of a filter. The decomposition of the image into $M \times M = (16)$ channels is illustrated in Fig. 2(a).

Since the spectral response to edges of an image is strongest in direction perpendicular to the edge, while it decreases as the look direction of the filter approaches that of the edge. Therefore we can perform edge detection by using 2D filtering as follows:

- *horizontal edges* are detected by high pass filtering on columns and low pass filtering on rows.

- *vertical edges* are detected by low pass filtering on columns and high pass filtering on rows.
- *diagonal edges* are detected by high pass filtering on columns and high pass filtering on rows.
- *horizontal-diagonal edges* are detected by high pass filtering on columns and low pass filtering on rows.
- *vertical-diagonal edges* are detected by low pass filtering on columns and high pass filtering on rows.

A typical edge detection filter corresponding to a particular direction, covers a certain region in the 2D spatial frequency domain. This is illustrated in Fig. 2(b), where f_x and f_y are the horizontal and vertical frequencies. Based on this concept several wavelet decomposition filters are possible which are given by, $\sum_{\text{Reg}} H_{\hat{m}_x,\hat{m}_y}$ for $\hat{m} = 1, \ldots, 4$, where "*Reg*" denotes the frequency fsector corresponding to a certain direction and scale.

Since the filter system we are using here is orthogonal and has QMF structure, therefore

$$\sum_{\hat{m}_x=1}^{4} \sum_{\hat{m}_y=1}^{4} H_{\hat{m}_x\hat{m}_y} H^*_{\hat{m}_x\hat{m}_y} = 1\,.$$

This signifies that the resulting 2D filters treats all the frequencies in a resolution cell equally. The number of channels as well as the number of possible filter combinations depend on the value of M. The decomposition filters $\sum_{\text{Reg}} H_{\hat{m}_x\hat{m}_y}$ are formed as follows for different directions in increasing level of resolutions.

⋆ Horizontal direction:

$$filt_{\text{hor}_1} = H_{12}\,,$$

$$filt_{\text{hor}_2} = H_{12} + H_{13}\,,$$

$$filt_{\text{hor}_3} = H_{12} + H_{13} + H_{14} + H_{24}\,.$$

⋆ Vertical direction:

$$filt_{\text{ver}_1} = H_{21}\,,$$

$$filt_{\text{ver}_2} = H_{21} + H_{31}\,,$$

$$filt_{\text{ver}_3} = H_{21} + H_{31} + H_{41} + H_{42}\,.$$

⋆ Diagonal direction:

$$filt_{\text{diag}_1} = H_{22}\,,$$

$$filt_{\text{diag}_2} = H_{22} + H_{33}\,,$$

$$filt_{\text{diag}_3} = H_{22} + H_{33} + H_{44}\,.$$

⋆ Horizontal-diagonal direction:

$$filt_{\mathrm{hdiag}_1} = H_{12}\,,$$

$$filt_{\mathrm{hdiag}_2} = H_{12} + H_{23}\,,$$

$$filt_{\mathrm{hdiag}_3} = H_{12} + H_{23} + H_{34}\,.$$

⋆ Vertical-diagonal direction:

$$filt_{\mathrm{vdiag}_1} = H_{21}\,,$$

$$filt_{\mathrm{vdiag}_2} = H_{21} + H_{32}\,,$$

$$filt_{\mathrm{vdiag}_3} = H_{21} + H_{32} + H_{43}\,.$$

These filter responses basically give a measure of signal energies at different directions and scales, the corresponding filtered images are denoted by F_{hor_j}, F_{ver_j}, F_{diag_j}, F_{hdiag_j} and F_{vdiag_j} for $j = 1, 2, 3$ as shown in Fig. 3.

Local energy estimator:

The next step is to estimate the energy in the filter output within a local region around each pixel. The local energy estimate is utilized for the purpose of identifying areas in each channel, where the bandpass frequency components are strong resulting in a high energy value and the areas where it is weak into a low energy value.

A nonlinearity is needed in order to discriminate texture pairs with identical mean brightness and second-order statistics. We have used the modulus operator as the nonlinearity. One reason for choosing this nonlinear operator is that it is parameter-independent; which means it is independent of the dynamic range of the input image and also of the filter amplification.

The raw wavelet coefficients are inefficient representation of texture cues. They can split textured information into different frequency channels, but without providing any local pixel information. To calculate local features of an image, we slide a fixed size window on the wavelet coefficients of an image. Compute the local statistics in each individual window. These values are used as feature values of the central pixels of these windows. A wide variety of textural measure is available. In this study, energy measure for texture segmentation is considered, which indicates the textural uniformity and pixel pairs repetitions. Energy is usually defined in terms of a squaring nonlinearity. However, in this study *average absolute deviation* (from the mean) has been used as a generalized energy definition.

This is followed by a Gaussian low pass (smoothing) filter of the form,

$$h_G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x^2+y^2)}\,,$$

where σ defines the spatial extent of the averaging filter.

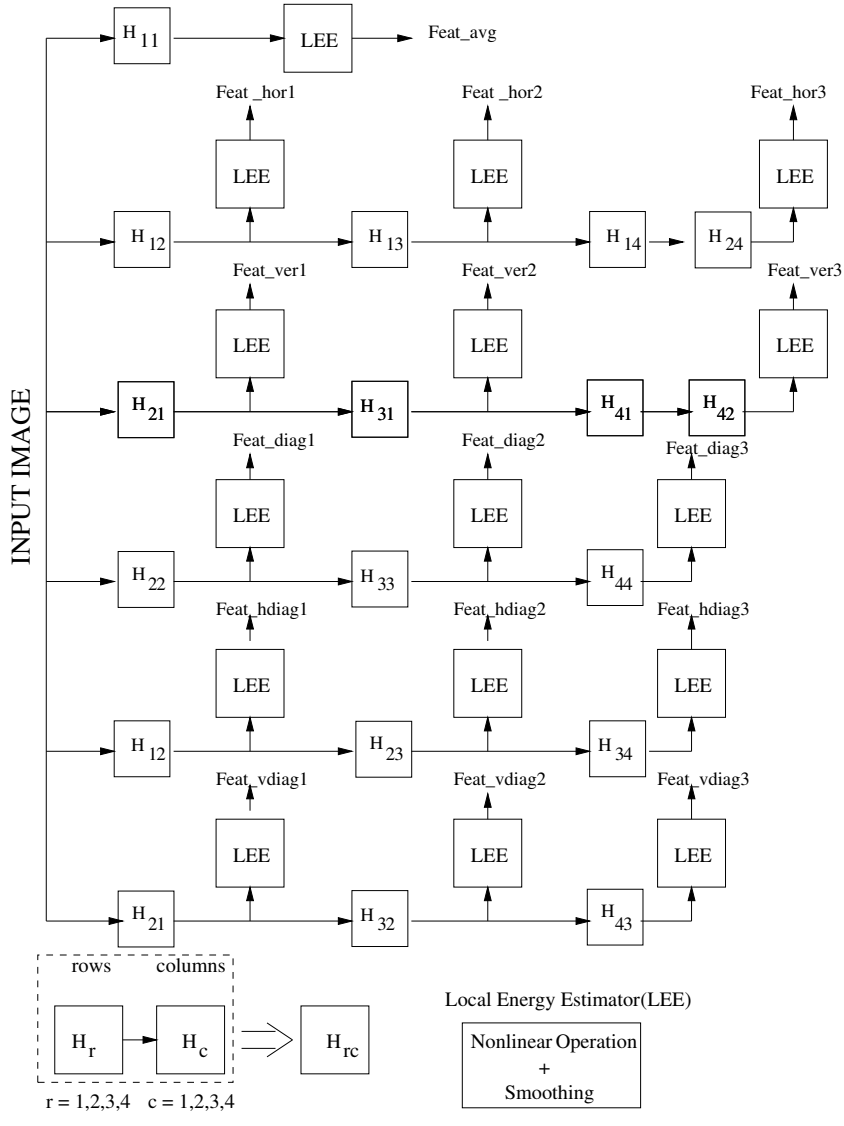


Fig. 3. Block diagram illustrating the algorithm of our scheme.

Formally, the feature image $Feat_b(x, y)$ corresponding to filtered image $F_b(x, y)$ is given by,

$$Feat_b(x, y) = \sum_{(x', y') \in G_{xy}} \Gamma(F_b(x', y') h_G(x - x', y - y')) \,, \tag{4.16}$$

where $b = \text{hor}_i, \text{ver}_i$ etc., $\Gamma(\cdot)$ is the nonlinear function and G_{xy} is a $G \times G$ window centered at pixel with coordinates (x, y).

An important parameter in this regard is the Gaussian averaging window size G. More reliable measurement of texture feature demands larger window size. On the other hand, more accurate localization of region boundaries requires smaller window. This is because averaging blurs the boundaries between textured regions.

Choice of σ of the smoothing filter:

The choice of the space constant σ of the averaging filter is very crucial. The problem is how to determine the size of the smoothing filter. If we want to estimate the local energy of an image with low spatial frequency, the smoothing filter must have a wide unit impulse response, while narrower filter is to be used for higher frequency content image.

In the present work we set the smoothing filter size based on the measure of the spectral content of the image. *Spectral flatness measure* (SFM) gives a measure of the overall image activity. The spectral flatness of a digital image of size $N1 \times N2$ is defined as the ratio of the arithmetic and the geometric mean of the Fourier coefficients.[60] For two-dimensional digital image this can be expressed as,

$$\text{SFM} = \frac{[\prod_{x=0}^{N1-1} \prod_{y=0}^{N2-1} |\hat{I}(x,y)|^2]^{\frac{1}{N1N2}}}{\frac{1}{N1N2} \sum_{x=0}^{N1-1} \sum_{y=0}^{N2-1} |\hat{I}(i,j)|^2} \tag{4.17}$$

$\hat{I}(i,j)$ is the (x,y)th Fourier coefficient of the two-dimensional image. SFM has a dynamic range of $[0,1]$.

Highly active image means SFM close to 1, then the image has many edges or has predominantly high frequencies. So the image requires a smaller window for smoothing. Moderately active image has SFM of somewhat moderate value within 0 and 1. That means the image contains moderate range of frequencies and requires a moderate window size for good feature extraction. Finally an image with low SFM is lowly active and has low spectral content. It is evident that this type of image would require larger window size of the smoothing filter.

We have found experimentally that the spatial extent of the windows, for these three categories of image activities range from 11×11 to 31×31. With these choices we have worked successfully on all the test images that have been experimented. Thus we can adaptively select the size of the averaging window depending on the spectral content of the image. Our scheme is adaptive in the sense that we do not use any fixed windowing operation and hence it can accommodate diverse set of textured images as input.

4.1.2. *Feature integration and post-processing*

Having obtained the feature images, the main task is to integrate these feature images to produce a segmented output. We emphasize on the feature extraction (representation) part in this work. So we have used a *k-means* clustering algorithm.

After the class maps are obtained, segmentation results can be improved by post processing. The *k-means* clustering algorithm labels each pixel independently and does not take into account the high correlation between neighboring pixels. A more sophisticated algorithm would incorporate some neighborhood constraint into the segmentation process, such as relaxation labelling. Here we have used median filtering to simulate the benefit of a local constraint.

Algorithm:

The two-class texture segmentation algorithm based on the M-band wavelet decomposition is illustrated in the block diagram Fig. 3.

This algorithm consists of the following steps:

- The input image is first decomposed into $M \times M$ channels by wavelet analysis without down sampling as referred in Sec. 4. In this work we have used an eight-tap 4-band wavelet,[4] so in all we get 16 decomposition channels as discussed in Sec. 4, which means the image is decomposed into 16 sub-bands, i.e. in all 16 features can be computed. But we ignore the filtered output corresponding to the low frequency channel H_{11}. Filtered outputs F_{hdiag_1} and F_{vdiag_1} are nothing but F_{hor_1} and F_{ver_1} respectively. So these two sub-bands are also ignored. In effect, now we are left with 13 frequency bands.
- These filter outputs are subject to the nonlinear operation followed by smoothing as discussed in Sec. 4, which then form the feature images $Feat_b$.
- We have a matrix of $n_F \times N_D$, where n_F is the number of feature elements (selected) in each vector and N_D is the total data size (the total number of pixels in the input image). The features are normalized between $[0, 1]$ along each column of the feature matrix and subject to the clustering algorithm. This step gives us the class map corresponding to the composite texture image.

4.1.3. *Unsupervised segmentation of document images*

In this section, the results on implementation of our algorithm on several document images are provided. The images considered are both structured and unstructured.

Test images:

The test documents are scanned from parts of pages of the "Times of India" (TOI) and "Hindustan Times" both of which are popular news dailies in India.

- Structured document image with non-overlapping text and non-text regions scanned from "Times of India", which are as follows:
  - — Fig. 4(a) shows a document image of size 512×512.
  - — Fig. 5(a) shows the same image rotated by an angle of $90°$.
  - — Fig. 6(a) shows the same image skewed by an angle of $25°$.

(a) (b)

Fig. 4. (a) A portion of a typical page of TOI, (b) the two-class segmentation.

(a) (b)

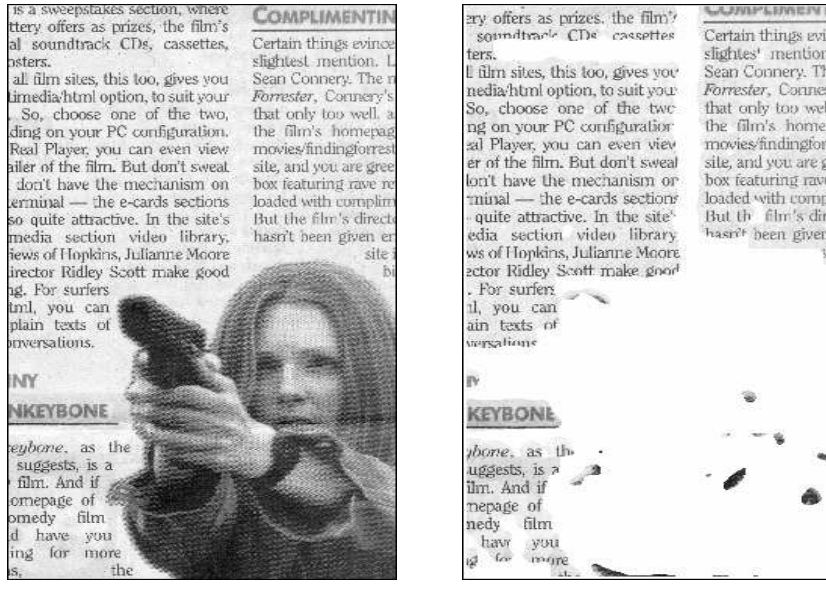
Fig. 5. (a) Same image rotated by 90°, (b) two-class segmentation.

- Highly unstructured images with overlapped/mixed classes of size 512 × 512 scanned from "Hindustan Times", these are:

  — Fig. 7(a) shows a test image with non-convex and overlapping object boundaries.
  — Fig. 8(a) the text portions have different orientations as well as gray values and have different font sizes.
  — Fig. 9(a) shows a test image of size 512 × 512 that has been used by Randen and Husøy.[40]

(a) (b)

Fig. 6. (a) Same image skewed by an angle 25°, (b) corresponding image segmentation.



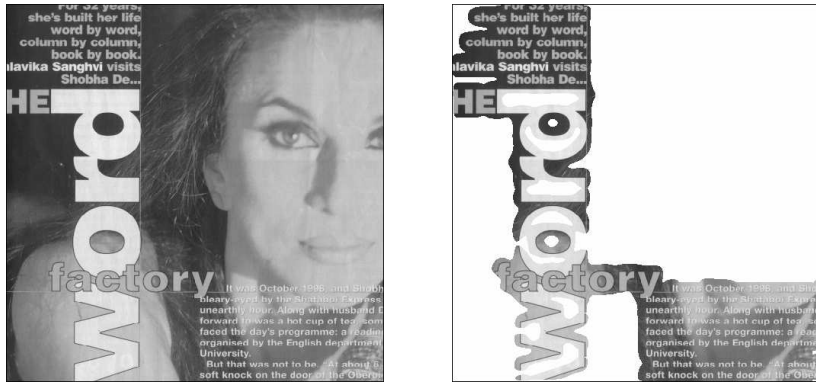
(a) (b)

Fig. 7. (a) Test image with non-convex and overlapping object boundaries, (b) segmented result.

Experimental results:

Out of the total 16 features possible in our decomposition scheme we have found that in most of the cases the number of features are limited between 3 to 5.

Figure 4(a) shows a 512×512 pixel scanned image of a portion of a typical page of TOI, and the successful two-class segmentation of the image (Fig. 4(b)).



(a) (b)

Fig. 8. (a) Test image with text portions having different orientations and gray values as well as different font sizes, (b) segmented result.

(a) (b)

(c) (d)

Fig. 9. (a) Segmentation results of the test image used in Ref. 24 using (b) our algorithm, (c) Randen and Husøy[18] and (d) Etemad *et al.*[25]

In order to prove the efficacy of our algorithm we apply this technique to segment the same image rotated by 90° (Fig. 5(a)). The image is rotated about an axis through the center of it in a clockwise direction. The transformed coordinates in the rotated plane for a rotation of θ are given by,

$$x' = x\cos\theta + y\sin\theta \quad \text{and} \quad y' = -x\sin\theta + y\cos\theta\,.$$

The segmentation results in Fig. 5(b) show that our method is almost invariant to any specific layout of the document.

The same image is also given a skewed transform about an axis through the center and the transformed coordinates for a skewness of ϕ are given by,

$$x' = x + y\tan\phi \quad \text{and} \quad y' = y\,.$$

The image is given a skewness by an angle of 25° (Fig. 6(a)) and the result is shown in Fig. 6(b), so we can also say that our algorithm is somewhat independent of skewness.

So long we have been concentrating on structured data with non-overlapping text and non-text regions. But there are several instances where documents are highly unstructured. We have experimented on several such data for an extensive study.

To prove the efficacy and robustness of our algorithm, different kinds of documents with various types of geometric transformation were used. We have also tested the algorithm on images with non-convex and overlapping object boundaries (Fig. 7(a)). The segmentation result (Fig. 7(b)) shows that the algorithm can efficiently identify the text and non-text regions in the documents.

In Fig. 8(a) the text portions have different orientations as well as gray values and have different font sizes. The corresponding text segmentation result is shown in Fig. 8(b).

Throughout the experiment, our basic objective has been to segment out the text part from the graphics part as accurately as possible. To compare the efficiency of our method with other methods, we have used the same data that has been used by Randen and Husøy[40] (Fig. 9(a)). Using the proposed algorithm, It is found that although some of the graphics part of Fig. 9(b) are misclassified as text data, But overall quality of the result is excellent as far as text identification is concerned. The headings of two different font sizes could not be identified very accurately by Randen's method[40] (Fig. 9(c)), but have been possible using our method. The segmentation result obtained using classical wavelet packets and features suggested by Etemad *et al.*[41] (Fig. 9(d)) is also presented here for a comparative study.

It is seen that there is a significant improvement in segmentation result using $M(M > 2)$-band wavelets compared to the classical wavelet packets, where $M = 2$. This may be explained by the additional information obtained by decomposing the image into a higher number of subbands for larger values of M. Also high frequency signals with relatively narrow bandwidth are better resolved using wavelets with higher values of M. This may also be attributed to the bandpass nature of higher

bands, that capture additional texture information. Moreover, in our method we have not subsampled the image, whereas in both the methods used here for the purpose of comparison, the image has been subsampled. Suitability of our method over the subsampled methods may be explained by the fact that subsampling reduces the size of the subbands at higher levels of decomposition and can possibly bias the decomposition.

Although the method used in Ref. 40 has a better computational savings than the approach of Jain and Bhattacharjee,[39] none of the above-mentioned works have considered test data that are unstructured, or have overlapping text and non-text regions. Whereas the work of Etemad *et al.*[41] have considered all these types of data. They have used fuzzy local decision information for classification. This step clearly reduces ambiguity between the various classes and hence gives better performance. But it has been shown that the set of features that our method has extracted is appreciably compact and authentic, which gives good enough segmentation. We however conjecture that incorporation of the concept of fuzziness may produce better results than that reported in this paper.

4.2. *Application of multi-class texture segmentation: Remotely sensed image segmentation*

All of these above methods use supervised classification where *a priori* knowledge about the images are supplied. We use an unsupervised technique to carry out this segmentation where no *a priori* knowledge about the image is available.

The next section discusses the filtering technique for extraction of features used in our investigation.

4.2.1. *Extraction of multi-class texture features*

This section provides the methodology for extracting multi-scale wavelet features of a texture image. It involves M-band wavelet packet filtering of an input image followed by adaptive basis selection. Subsequently, features are computed from this set of selected basis by using a nonlinear operator and smoothing filter. These features are then evaluated and selected using a neuro-fuzzy algorithm (described in Sec. 4.2.3). The entire methodology is depicted in Fig. 10.

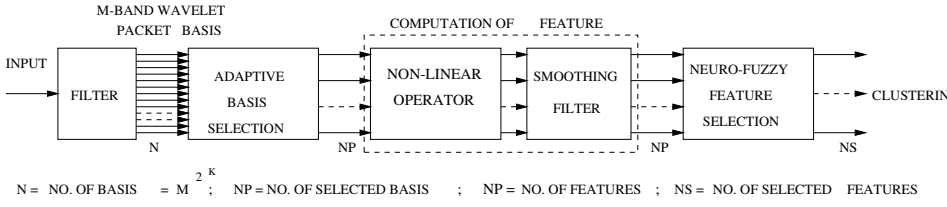


Fig. 10. Experimental setup.

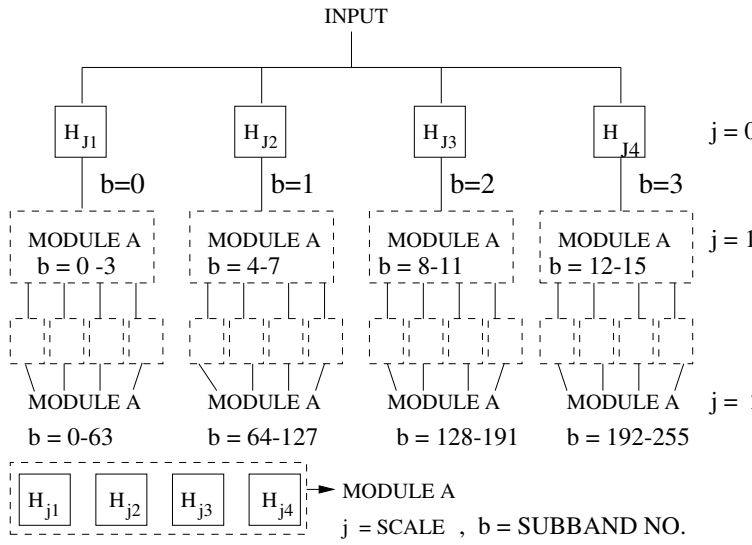


Fig. 11. Tree structure of DM-bWPF transform and related indices.

M-band wavelet packet filters:

In the filtering stage, we make use of the same M-band ($M = 4$) wavelet used in the previous section. The discrete M-band wavelet packet transform (DMbWPT) used in this work results in a tree structured multi-band extension of the M-band wavelet transform. Thus a finer and adjustable resolution is allowed at high frequencies as compared to the 2-band case.

The classical decomposition method requires a down sampling by a factor M (where $M = 2$ for standard wavelet) at each scale. But this decomposition is not translation invariant. As mentioned earlier, a discrete M-band wavelet packet frame (DMbWPF) is used, which is similar to DMbWPT, except that no down sampling occurs between scales (levels of decomposition).

Figure 11 shows a general tree structure of discrete M-band wavelet packet frame decomposition. Scale $j = 0$ corresponds to the highest resolution of the image, that is the image itself before decomposition.

The filter responses in the frequency domain $H_{j,r}(\omega)$ (for $r = 1, \ldots, 4$) and at level j are generated as follows:

$$H_{j,r} = H_{0,r}(M^j\omega) \quad \text{for} \quad r = 1, \ldots, 4\,. \tag{4.18}$$

Let $\hat{I}_b^j(\omega)$ be the Fourier transform of the input signal $I(x)$ for subband (frequency channel) b at decomposition level j. For $0 \le b \le 4^j - 1$ and $r = 1, 2, 3, 4$, we have

$$\hat{I}^{j+1}_{4b+(r-1),r}(\omega) = H_{j,r}(\omega)\hat{I}^j_{b,r}(\omega)\,. \tag{4.19}$$

From the filter bank theoretic point of view,[61] this corresponds to a filter bank with channel filters $\{filt^j_{b,r}(\omega)|r = 1, \ldots, 4\}$. $filt^j_{b,r}(\omega)$ are given by the following

recursive relation

$$filt^{0}_{0,r}(\omega) = H_{0,r}(\omega)\,,$$

and

$$\begin{aligned} filt^{j+1}_{4b+(r-1),l}(\omega) &= H_{j,r}(\omega) filt^{j}_{b,r}(\omega)\,, \\ &= H_{0,r}(M^{j}\omega) filt^{j}_{b,r}(\omega)\,. \end{aligned} \tag{4.20}$$

For images we simply use tensor product extension for which the channel filters are written as

$$filt^{j}_{b,r[x\times y]}(\omega_x, \omega_y) = filt^{j}_{b,r[x]}(\omega_x) filt^{j}_{b,r[y]}(\omega_y)\,.$$

At scale $j = 1$, the image is first decomposed into $M \times M$ channels using all the filters $H_{0,r}$ with $r = 1, 2, 3, 4$, and without down sampling. The process is repeated for each of the subbands for subsequent scales. Figure 11 shows a general tree structure of discrete M-band wavelet packet frame decomposition. Module-A in Fig. 11 comprises of all the filters $H_{0,r}$ with $r = 1, 2, 3, 4$.

4.2.2. *Adaptive basis selection*

The objective of wavelet decomposition of a signal/image is to collect most of the energy in a fewer number of subbands. An appropriate approach for obtaining a suitable wavelet transform, for texture feature extraction is to detect the most significant frequency channels and then to decompose those channels further. This leads naturally to a tree structured wavelet transform. It is also usually not essential to decompose all the subbands in each scale to achieve the full tree of decomposition. Moreover, an M-band wavelet packet decomposition gives M^{2^J} number of bases, for a decomposition depth of J. It is quite evident that an exhaustive search to determine the optimal basis from this large set is computationally very expensive.

In order to find out a suitable basis without going for a full decomposition, we propose an adaptive decomposition algorithm using a criterion of maximum textural measures extracted from each of the subbands. Then the most significant subbands are identified and it is decided whether further decomposition of the particular channel would generate more information or not. This search is computationally efficient and enables one to zoom into any desired frequency channel for further decomposition.[2]

For this purpose, the image is first decomposed into $M \times M$ channels using the 2D M-band wavelet transform without down sampling (oversampled) as shown in Fig. 11. Energy for each subband is then computed. Among these subbands, those for which energy values exceed $\varepsilon_1\%$ 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 $\varepsilon_2\%$ 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 $\varepsilon_1 = 2 - 5\%$ and $\varepsilon_2 = 50\%$ are good choices for the images we have considered here. This simple top-down splitting technique performs well for most images.

4.2.3. *Selection of wavelet features using a neuro-fuzzy method and segmentation*

The wavelet features extracted (as mentioned in the previous section), are evaluated and then selected using a neuro-fuzzy feature selection under unsupervised learning. The method is a modification of an earlier approach.[22] This modification enables one to handle large data sets in an efficient manner. These selected features are then used for the purpose of segmentation of multi-texture images.

Fuzzy feature evaluation index and membership function:

The neuro-fuzzy feature selection method[22] involves formulation of a fuzzy feature evaluation index E followed by its minimization in connectionist framework. E for a set of transformed features is defined as

$$E = \frac{2}{s(s-1)} \sum_{p} \sum_{q \neq p} \frac{1}{2} [\mu_{pq}^{T}(1 - \mu_{pq}^{O}) + \mu_{pq}^{O}(1 - \mu_{pq}^{T})] \,. \tag{4.21}$$

Here μ_{pq}^{O} and μ_{pq}^{T} are the degree that both the pth and qth patterns belong to the same cluster in the n-dimensional original feature space, and in the n'-dimensional ($n' \leq n$) transformed feature space respectively. μ values determine how similar a pair of patterns are in the respective features spaces. s is the number of samples on which the feature evaluation index is computed.

E decreases as the membership value μ representing the degree of belonging of pth and qth patterns to the same cluster in the transformed feature space tends to either 0 (when $\mu^{O} < 0.5$) or 1 (when $\mu^{O} > 0.5$). This means, if the inter-cluster/intra-cluster distances in the transformed space increase/decrease, the feature evaluation index of the corresponding set of features decreases. Therefore, our objective is to select those features for which the evaluation index becomes minimum; thereby optimizing the decision on the similarity of a pair of patterns with respect to their belonging to a cluster.

The membership function μ_{pq} in a feature space, satisfying the characteristics of E (4.21), may be defined as[22]

$$\mu_{pq} = \begin{cases} 1 - \dfrac{d_{pq}}{D} & \text{if } d_{pq} \leq D \,, \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{4.22}$$

d_{pq} is a distance measure which provides similarity (in terms of proximity) between the pth and qth patterns in the feature space. Note that the higher the value of d_{pq}, the lower is the similarity between pth and qth patterns, and vice versa. D

is a parameter which reflects the minimum separation between a pair of patterns belonging to two different clusters. When $d_{pq} = 0$ and $d_{pq} = D$, we have $\mu_{pq} = 1$ and 0, respectively.

The term D in (4.22) may be expressed as $D = \alpha d_{\max}$ 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 membership function (4.22). The higher the value of α, more will be the degree, and vice versa.

The distance d_{pq} (4.22) can be defined in many ways. Considering weighted Euclidean distance, we have

$$\begin{aligned} d_{pq} &= \left[\sum_i w_i^2 (x_p i - x_q i)^2\right]^{1/2}, \\ &= \left[\sum_i w_i^2 \chi_i^2\right]^{1/2}, \quad \chi_i = (x_{pi} - x_{qi}), \end{aligned} \tag{4.23}$$

where $w_i \in [0, 1]$ represents weighing coefficient corresponding to ith feature. The terms x_{pi} and x_{qi} are values of ith feature (in the corresponding feature space) of pth and qth patterns, respectively. $d_{\max}$ is defined as

$$d_{\max} = \left[\sum_i (x_{\max\, i} - x_{\min\, i})^2\right]^{1/2}, \tag{4.24}$$

where $x_{\max\, i}$ and $x_{\min\, i}$ are the maximum and minimum values of the ith feature in the corresponding feature space.

The membership value μ_{pq} depends on w_i. The values of w_i (< 1) make the μ_{pq} function of (4.22) flattened along the axis of d_{pq}. The lower the value of w_i, the higher is the extent of flattening. The weight w_i in (4.23) reflects the relative importance of the feature x_i in measuring the similarity (in terms of distance) of a pair of patterns. The higher the value of w_i, the more is the importance of x_i in characterizing a cluster or discriminating various clusters. $w_i = 1(0)$ indicates most (least) importance of x_i.

Our objective is to minimize the evaluation index E (4.21) which involves the terms μ^O and μ^T. The evaluation index E (4.21) is a function of $\mathbf{w}$, if we consider ranking of n features, in a set. The problem of feature selection/ranking thus reduces to finding a set of w_i's for which E becomes minimum; w_i's indicating the relative importance of x_i's. The task of minimization as in Ref. 22, is performed using gradient-descent technique in a connectionist framework under unsupervised mode. This is described in the appendix.

4.2.4. *Segmentation of IRS images*

The features thus selected are used for segmenting a texture image. In this section, the results on implementation of our algorithm on several remotely sensed images are provided. The images considered are two IRS-1A images.

Fig. 12. IRS-1A image of Calcutta (band-4).

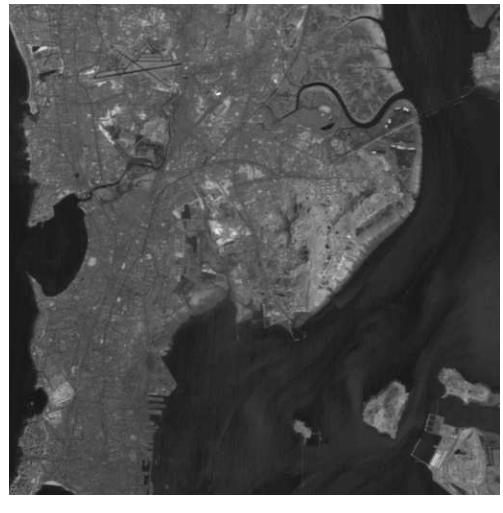


Fig. 13. IRS -1A image of Bombay (band-4).

Data sets:

The IRS-1A images (Figs. 12 and 13) were obtained from Indian Remote Sensing Satellite which is a circular sun-synchronous satellite, rotating around the earth at the rate of 14 orbits per day, at an altitude of 904 km and a repetition cycle of 22 days (NRSA 1986). This satellite is equipped with two different sensors LISS (Linear Imaging Self Scanner)-I and LISS-II. LISS-I has a spatial resolution of 72.5 m $\times$ 72.5 m while that for LISS-II is 36.25 m $\times$ 36.25 m. The IRS-1A images

used for this work were taken using the scanner LISS-II in the wavelength range 0.45–0.86 μm. The whole spectrum range is decomposed into four spectral bands namely,

- band 1 — blue band of wavelength 0.45–0.52 μm,
- band 2 — green band of wavelength 0.52–0.59 μm,
- band 3 — red band of wavelength 0.62–0.68 μm and
- band 4 — near infra red wavelength 0.77–0.86 μm.

The image in Fig. 12 covers an area around the city of Calcutta in the near infrared band. In this figure the prominent black stretch across them is the river *Hoogly*. There is a prominent light patch on the bottom right corner, which is the *Salt Lake stadium* and the black patches nearby are the *fisheries*. In the upper right part of the image, there is a distinct line structure corresponding to the *airport runway*. In total there are five major classes in which the regions of Calcutta IRS images can be classified. These are *water bodies* (WB), *vegetation* (VEG), *habitation* (HAB), *city area* (CA) and *open spaces* (OS).

Figure 13, shows a part of the city of Bombay in the near infrared band. The elongated city area is surrounded by the Arabian sea. There is a *concrete structure* (on the right side top corner) connecting Bombay to New Bombay. On the southern part of the city, there are several islands including the famous *Elephanta islands*. The *dockyard* is situated on the south-eastern part of Bombay, which can be seen as a set of three finger like structure. On the upper part of the image, towards left, there is a distinct criss-crossed structure, which corresponds to the *Santa Cruz airport*. In total the regions of the Bombay image can be classified into six major classes, *viz.*, *turbid water* 1 (TW1), *turbid water* 2 (TW2), *concrete* (CONCR), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS). The sea water is decomposed into two classes TW1 and TW2 for better classification since they have somewhat different reflectance properties due to variation in sea water density (as seen in Fig. 13).

All the images considered in this investigation are of size 512×512. Due to poor illumination, the actual object classes present in the input images are not visible clearly. So we have presented the histogram equalized images in Figs. 12 and 13, which highlight the different land cover regions. But the algorithms are implemented on the actual inputs.

Experimental results:

Here we demonstrate the effectiveness of our methodology over several remotely sensed images as described in Sec. 4.2.4. In order to validate the importance of neuro-fuzzy feature evaluation, we show how the feature dimensionality can be greatly reduced after feature evaluation. The test images have several fine (line) structures (roads and bridges). In order to detect these structures, the local window size has been kept small (3×3).

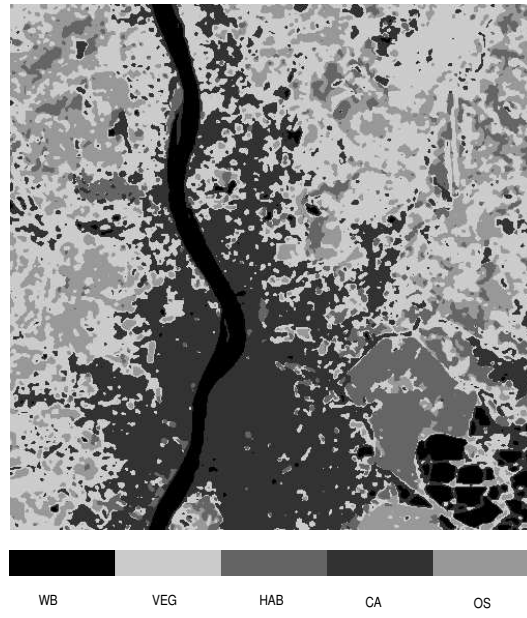


Fig. 14. Segmented output of Calcutta IRS-1A image 12 with $c = 5$ and $\beta = 3.65887$.

The total number of features considering all the bands of IRS-1A Calcutta image, are found to be 10 using the proposed feature extraction methodology described in Sec. 4.2.1. Figure 14 shows the segmented output of the IRS-1A image using all these 10 features considering the number of classes to be $c = 5$. The β index is found to be 3.65887. The neuro-fuzzy feature evaluation algorithm reduces the number of features to only one for which the segmentation result is given in Fig. 15. The value of β index is found to be 3.83431 signifying an improvement in segmentation quality. The *stadium* and *fisheries* as well as the *airport runway* are quite distinctly discernible in the segmented output (Fig. 15) as compared to Fig. 14. Comparing Figs. 14 and 15 we find that we get better segmentation output in the *Salt lake stadium* area, also the compactness of the various classes present in the IRS-1A Calcutta image increases after the feature evaluation step.

The segmentation result of the IRS-1A image of Bombay is shown in Fig. 16 considering number of classes to be $c = 6$, and the β value is found to be 8.01462. The number of extracted features is found to be 12 as obtained by Sec. 4.2.1. The feature dimensionality is reduced to one after neuro-fuzzy feature evaluation. As in the previous case, β index (8.21309) shows better segmentation quality. In the segmented output after the feature evaluation step (Fig. 17), it can be seen that the *dockyard*, the *concrete structure* connecting Bombay and New Bombay and *Santa Cruz airport* are very well detected. Even for this image of Bombay we can reiterate the above comments. Comparing Figs. 16 and 17 we find that various objects and regions (*santa cruz airport*, *dockyard*, *roads and bridges* etc.) are identified more

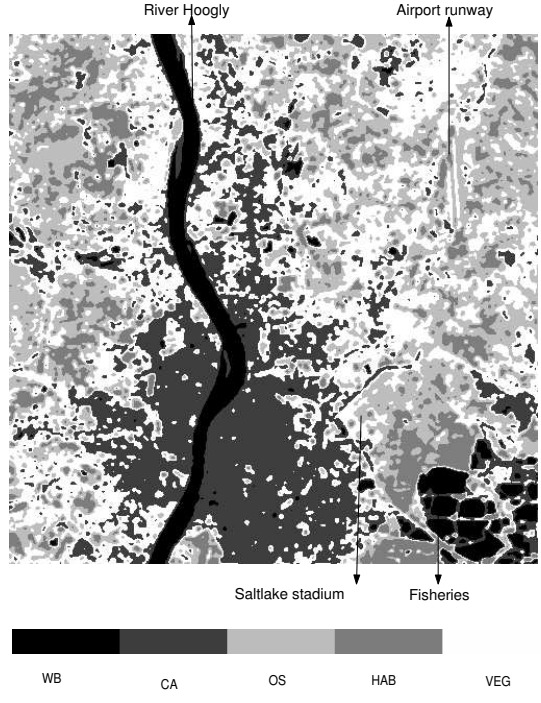


Fig. 15. Segmented output of Calcutta IRS-1A image 12 with $c = 5$ and $\beta = 3.84578$.

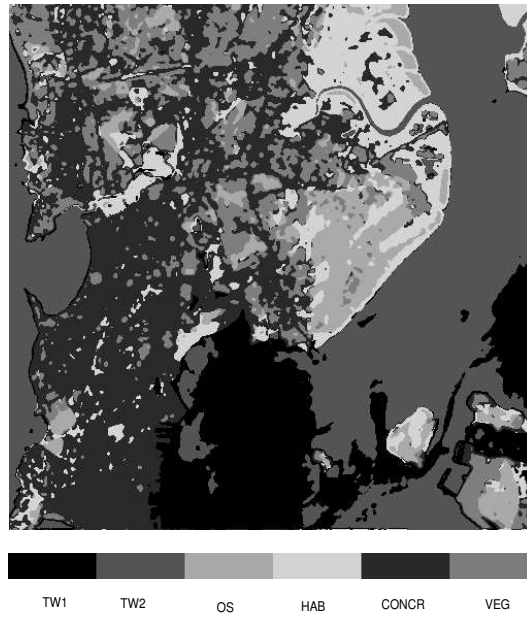


Fig. 16. Segmented output with $c = 6$ and $\beta = 8.01462$.

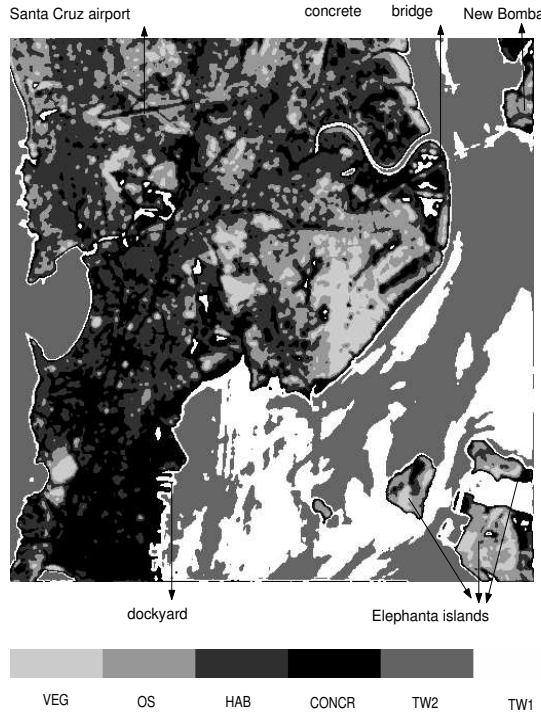


Fig. 17. Segmented output with $c = 6$ and $\beta = 8.21309$.

prominently. Moreover, the feature evaluation step increases the compactness of the various classes similar to the case of IRS-1A Calcutta image.

5. Conclusion

In the first part of this paper we have presented a technique for segmenting out the text part from the non-text part of a document image based on textural cues using M-band wavelet filters. Both structured and unstructured documents have been studied. Several of the difficulties that occur in document segmentation have been addressed in this investigation. These include: different fonts, orientation, other textual attributes; skewed document and text regions have different orientations; text regions touching or overlapping with non-text regions etc. The algorithm has been able to successfully segment several document types mentioned above. In this study, no information about the input have been considered.

In contrast to most traditional methods for text-graphics segmentation, we do not make any *a priori* assumptions about the font size, scanning resolution, column layout, orientation etc. of the input, that is our approach is purely unsupervised. The results indicate that M-band wavelets have the efficacy to discriminate between textures, and can be effectively applied for document segmentation.

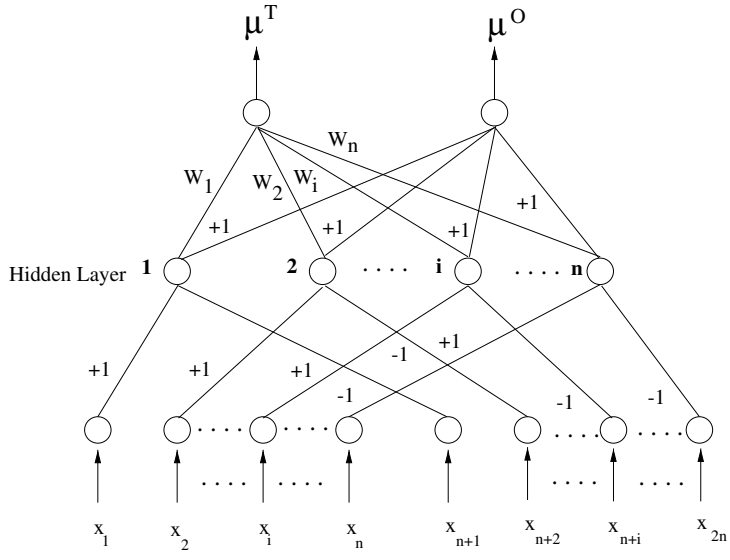


Fig. 18. A schematic diagram of the neural network model for feature selection.

In the second part of this paper, we have presented a technique based on the theory of wavelet and neuro-fuzzy hybridization paradigm taken together for image segmentation. It is shown that this method can be applied in the domain of remotely sensed images for their segmentation. The feature extraction method simultaneously splits the lower as well as the higher frequency bands, results in a tree structure. This enables efficient characterization of remotely sensed images. The neuro-fuzzy feature evaluation method helps in searching more efficiently most significant features from a remotely sensed image where the various classes are overlapping in nature. Both the feature extraction and neuro-fuzzy feature evaluation schemes are unsupervised and do not require the knowledge of the number of classes. Moreover, the methodology does not require *a priori* knowledge about the spatial relationship of different regions in these images.

The features obtained by the feature extraction method have been able to segment the remotely sensed images satisfactorily. It has been found for the IRS-1A images that although ultimately only one significant feature is selected for segmentation. But almost all the desired classes have been obtained satisfactorily in the segmented output. Moreover, the compactness of the various class regions increases after the neuro-fuzzy feature evaluation step.

Appendix A.

Connectionist model:

The network (Fig. 18) consists of an input, a hidden and an output layer.[22] The input layer consists of a pair of nodes corresponding to each feature, i.e. the number

of nodes in the input layer is $2n$, for n-dimensional (original) feature space. The hidden layer consists of n number of nodes which compute the part χ_i^2 of (4.23) 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 (4.21) is computed from these μ-values off the network.

Input nodes receive activations corresponding to feature values of each pair of patterns. A jth node in the hidden layer is connected only to an ith and $(i+n)$th input nodes via connection weights $+1$ and -1, respectively, where $j, i = 1, 2, \ldots, n$ and $j = i$. The output node computing μ^T-values is connected to a jth node in the hidden layer via connection weight W_j $(= w_j^2)$, whereas that computing μ^O-values is connected to all the nodes in the hidden layer via connection weights $+1$ each.

During learning, each pair of patterns is presented at the input layer and the evaluation index is computed. The weights W_j's are updated in order to minimize the index E. The task of minimization of E (4.21) with respect to $\mathbf{W}$ is performed using gradient-descent technique, where the change in W_j $(\triangle W_j)$ is computed as

$$\triangle W_j = -\eta \frac{\partial E}{\partial W_j}, \quad \forall\, j\,, \tag{A.1}$$

where η is the learning rate. Note that $d_{\max}$ is directly computed from the unlabelled training set. The values of $d_{\max}$ and α are stored in both the output nodes for the computation of D. The details concerning the operation of the network are given in Ref. 22.

Modification of the neuro-fuzzy algorithm for handling large data:

As we have seen in Sec. 4.2.1, the number of patterns generated for an $N \times N$ image is $N^2 = s$. Each of these patterns corresponds to a pixel and has all the multiscale wavelet feature. Therefore, for selecting an optimal set of features, 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 very large number. This requires a very high computational time. In order to avoid such situation, i.e. in order to make the neuro-fuzzy algorithm computationally more efficient, we, first of all, apply a clustering algorithm (e.g., k-means clustering algorithm) on the entire feature space, for grouping the data, and the cluster centers $\mathbf{cen}_q$'s are noted. Then two sets of samples, namely, $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p, \ldots, \mathbf{x}_{N^2}\}$ and $S_c = \{\mathbf{cen}_1, \mathbf{cen}_2, \ldots, \mathbf{cen}_c\}$ are formed. That is, S is the entire training set, and S_c is the set of c cluster centers (for c clusters) obtained by the clustering algorithm. Now the similarity between the patterns and these cluster centers are computed, instead of computing it for every pair of patterns. These cluster centers are considered as representatives (prototypes) of all the 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 $s_c = |S_c| \ll s$. This is a way of making the algorithm scalable.

Note that the clustering algorithm at this stage provides a first-hand knowledge about the various objects in an image. Based on this knowledge, neuro-fuzzy feature selection algorithm is applied, in order to remove the redundant and confusing features, and thereby improving the segmentation result.

E, after convergence, attains a local minimum and then the weights ($W_j = w_j^2$) of the links connecting hidden nodes and the output node computing μ^T-values, indicate the order of importance of the features. Note that this unsupervised method performs the task of feature selection without clustering the feature space explicitly and does not need to know the number of clusters present in the feature space.

Acknowledgment

M.A. is grateful to the Council of Scientific & Industrial Research (CSIR), New Delhi, India for the Research Associateship, vide grant No. 9/93(66)2000-EMR-I, in carrying out her research work.

References

1. M. Acharyya and M. K. Kundu, *Two texture segmentation using m-band wavelets*, in *Proc. 15th. Int. Conf. Pattern Recognition*, Barcelona, Spain, September 2000 (IEEE Press, 2001), pp. 405–408.
2. M. Acharyya and M. K. Kundu, *Adaptive basis selection for multitexture segmentation by m-band wavelet packet frame*, in *Proc. 2001 Int. Conf. Image Processing*, Thessaloniki, Greece, October 2001 (IEEE Press, 2002), pp. 622–625.
3. P. Steffen, P. N. Heller, R. A. Gopinath and C. S. Burrus, *Theory of regular m-band wavelet bases*, *IEEE Trans. Signal Processing* **41** (1993) 3497–3510.
4. O. Alkin and H. Caglar, *Design of efficient m-band coders with linear phase and perfect reconstruction properties*, *IEEE Trans. Signal Processing* **43** (1995) 1579–1590.
5. T. Greiner, J. P. Casel and M. Pandit, *Texture analysis with a texture matched m-channel wavelet approach*, *IEEE Int. Conf. Acoustics, Speech, Signal Processing* (IEEE, 1993), Vol. 5, pp. V-129-V-132.
6. Y. Chitre and A. P. Dhawan, *M-band wavelet discrimination of natural textures*, *Pattern Recognition* **32** (1999) 773–789.
7. M. Acharyya and M. K. Kundu, *Multiscale segmentation of document images using m-band wavelets*, in *Proc. 9th Int. Conf. Computer Analysis of Images and Patterns*, Vol. LNCS 2124, Warsaw, Poland, September 2001 (Springer-Verlag, 2002), pp. 510–517.
8. M. Acharyya and M. K. Kundu, *An adaptive approach to unsupervised texture segmentation using m-band wavelet*, *Signal Processing* **81** (2001) 1337–1356.
9. M. Acharyya and M. K. Kundu, *Wavelet-based texture segmentation of remotely sensed images*, in *Proc. 2001 Int. Conf. Image Analysis and Processing*, Palermo, Italy, September 2001 (IEEE Press, 2002), pp. 69–74.
10. X. D. Huang, Y. Akiri and M. A. Jack, *Hidden Markov Models for Speech Recognition* (Edinburgh Univ. Press, 1990).
11. R. Chellapa, *Two-dimensional discrete gaussian markov random field models for image processing*, in *Prog. Pattern Recog.*, eds. L. N. Kanal and A. Rosenfeld (North-Holland, 1985), pp. 79–112.
12. P. C. Chen and T. Pavlidis, *Segmentation by texture using correlation*, *IEEE Trans. Pattern Anal. Machine Intell.* **5** (1983) 64–69.

13. F. Farrokhnia, *Multichannel filtering techniques for texture segmentation and surface quality inspection*, PhD thesis, Michigan State University, 1990.
14. N. Saito and R. R. Coifman, *Local discriminant bases and their applications*, *J. Math. Imaging Vision* **5** (1995) 337–358.
15. R. R. Coifman and M. V. Wickerhauser, *Entropy based algorithms for best basis selection*, *IEEE Trans. Inform. Theory* **38** (1992) 713–718.
16. S. Mallat, *Zero-crossings of a wavelet transform*, *IEEE Trans. Pattern Anal. Machine Intell.* **37** (1993) 1019–1033.
17. J. M. Steppe and K. W. Bauer, Jr., *Improved feature screening in feedforward neural networks*, *Neurocomputing* **13** (1996) 47–58.
18. R. K. De, N. R. Pal and S. K. Pal, *Feature analysis*: *Neural network and fuzzy set theoretic approaches*, *Pattern Recog.* **30** (1997) 1579–1590.
19. M. Pregenzer, G. Pfurtscheller and D. Flotzinger, *Automated feature selection with a distinctive sensitive learning vector quantizer*, *Neurocomputing* **11** (1996) 19–29.
20. W. A. C. Schmidt and J. P. Davis, *Pattern recognition properties of various feature spaces for higher order neural networks*, *IEEE Trans. Pattern Analy. Machine Intell.* **15** (1993) 795–801.
21. K. Etemad and R. Chellapa, *Separability-based multiscale basis selection and feature extraction for signal and image classification*, *IEEE Trans. Image Processing* **7** 1453–1465 (1998).
22. S. K. Pal, R. K. De and J. Basak, *Unsupervised feature evaluation*: *A neuro-fuzzy approach*, *IEEE Trans. Neural Networks* **11** (2000) 366–376.
23. S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition*: *Methods in Soft Computing* (John Wiley, 1999).
24. Eds. S. K. Pal, A. Ghosh and M. K. Kundu, *Soft Computing for Image Processing* (Physica Verlag, 2000).
25. I. Daubechies, *Orthogonal bases for compactly supported wavelets*, *Comm. Pure Appl. Math.* **41** (1988) 909–996.
26. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, 1992).
27. S. Mallat, *A theory for multiresolution signal decomposition*: *The wavelet representation*, *IEEE Trans. Pattern Anal. Machine Intell.* **11** (1989) 674–693.
28. S. N. Srihari, *Document image understanding*, in *Proc. IEEE Computer Society Fall Joint Computer Conf.* (IEEE, 1986), pp. 87–96.
29. P. Chauvet, J. Lopez-Krahe, E. Taflin and H. Maitre, *System for an intelligent office document analysis*, *recognition and description*, *Signal Processing* **32** (1993) 161–190.
30. F. M. Wahl, K. Y. Wong and R. G. Kasey, *Block segmentation and text extraction in mixed text/image documents*, *Comput. Graphics Image Processing* **20** (1982) 375–390.
31. D. Wang and S. N. Srihari, *Classification of newspaper image blocks using texture analysis*, *Computer Vision*, *Graphics*, *Image Processing* **47** (1989) 327–352.
32. M. Krishnamoorthy, G. Nagy, S. Seth and M. Viswanathan, *Syntactic segmentation and labelling of digitized pages from technical journals*, *IEEE Trans Pattern Anal. Machine Intell.* **15** (1993) 737–747.
33. R. M. Haralick, *Document image understanding*: *Geometric and logical layout*, in *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, June (IEEE, 1994), 385–390.
34. K. Konstantinides and D. Tretter, *A method for variable quantization in jpeg for improved text quality in compound documents*, in *Proc. IEEE Intl. Conf. Image Processing* (IEEE, 1998), pp. 565–568.

35. S. Sural and P. K. Das, *A two-step algorithm and its parallelisation for the generation of minimum containing rectangles for document image segmentation*, in *Proc. Int. Conf. Document Analysis and Recognition* (IEEE Press, 1999), pp. 173–176.
36. F. Lebourgeois, Z. Bublinski and H. Emptoz, *A fast and efficient method for extracting text paragraphs and graphics from unconstrained documents*, in *Proc. Int. Conf. Pattern Recognition* (IEEE Press, 1992), pp. 272–276.
37. C. L. Tan, B. Yuan, W. Huang and Z. Zang, *Text/graphics separation using pyramid operations*, in *Proc. Int. Conf. Document Analysis and Recognition* (IEEE Press, 1999), pp. 169–172.
38. A. Antaonacopulos, *Page segmentation using the description of the background, Computer Vision and Image Understanding* **70** (1998) 350–369.
39. A. K. Jain and S. Bhattacharjee, *Text segmentation using gabor filters for automatic document processing*, *Machine Vision Appl.* **5** (1992) 169–184.
40. T. Randen and J. H. Husøy, *Segmentation of text/image documents using texture approaches*, in *Proc. NOBIM-konferansen-94* (1994), pp. 60–67.
41. K. Etemad, D. Doermann and R. Chellappa, *Multiscale segmentation of unstructured document pages using soft decision integration*, *IEEE Trans. Pattern Anal. Machine Intell.* **19** (1997) 92–96.
42. H. Choi and R. G. Baraniuk, *Multiscale image segmentation using wavelet-domain hidden markov models*, *IEEE Trans. Image Processing* **10** (2001) 1309–1321.
43. Y. Y. Tang, Y. Hou, J. Song and X. Yang, *Mixture-state document segmentation using wavelet-domain hidden Markov tree models*, in *Proc. 2nd Int. Conf. Wavelet Analysis and its Application*, *ICWAA 2001*, Hong Kong (December 2001), pp. 230–236.
44. J. Li and R. M. Gray, *Context-based multiscale classification of document images using wavelet coefficient distributions*, *IEEE Trans. on Image Processing* **9** (2000) 1604–1616.
45. G. Harit, S. Chaudhury, P. Gupta, N. Vohra and S. D. Joshi, *A model guided document image analysis scheme*, in *Proc. Sixth Int. Conf. Document Analysis and Recognition* (2001) 1137–1141.
46. R. M. Haralick, K. Shanmugam and I. Dinstein, *Texture feature for image classification*, *IEEE Trans. Systems*, *Man and Cybernatics* **8** (1973) 610–621.
47. E. Rignot and R. Kowk, *Extraction of textural features in sar images*: *Statistical model and sensitivity*, in *Proc. IEEE Geoscience and Remote Sensing Symposium* (IEEE, 1990).
48. L. J. Du, *Texture segmentation of sar images using localized spatial filtering*, in *Proc. Int. Geoscience and Remote Sensing Symposium* (1990), pp. 1983–1986.
49. A. Mecocci, P. Gamba, A. Marazzi and M. Barni, *Texture segmentation in remote sensing images by means of packet wavelets and fuzzy clustering*, in *Proc. European Symposium on Satellite and Remote Sensing II* **SPIE 2584** (1995) 142–157.
50. R. W. Lindsay, D. B. Percival and D. A. Rothrock, *The discrete wavelet transform and the scale analysis of the surface properties of sea ice*, *IEEE Trans. Geosc. Remote Sensing* **34** (1996) 771–787.
51. S. Fukuda and H. Hirosawa, *A wavelet-based texture feature set applied to classification of multifrequency polarimetric sar images*, *IEEE Trans. Geosc. Remote Sensing* **37** (1999) 2282–2286.
52. M. Simard, S. S. Saatchi and G. D. Grandi, *The use of decision tree and multiscale texture for classification of jers-1 sar data over tropical forest*, *IEEE Trans. Geosci. Remote Sensing* **38** (2000) 2310–2321.
53. A. Niedermeier, E. Romaneesen and S. Lehner, *Detection of coastline sar images using wavelet methods*, *IEEE Trans. Geosci. Remote Sensing* **38** (2000) 2270–2281.

54. S. Mallat and S. Zhong, *Characterization of signals from multiscale edges*, *IEEE Trans. Pattern Anal. Machine Intell.* **14** (1992) 710–732.
55. D. C. Tseng, H. M. Tsai and C. C. Lai, *Unsupervised texture segmentation for multispectral remote-sensing images*, in *Proc. Int. Conf. Pattern Recognition*, *ICPR'98* (IEEE Press, 1998), pp. 1630–1632.
56. P. Thitimajshima, *Multiresolution fuzzy clustering for sar image segmentation*, in *Proc. Int. Geoscience and Remote Sensing Symposium*, *IGARSS '99* (1999), Vol. 5, pp. 2507–2509.
57. S. K. Pal, A. Ghosh and B. U. Shankar, *Segmentation with remotely sensed images with fuzzy thresholding, and quantitative evaluation*, *Int. J. Remote Sensing* **21** (2000) 2269–2300.
58. S. Bandyopadhyay and S. K. Pal, *Pixel classification using variable string genetic algorithms with chromosome differentiation*, *IEEE Trans. Geosci. Remote Sensing* **39** (2001) 303–308.
59. D. Muchoney and J. Williamson, *A Gaussian adaptive resonance theory neural network classification algorithm applied to supervised land cover mapping using multitemporal vegetation index data*, *IEEE Trans. Geosci. Remote Sensing* **39** (2001) 1969–1977.
60. N. Jayant and P. Noll, *Digital Coding of Waveforms*: *Principles and Applications to Speech and Video* (Prentice Hall, 1984).
61. M. Vetterelli and C. Herley, *Wavelets and filter banks*: *Theory and design*, *IEEE Trans. Signal Processing* **40** (1992) 2207–2232.