Applied Soft Computing 11 (2011) 2999–3011

Wavelet-fuzzy hybridization: Feature-extraction and land-cover classification of remote sensing images

B. Uma Shankar*, Saroj K. Meher, Ashish Ghosh

Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 18 December 2009
Received in revised form 4 August 2010
Accepted 28 November 2010
Available online 4 December 2010

Keywords:
Remote sensing
Land cover classification
Wavelet transform
Wavelet features
Fuzzy classification
Soft computing

ABSTRACT

A wavelet feature based supervised scheme for fuzzy classification of land covers in multispectral remote sensing images is proposed. The proposed scheme is developed in the framework of wavelet-fuzzy hybridization, a soft computing approach. The wavelet features obtained from wavelet transform on an image provides spatial and spectral characteristics (i.e., texture information) of pixels and hence can be utilized effectively for improving accuracy in classification, instead of using original spectral features. Four different fuzzy classifiers are considered for this purpose and evaluated using different wavelet features. Wavelet feature based fuzzy classifiers produced consistently better results compared to original spectral feature based methods on various images used in the present investigation. Further, the performance of the Biorthogonal3.3 (Bior3.3) wavelet is observed to be superior to other wavelets. This wavelet in combination with fuzzy product aggregation reasoning-rule outperformed all other methods. Potentiality of the proposed soft computing approach in isolating various land covers are evaluated both visually and quantitatively using indexes like β measure of homogeneity and *Xie-Beni* measure of compactness and separability.

1. Introduction

Classification of different land cover regions of remote sensing images is essential for efficient interpretation of them [1–3]. Normally, the images acquired from satellite mounted camera suffers from lot of problems including low illumination quality and rapid changes in environmental conditions. Also many times the spatial resolution is not very high. This makes the analysis of remote sensing images more complex and difficult. Basically the regions like vegetation, soil, water bodies, concrete structure, etc. of a natural scene are often not well separated leading to overlapping regions. Moreover, the gray value assigned to a pixel is an average reflectance of different types of land cover classes present in the corresponding pixel area. Therefore, a pixel may represent more than one class with varying degree of belonging. Thus assigning unique class label to a pixel with certainty is a difficult problem. Conventional methods (particularly non-fuzzy methods) cannot deal with the imprecise representation of geological information, normally encountered in processing of remote sensing images. To improve the representation, a graded belonging approach like fuzzy sets, a soft computing, is used which provide an effective solution to this problem [4,5].

Fuzzy set theory provides some useful techniques in overcoming the above shortcomings and allows a pixel to be a member of more than one category or class with graded membership [6–8]. Thus the fuzzy concept is a useful tool to deal with imprecise information and to get a reasonable classification of the land covers in remote sensing imagery. Many attempts have been made for remote sensing image analysis and classification using fuzzy sets [2,7,9–11]. Fuzzy classification of multispectral remotely sensed data was applied to estimate sub-pixel components in [7,12]. The fuzzy supervised method described by Melgani et al. [9] to classify land cover types used the concept of fuzzification, MINIMUM (*MIN*) reasoning rule (RR) and defuzzification on each pixel. Wang and Jamshidi [11] and Shackelford and Davis [13] described a hierarchical fuzzy classification method for high resolution multispectral data of an urban area. Bardossy and Samaniego [10] described a fuzzy rule based classification technique. In some other approaches [5,14] aggregation operators are used on the fuzzified value to get an aggregated decision on the available information. Combination of features/bands for a multispectral pixel pattern is also important with respect to different classes, particularly in the classification of ill-defined land cover classes. In this regard Ghosh et al. [15] proposed a fuzzy aggregation based classification approach for remotely sensed images.

Multispectral remotely sensed images comprise information over a large range of variation on frequencies (information), and these frequencies change over different regions (i.e., non-stationary behavior of the signal) which needs to be estimated properly for

* Corresponding author. Tel.: +91 33 2575 3107; fax: +91 33 2578 3357.
E-mail address: uma@isical.ac.in (B. Uma Shankar).

improved classification. The multispectral remote sensing image data have both spectral features with correlated bands and spatial features correlated in the same band (also known as spatial correlation). An efficient utilization of these spectral and spatial (contextual) information can improve the classification performance significantly compared to the conventional non-contextual information based classification methods (*k*-nearest neighbors, maximum likelihood, neural networks, etc.) [2,16]. Such conventional (non-contextual) approaches may be reasonable if spatial resolution is high or when the spectral intensities are well separated for different classes, which is rarely found in any real life data. For example, in the urban areas, the densities of the spectral intensities are seldom well separated. Thus it is important to decide whether the arrangements of spatial data or a transformation of the pixels to a different space that uncorrelated the information to separate the data into desired classes can be used as features.

Many efforts have been made to take the advantages of neighboring pixel information in [2,17–19]. These include texture features extracted from angular second moments, contrast, correlation, entropy, variance, etc., computed from the gray level co-occurrence matrices [17]. However, these methods are computationally expensive due to the estimation of autocorrelation parameters and transition probabilities. Also, the texture elements are not easy to quantify and it deals with the spatial distribution of the gray levels over a small portion of an image. Later on Gaussian Markov random fields (GMRF) [20] and Gibbs random fields [21] were used to characterize textures. Further, local linear transformations are also used to compute textural features [22]. These methods can model spatial interactions over relatively small neighborhoods on a single scale. Again the multispectral pixel information are usually insufficient to discriminate properly the different land cover classes having ill-defined and overlapping regions. Hence the use of both spectral and textural information in classification problems is more appropriate.

An efficient way to deal with a problem where use of the spectral and spatial (contextual) information is necessary to get proper classification is to recognize the images by a number of subsampled approximation in different resolution and desired scales. This approach is conventionally known as multiresolution analysis [23]. This scheme analyzes the coarse image first and gradually increases the resolution to analyze the finer details. To capture the local variations in the orientation and frequency of texture elements that lead to the nonstationary behavior of remotely sensed image textures, we need a joint spatial and frequency representation. Recently wavelet transform (WT) has received much attention as a promising tool for texture analysis in both spatial and frequency domains [23–27], as it has the ability to examine the signal at different scales. This means that the WT represents the input information in both spatial and frequency domains. Basically the WT coefficients represent the characteristics in frequency bands indicating the characteristic of the original pixel like frequency, spatial location and both; from where the WT coefficients originated. Thus the WT coefficients gather information of neighboring pixels and these are uncorrelated in spatial domain.

These characteristics of the WT motivated us to use it for extraction of non-stationary behavior of pixel information from remote sensing images. Some research studies have been made to use WT for classification of remote sensing images [24,28,29]. A comparative study of multiwavelet [30], wavelet, and shape features [17], for microcalcification classification in mammogram has been described in [31], where it is experimentally shown that the results of multiwavelet based approach were better compared to wavelet and statistical based feature. However, the computational complexities of multiwavelet based classification is much higher than the wavelet based ones. Recently, Huang et al. [32] used multiscale spectral and spatial information fusion using wavelet transform, to classify very high resolution satellite imagery with the support vector machine (SVM) as a classifier. Some recent works include the paper of Wong et al. [33] on the stitching defect detection, Zegarra et al. [34] on finger-print image retrieval, and the MRI image classification using wavelet in a Hybrid framework by El-Dahshan et al. [35].

Daugman pioneered the work in biometric identification by developing the Gabor wavelet based iris recognition [36]. Even today most of the research on iris recognition systems are centered around the wavelet feature based methods [37,38]. Apart from these, WT have been used in other applications. Zegarra et al. [39] describe a novel approach for personal identification based on a wavelet-based fingerprint retrieval system, which encompasses three tasks, namely, feature extraction, similarity measurement, and feature indexing. Grzegorzek et al. [40] described a system for texture-based probabilistic classification and localization of 3D objects in 2D digital images using wavelet features. In another study, a multiresolution analysis system for interpreting digital mammogram is proposed and tested by Rashed et al. [41]. This system is based on using fractional amount of the biggest wavelet coefficients in multilevel decomposition. Murtagh and Starck [42] described the method for extracting wavelet and curvelet features for image classification, and applied to aggregate mixture grading. In the field of remote sensing, WT has also acquired important place [43]. Hsu et al. [44] described a method for feature extraction of hyperspectral images using WT. Zhang et al. [45] propose a technique for fusion of multispectral and hyperspectral images to enhance the spatial resolution of the latter. The technique works in the wavelet domain and is based on a Bayesian estimation of the hyperspectral image, assuming a joint normal model for the images and an additive noise imaging model for the hyperspectral image.

The aim of the present work is three fold. First, we exploit the WT in extracting features (i.e., wavelet features) of the input patterns/pixels. Second, we use the fuzzy classifiers for land cover classification of remote sensing images, to deal with overlapping class problem. Third, we developed a hybrid approach in the framework of soft computing paradigm and utilize the advantages of both WT and fuzzy classifiers. We use WT by incorporating it as a preprocessor to extract both spatial and spectral information. The wavelet transform is quite useful in extracting features of the input patterns/pixels, which can be used in the next step of processing as input features for classification. In this experiment we have considered four fuzzy classifiers; (1) fuzzy product aggregation reasoning rule (FPARR), (2) fuzzy explicit (FE), (3) fuzzy maximum likelihood (FML), and (4) fuzzy *k*-nearest neighbor (F*k*-NN), because of their ability to perform reasonably well in all conditions [7–9,15]. The performance of these fuzzy classifiers are compared with different wavelets. The proposed scheme has been tested on various multispectral remote sensing images. Since the results were similar with respect to performance, we have reported the results of three images as typical examples, two images from four-band Indian Remote Sensing (IRS) satellite images [46] and one from three-band SPOT image [1]. Comparison of results showed that among the used wavelet features the performance of the Biorthogonal3.3 (Bior3.3) wavelet is superior to other wavelets. In the proposed wavelet-features based fuzzy classification scheme, the FPARR is yielding superior results compared to others. As a result, the hybridization of FPARR with Biorthogonal3.3 (Bior3.3) wavelet-features based scheme, boosted the performance to maximum.

The rest of the paper is organized as follows. Wavelet transform and extraction of wavelet features are described in Section 2. Fuzzy classification and the considered fuzzy classifier are briefly described in Section 3. A brief discussion on the performance measurement indexes are given in Section 4. A detail discussion on the results are presented in Section 5. The paper ends with a con-

clusion on the findings in Section 6, along with scope for future work.

2. Wavelet transform based feature extraction

Many research works have been carried out to deal with the non-stationary behavior of signals in an appropriate way. In this regard, efforts are made to overcome the disadvantages of the Fourier transform (FT), which assumes the signal to be stationary within its total range of analysis. WT [23,47] is an example of such a transform, which does not lose the spatial localization of the signal frequencies. The multiscale behavior of the WT analyzes or decomposes the signal in multiple scales, where each scale represents a particular coarseness of the analyzed signal. Thus the decomposition steps divide the signal into a set of signals of varying coarseness ranging from low frequency to high frequency components. Accordingly the WT tries to identify both the scale and space information of the events simultaneously. This makes the WT useful for signal feature analysis. This property is more useful for remote sensing data analysis, where any characteristic of the scene is first analyzed using low resolution and then analyze an area of interest in detail using an appropriate higher resolution level.

Since we are dealing with different textures of remotely sensed images, the decomposition of the signal into different scales, which can uncorrelate the data as much as possible without losing their distinguishable characteristics, is particularly more useful when the WT is done on an orthogonal basis [23]. The distinguishable characteristics of the original information preserved in the WT decomposition are spatio-geometrical information, energy at different scales, etc.; normally called the signature of the particular land cover. WT is also suitable for operations like feature extraction [48], parameter estimation [49] and reconstruction of the signal series because of its invertible properties [23].

The wavelets are the functions used in the transformation which act as a basis for representing many functions of the same family. A series of functions can be generated by translation and dilation of these functions called mother wavelets $\psi(t)$. The translation and dilation of the mother wavelet can be given by

$$\psi_{\gamma,\tau}(t) = |\gamma|^{-1/2}\psi\left(\frac{t-\tau}{\gamma}\right), \quad \gamma \neq 0 \text{ and } \gamma \varepsilon \mathbb{R}, \tau \varepsilon \mathbb{R} \tag{1}$$

where τ and γ are the translation and dilation parameters.

The following multiresolution scheme, given by Mallat [23], is used for implementation of wavelets.

2.1. Discrete WT and multiresolution analysis

2.1.1. 1-Dimensional

The discrete WT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into low frequency (approximation) and high frequency (details) band information. The decomposition of the signal into different frequency bands is obtained by successive high-pass and low-pass filtering of the signal.

In the reverse process of discrete WT, normally called inverse WT, the reconstructed signal may not be exactly the same as original one; however, it will be a good approximation. Due to these properties, WT is widely used for signal and image compression [50].

The decomposition operation in the WT halves the spatial resolution since only half of the number of samples now characterizes the entire signal. However, this operation doubles the frequency resolution, since the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency by half. The above procedure, which is also known as the subband coding, can be repeated for further decomposition. At every level, the filtering and *downsampling* will result in half the number of samples (and hence half the spatial resolution) and half the frequency band spanned (and hence double the frequency resolution). This process of subband coding is also known as multiresolution analysis [51].

2.1.2. 2-Dimensional

The two-dimensional (2-D) WT is performed by consecutively applying 1-D WT on rows and columns of the 2-D data. A 2-D WT, which is a separable filter bank in row and column directions, decomposes an image into four sub-images [23]. Fig. 1 shows this dyadic decomposition of a 2-D image for a 2-level decomposition. H and L in Fig. 1 denote a high-pass and low-pass filter, respectively. ↓2 denotes the *downsampling* (decrease the sample occurrence rate) by a factor of two. Thus in one level decomposition of the 2-D WT, four sub-sampled versions of the original image are obtained. Among them, one contains the WT coefficients in low frequency range called the approximation part, i.e., LL (Fig. 1) and three in high frequency range in three directions: vertical (LH), horizontal (HL) and diagonal (HH) called the detail parts. The wavelet coefficients in these subbands provide frequency information of the original signal in four different frequency bands maintaining their spatial position same as the original. In addition to this the coefficient values also preserve the neighborhood information. The decomposition of the wavelet coefficients can be extended to more than one level. The next level of decomposition can be performed on the approximation coefficient of the previous level. The other detailed coefficients are assumed to be redundant as they contain information in the high frequency bands and normally considered as noise. However, the next step of decomposition can be performed on all the coefficients obtained from the previous level. This decomposition method is called wavelet packet [23].

At the higher level of decomposition detailed information can be obtained. Thus the level of decomposition depends on the type of requirement and it varies with the problem in hand. To have an objective evaluation, we computed the average entropy [52], which provides a measure of information, of the images for each level. We found that the average entropy value is not changing much after a certain level of decomposition; and we decided to decompose up to that level. For the present experiments we stopped decomposition after second level only, as the entropy measure was not changing much after this.

From the WT coefficients the corresponding reconstructed images are obtained using inverse WT, which will be used subsequently as the extracted features of the original image for classification purpose. Thus we will have many extracted features from the input images and the number of images (features) will depend on the level of decomposition. This procedure is extended to reconstruct all the sub-images in different scales.

2.2. Feature extraction

As discussed above the decomposed WT coefficients at different levels in different sub-images represent the information of the original pixel values as well as the information of the neighboring pixels, we have used these coefficients to construct features. In this regard, different bands of images are decomposed into the desired level using the 2-D WT, which provides four subband images from each band. As a whole 16 subband images can be obtained from a four-band image (original input) after one level of decomposition. It becomes 28 band sub-images with two levels of decomposition and so on. However, the pixels of the sub-images are reconstructed to get the image information from the corresponding subband. In this operation the representation of the original image for a particular band is obtained through the reconstruction process. The sub-images are then cascaded as shown in Fig. 2 so that the extracted

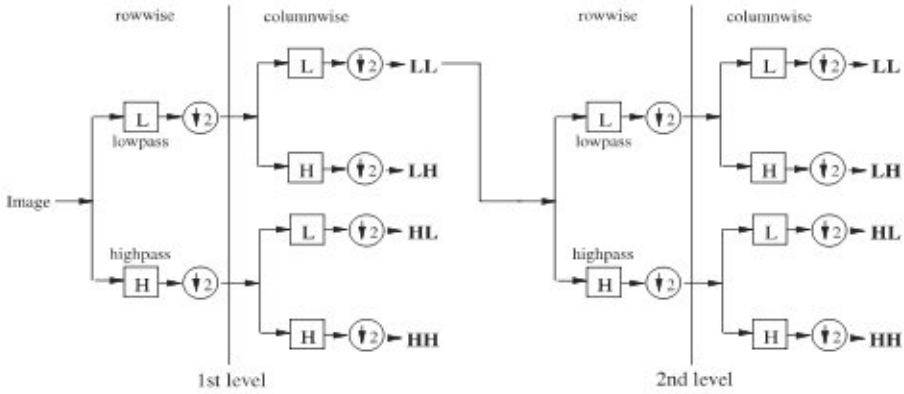

Fig. 1. A schematic representation of two-level discrete wavelet transform.

feature vectors of the original multispectral image can be obtained for the next step of classification. Cascading of different bands for generation of feature vector with Q-level of decomposition can be performed as

$$(I^1_{LL-Q}, I^1_{LH-Q}, I^1_{HL-Q}, I^1_{HH-Q}, \ldots, I^1_{LH-1}, I^1_{HL-1}, I^1_{HH-1}, \ldots, I^B_{LL-Q}, I^B_{LH-Q}, I^B_{HL-Q}, I^B_{HH-Q}, \ldots, I^B_{LH-1}, I^B_{HL-1}, I^B_{HH-1}).$$

where, I^B_{LH-1} denotes the sub-image at first level for first band with B as the number of spectral bands of the original image. Hence the feature vector of each pattern of the above decomposition will be of length $B(3Q+1)$. Thus, a two-band multispectral image with three-levels of decomposition creates a feature vector of length 20. Fig. 2 shows the way of cascading sub-images of a single band image, which can be extended to the desired level of decomposition.

3. Classification techniques

In this article since we are dealing with images, particularly remote sensing images having overlapping classes, fuzzy classification methods [4,6,53] are more useful as it allows imprecise class definition and recognize patterns belonging to more than one class [6,9] with varying degree of membership values. Thus the partitions in fuzzy classes are soft and gradual rather than hard and crisp. Popular fuzzy classifiers include fuzzy k-nearest neighbor algorithm by Keller et al. [8], fuzzy rule based algorithms by Ishibuchi et al. [54], Abe and Lan [55] and fuzzy ML classifier by Wang [7]. Fuzzy and neural networks based techniques are applied successfully to various areas including land cover classification of remote sensing images [2,56,57]. A summary of different fuzzy classifiers and their applications are described in [6,58].

The proposed wavelet-features (WF) based fuzzy classification scheme has been implemented with four different fuzzy classification methods. These are: fuzzy k-nearest neighbor (k-NN) [8], fuzzy maximum likelihood (FML) [7], fuzzy product aggregation reasoning rule (FPARR) [15] and fuzzy explicit (FE) [9]. A brief description of these classifiers are described below. The motivation behind the selection of the classifiers for present study is as follows. The FML is an extension of the maximum-likelihood (ML) classifier, it is always an advantage to evaluate the performance with FML (fuzzy equivalent of a standard ML). Similarly, it is also an advantage to consider Fk-NN (standard classifier in the fuzzy domain, equivalent k-NN). Among the other fuzzy classifiers, we found that the FE performs better. The classifier is simple and its modular architecture involves a high flexibility for easily inserting new bands or removing bands without disturbing the remaining parts of the classifier. The strength of the method is certainly its optimal extraction of the data through the explicit fuzzification, exploited efficiently by the MIN fuzzy reasoning rule. The FPARR is recently proposed fuzzy classifier and it has the ability to perform better than other three classifiers. The FPARR explores three important aspects: these are (i) extracting feature-wise information for different classes, (ii) generalization capability and (iii) combined contribution of individual features to a particular class. Therefore, keeping in mind all these advantages, we have selected the above mentioned four fuzzy clas-

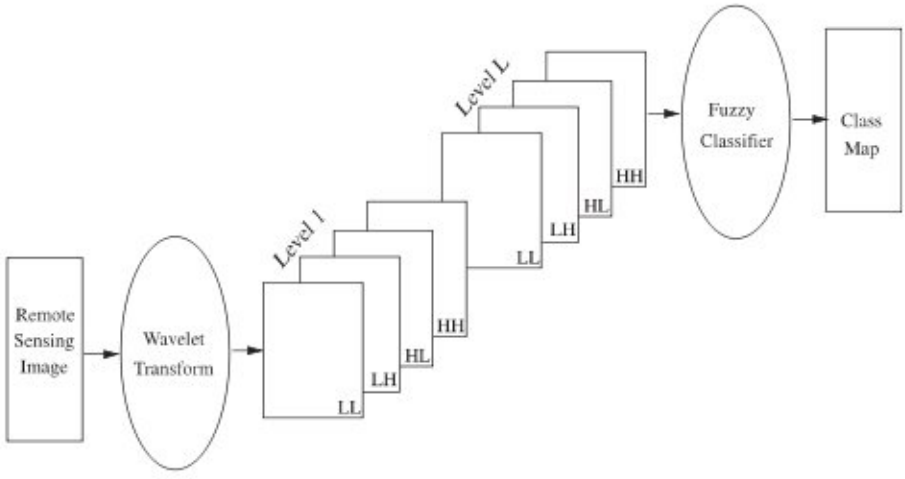

Fig. 2. Classification procedure.

sifiers. However one can extend it to some other fuzzy classifiers and may get even better results.

3.1. Fuzzy k-nearest neighbor (Fk-NN)

The k-NN classifier is a non-parametric method and based on the determination of k number of nearest neighbors of a test pattern and allocate the class map that has the majority of neighbors belonging to a particular class. Keller et al. [8] incorporated the concepts of fuzzy set theory into the k-NN voting procedure and proposed a fuzzy version of k-NN rule. The translation to fuzzy sets requires two modifications: the introduction of distance from the test pattern to the centroid of classes of all neighbors to measure the membership values, and the introduction of a membership vector for each pattern. The membership degree of a test pattern $\mathbf{x}$ to class c is calculated by

$$\mu_c(\mathbf{x}) = \frac{\sum_{j=1}^{k} \mu_{cj} \left(1/\|\mathbf{x}-\mathbf{x}_j\|^{2/(m_f-1)}\right)}{\sum_{j=1}^{k} \left(1/\|\mathbf{x}-\mathbf{x}_j\|^{2/(m_f-1)}\right)} \tag{2}$$

where $c = 1, 2, \ldots, C$, and $j = 1, 2, \ldots, k$, with C number of classes and k number of nearest neighbors. μ_{cj} is the membership degree of pattern $\mathbf{x}_j$ to class c, among the k-nearest neighbors of $\mathbf{x}$.

3.2. Fuzzy maximum likelihood (FML)

The FML [7] is a fuzzy evaluation of the conventional maximum likelihood parameters. The mean-vector and covariance matrix estimated using the fuzzy membership values for each pattern are called *fuzzy mean-vector* and *fuzzy covariance-matrix*. The fuzzy mean-vector ($\tilde{\mathbf{x}}_c$) for the c th class can be defined as

$$\tilde{\mathbf{x}}_c = \frac{\sum_{i=1}^{M_c} \mu_c(\mathbf{x}_i)\mathbf{x}_i}{\sum_{i=1}^{M_c} \mu_c(\mathbf{x}_i)} \tag{3}$$

where M_c is the total number of patterns in the c th class and $\mu_c(\mathbf{x}_i)$ is the membership value of i th pattern ($\mathbf{x}_i$) to the c th class. The fuzzy covariance-matrix is defined as

$$\Sigma_c = \frac{\sum_{i=1}^{M_c} \mu_c(\mathbf{x}_i)(\mathbf{x}_i-\tilde{\mathbf{x}}_c)(\mathbf{x}_i-\tilde{\mathbf{x}}_c)^T}{\sum_{i=1}^{M_c} \mu_c(\mathbf{x}_i)}. \tag{4}$$

The fuzzy mean-vector and fuzzy covariance-matrix can be considered as extensions of the conventional mean-vector and covariance-matrix with $\mu_c(\mathbf{x}_i) = 1$ for class c to which the pattern belongs, and 0 for other classes, Eqs. (3) and (4) then correspond to the conventional mean-vector and covariance-matrix. The fuzzy partition matrix is then evaluated using the fuzzy mean-vector and fuzzy covariance-matrix. Thus the MF for a pattern $\mathbf{x}$ to class c can be expressed as

$$\mu_c(\mathbf{x}) = \frac{p_c(\mathbf{x})}{\sum_{j=1}^{C} p_j(\mathbf{x})}, \tag{5}$$

where $p_j(\mathbf{x})$ is the probability density function. Assuming Gaussian distribution it can be computed as

$$p_j(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_j|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\tilde{\mathbf{x}}_j)^T \Sigma_j^{-1}(\mathbf{x}-\tilde{\mathbf{x}}_j)\right], \tag{6}$$

for $j = 1, 2, \ldots, C$ and $|\Sigma|$ is the determinant of the covariance-matrix Σ and D is the dimension of the pattern.

For the estimation of the MF in FML, the parameters mean-vector and covariance-matrix require the fuzzy representation of the pattern to different classes. Chen [59] described a suitable method, which estimates the fuzzy representation of the land covers in an iterative manner and does not require the prior information. We have implemented this method in the following way. Initial representation for the land covers are made randomly, by picking up values in [0, 1] representing its membership to a class. Using this representation the fuzzy mean-vector and covariance-matrices are estimated. These parameters then estimate the new fuzzy matrix following Eq. (5). This process repeats until a stable mean-vector or covariance-matrix are estimated. The stability condition assumes that there is no significant variation in the present and previous estimation of parameters in consecutive iterations, i.e. mean square error is less than ϵ (>0). In the present experiment we have used $\epsilon = 0.001$.

3.3. Fuzzy product aggregation reasoning rule (FPARR)

The FPARR classification method uses three steps [15]. In the first step, it takes the input feature vector and fuzzifies the feature values. This step uses a π-type membership function (MF) [4] to get the degree of support of a pattern into different classes based on different features. The membership value $\mu_{d,c}(x_d) = \pi_{d,c}(x_d)$, thus generated, expresses the degree of belonging of d th feature of a pattern $\mathbf{x}$ ($\mathbf{x} = [x_1, x_2, \ldots, x_d, \ldots, x_D]^T$) to c th class, with $d = 1, 2, \ldots, D$ and $c = 1, 2, \ldots, C$. The π-type MF is given by

$$\begin{aligned}
\pi(x; a, r, b) &= 0, && x \le a,\\
&= 2^{m-1}[(x-a)/(r-a)]^m, && a < x \le p,\\
&= 1 - 2^{m-1}[(r-x)/(r-a)]^m, && p < x \le r,\\
&= 1 - 2^{m-1}[(x-r)/(b-r)]^m, && r < x \le q,\\
&= 2^{m-1}[(b-x)/(b-r)]^m, && q < x < b,\\
&= 0, && x \ge b,
\end{aligned} \tag{7}$$

with m as the fuzzifier of the MF (shown in Fig. 3). In the present study we have taken the value of $m = 2.0$. The MF can be estimated with center at r. We have considered *mean* as the center (i.e., $r = mean(y)$, where y is one of the features of the training data and the two crossover points are $p = mean(y) - [max(y) - min(y)]/2$,

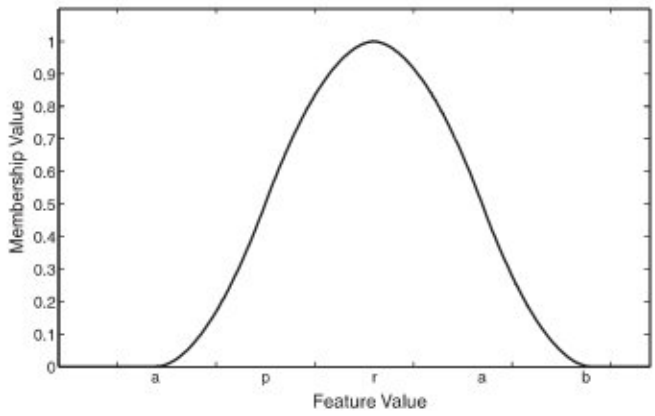

Fig. 3. π-type membership function.

and $q = mean(y) + [max(y) - min(y)]/2$. The membership value at the crossover point is 0.5 and at the center r, its value is maximum (i.e., 1). Assignment of membership value is made in such a way that all the patterns in the training data get a membership value closer to 1 for the patterns that are nearer to the center of the MF, and when they are away from the center they get a membership value closer to 0.5. Thus for a multi-featured pattern $\mathbf{x}$, the membership matrix after the fuzzification process can be expressed as

$$F(\mathbf{x}) = \begin{bmatrix} \mu_{1,1}(x_1) & \mu_{1,2}(x_1) & \dots & \mu_{1,C}(x_1) \\ \mu_{2,1}(x_2) & \mu_{2,2}(x_2) & \dots & \mu_{2,C}(x_2) \\ \dots & \dots & \dots & \dots \\ \mu_{D,1}(x_D) & \mu_{D,2}(x_D) & \dots & \mu_{D,C}(x_D) \end{bmatrix}. \quad (8)$$

In the second step, the fuzzified feature values are aggregated using PRODUCT reasoning rule (PARR). It is applied on the membership matrix to get the combined membership grade of features of a pattern to various classes. After applying the PARR, we obtain the output as a vector given by

$$F'(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_C(\mathbf{x})]^T \quad (9)$$

and

$$F_c(\mathbf{x}) = \prod_{d=1}^{D} \mu_{d,c}(x_d), \quad \forall c = 1, 2, \dots, C. \quad (10)$$

The vector $F'(\mathbf{x})$ represents the fuzzy classification expressing the class belonging, from which a hard classification output can be obtained by a defuzzification process.

The last step of the present classifier is a hard classification and is performed through a MAXIMUM (*MAX*) operation to defuzzify the output of the FPARR classification method. Here the pattern is classified to class c with the highest class membership value defined by

$$\forall_{j \in 1,2,\dots,C} \quad \text{and} \quad j \neq c, F_c(\mathbf{x}) \geq F_j(\mathbf{x}). \quad (11)$$

3.4. Fuzzy explicit (FE)

The FE classification method [9] is also processed through three steps. In the first step, it finds the membership matrix for each of the pixels/patterns with a Gaussian MF. The Gaussian MF is given as

$$\mu_{d,c}(x_d) = \exp\left[-\frac{(x_d - \bar{x}_{d,c})^2}{2\sigma_{d,c}^2}\right] \quad (12)$$

where $\bar{x}_{d,c}$ is the mean of class c with respect to feature d and $\sigma_{d,c}$ is the standard deviation of class c for feature d.

Using the above equations the fuzzy membership matrix can be evaluated as shown in Eq. (8). The membership matrix is then processed using a *MIN* RR in the second step, which selects the minimum among all the feature-wise membership values. It is calculated for all classes. A re-scaling operation is also performed for the output of the membership values (after applying RR) for all classes in this step. Then the defuzzification step is performed on the re-scaled value using the *MAX* operation. In this step a pixel is assigned a class label for which the re-scaled membership value is maximum.

4. Performance measurement indexes

In this study we have used two performance indexes to measure the quality of the classified images. After training the classifiers with the limited number available training samples, the unlabeled data has been used for testing. Hence we may call this data sets as partially labeled. Due to this limited information the performance measures like the Kappa [60] are not suitable for evaluation of the classification results. Thus two clustering indexes namely, β index [61] *based on hard labels* and *Xie-Beni* index [62] *based on soft labels* are used as the quantitative index (QI). However, to strengthen our claim for the superiority of the proposed method with the available training samples, we calculate the classification accuracy with ten-fold cross validation.

One can use other indexes like Partition Coefficient (PC), Partition Entropy (PE), Fukunyam-Sugeno (FS) index, and Fuzzy Hyper Volume (FHV) [63] for evaluation of the results depending on the problem at hand. Wu and Yang [63] analyzed various fuzzy cluster validity indexes which includes PC, PE, FS and FHV. It is understood that PC, PE and FHV does not have any connection to the geometrical structures of the data, it only measures the compactness of the clusters. On the other hand, FS measures the error within the clusters and is not a good separation measure. Contrarily XB index addresses both the objectives, i.e., within cluster compactness and separation between the clusters. Due to this advantage and wide usability we have chosen the XB index in the present study. On the other hand the β index is quite effective when number of classes are fixed and labels are hard. Computationally it is very simple. It has been used very successfully to evaluate the quality of the segmentation/classification of images [15,56,57,61,64].

4.1. β index

β is defined [61] as the ratio of the total variation and within-class variation (Eq. (13)). Since the numerator is constant for a given image, β value is dependent only on the denominator. The denominator decreases with increase in homogeneity within the class for a fixed number of classes (C). Thus for a given image and given number of classes, the higher the homogeneity within the classes, the higher would be the β value. Mathematically β can be represented as,

$$\beta = \frac{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}})^2}{\sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^2} \quad (13)$$

where $\bar{\mathbf{x}}$ is the mean (vector) grey value of the image pixels (pattern vector), M_i is the number of pixels in the ith ($i = 1, 2, \dots, C$) class, x_{ij} is the grey value of the j th pixel ($j = 1, 2, \dots, M_i$) in class i, and $\bar{\mathbf{x}}_i$ is the mean-vector of M_i grey values of the i th class.

4.2. Xie-Beni index

This measure, popularly known as the *Xie-Beni* (*XB*) was first proposed in [62]. The *XB* measure provides a validity criterion based on a validity function that identifies overall compactness and separation of fuzzy classification without any assumptions to the number of substructures inherent in the data. The index depends on the data set, geometric distance measure, distance between centroid of class and pattern, and more importantly on the fuzzy classification generated by an algorithm. It is mathematically expressed, in Eq. (14), as the ratio of compactness and separation

$$XB = \frac{1}{M}\frac{\sum_{c=1}^{C}\sum_{i=1}^{M}\mu_{ci}^2\|V_c - \mathbf{x}_i\|^2}{\min_{c \neq j}\|V_c - V_j\|^2}, \quad (14)$$

where V_c and V_j, $\forall c, j = 1, \dots, C$, are the centroid of the c th and j th *class*, respectively, and $\mathbf{x}_i$ is the i th pattern in the data set. M is the total number of data points in the data set and μ_{ci} is the

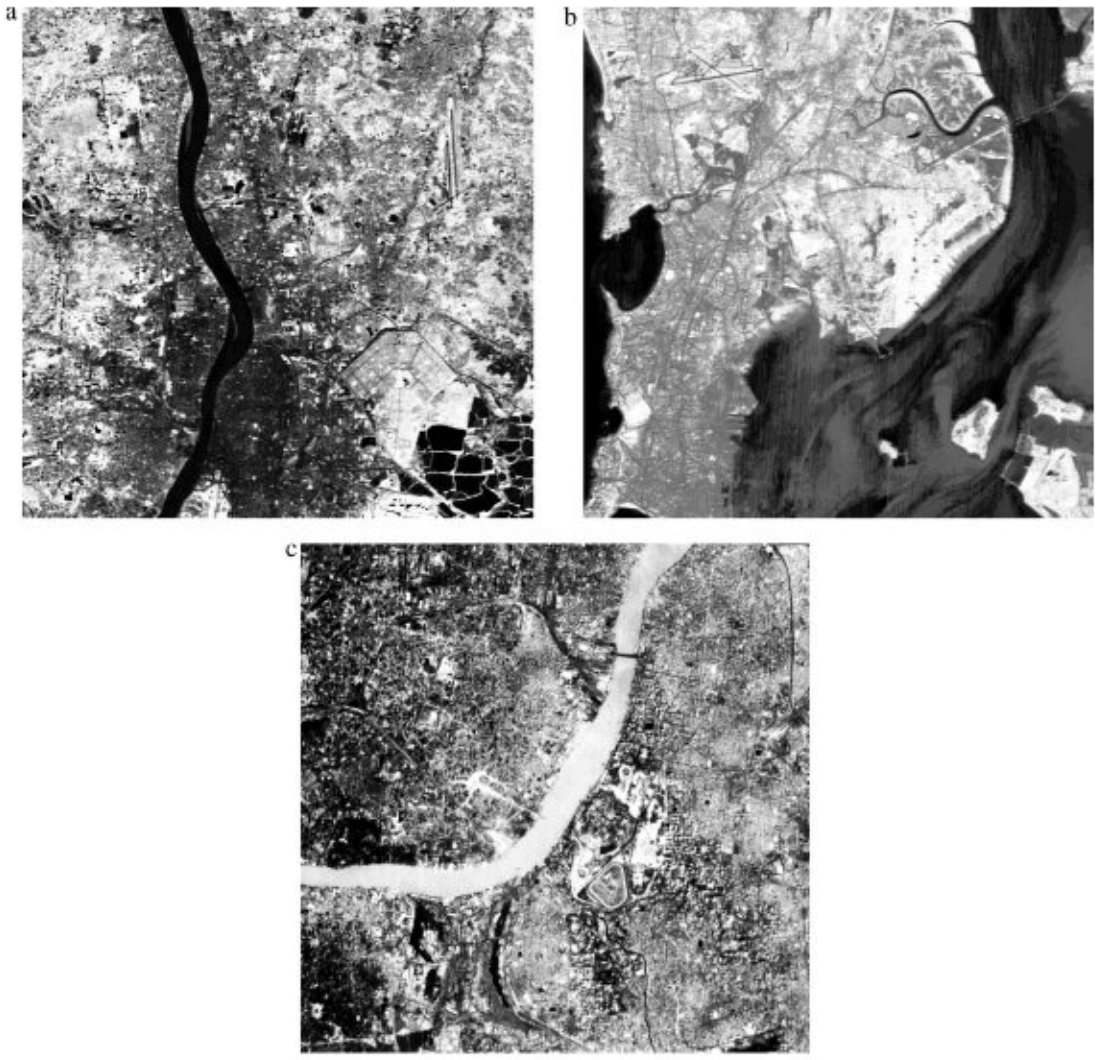

Fig. 4. (a) and (b) are IRS (band-4) Calcutta and Bombay image, respectively, and (c) is SPOT (band-3) Calcutta image.

membership value of the i th pattern to c th *class*. The *smaller* the *XB* value, the better is the classification [62].

5. Results and discussion

Selection of the training samples for all classes are made according to the ground truth available for land cover regions. These training samples are used to estimate the parameters of the classifiers. After learning the classifier, it is used to classify the land covers of the whole image.

5.1. Classification process

Four fuzzy classifiers are used in the present investigation to classify the land covers of the remote sensing images and tested using wavelets from different groups (i.e., Daubechies, Biorthogonal, Coiflets, Symlets) [47]. However, results are given for four wavelets as their performances are found (empirically) to be comparatively better than others. These are Daubechies 3 (Db3), Daubechies 6 (Db6), Biorthogonal 3.3 (Bior3.3) and Biorthogonal 3.5 (Bior3.5) wavelets [47]. In Daubechies wavelets, N (3 or 6) is related to support width, filters length, regularity and number of vanishing moments for the wavelets. For example, DbN is having support width of $2N-1$, filter length (number of filter coefficients) of $2N$, regularity about $0.2N$ for large N and number of vanishing moments N. Similarly, 3.3 and 3.5 of biorthogonal wavelets indicate some specific properties like regularities, vanishing moments, filter length, etc.

Classification accuracies of the WF based classifiers are provided in tabular form whereas the classified images with Bior3.3 wavelet and FPARR are shown in the figures also. In the present study we have used a two-level decomposition of the WT as the complexity increases proportionally with the level of decomposition with insignificant increase in the performance. The fuzzy k- NN algorithm was implemented with $m_f = 2$ and $k = 8$ (selected on the basis of performance).

5.2. Description of images

Three different remote sensing images (size 512×512) are used for the simulation study of the proposed WF based classification scheme. Among them two are from IRS and one from SPOT image. Due to poor illumination, the actual classes present in the input images are not visible clearly. So we have presented enhanced images in Fig. 4, which highlight the different land cover regions properly. However, the algorithm are implemented on actual (original) images.

5.2.1. IRS images

The IRS images (shown in Fig. 4a and b) were obtained from Indian Remote Sensing Satellite [46]. We have used the images

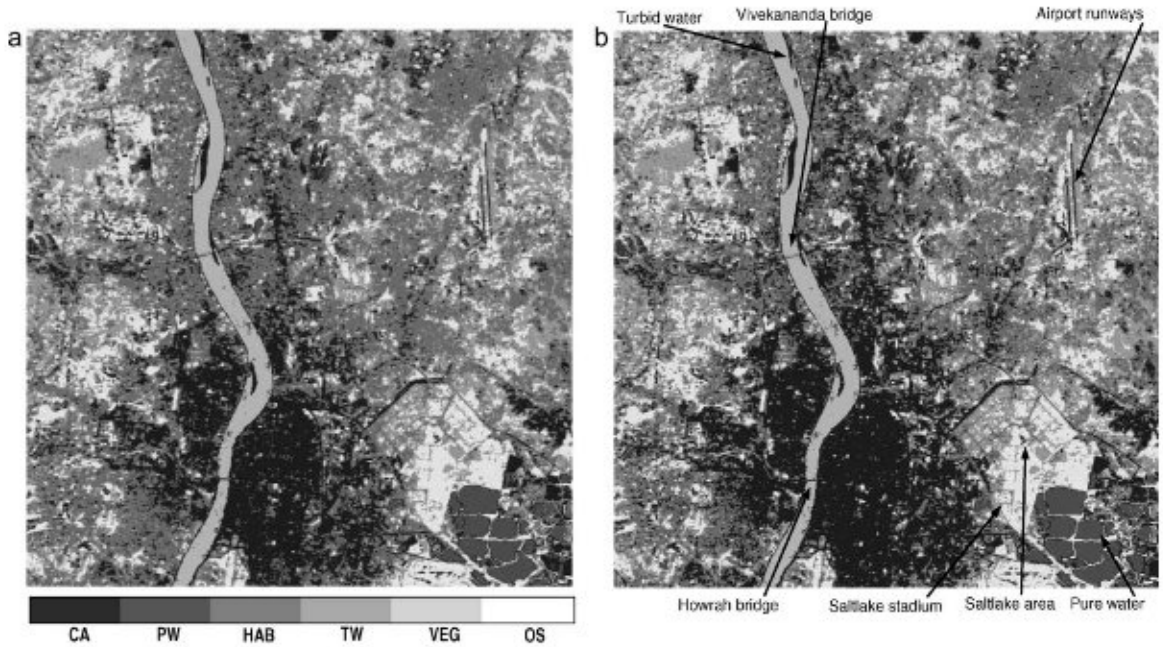

Fig. 5. Classified IRS Calcutta image by (a) FPARR with spectral features and (b) FPARR with Bior3.3 wavelet.

taken from the Linear Imaging Self Scanner (LISS-II). LISS-II has a spatial resolution of 36.25 m × 36.25 m and works in the wavelength range of 0.45–0.86 μm. The whole spectrum range is decomposed into four spectral bands namely, blue band (band-1), green band (band-2), red band (band-3) and near infrared band (band-4) with wavelengths 0.45–0.52 μm, 0.52–0.59 μm, 0.62–0.68 μm, and 0.77–0.86 μm, respectively. The image in Fig. 4a covers an area around the city of Calcutta in the near infrared band having six major land cover classes. These are *pure water* (PW), *turbid water* (TW), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS). Fig. 4b shows a part of the Bombay city in the near infrared band. The elongated city area is surrounded by the *Arabian Sea*. The total region of the Bombay image can be classified into five major classes namely, *water* (W), *concrete area* (CA), *habitation* (HAB), *vegetation* (VEG) and *open spaces* (OS).

5.2.2. SPOT image

The SPOT image (enhanced) shown in Fig. 4c is obtained from SPOT (Systeme Pour d'Observation de la Terre) satellite [1], which carries an imaging device referred to as HRV (High Resolution Visible). The Calcutta image used here has been acquired from the HRV that uses the wavelength range 0.50–0.89 μm. The whole spectrum range is decomposed into three spectral bands namely, green band (band-1), red band (band-2) and near infrared band (band-3) of wavelengths 0.50–0.59 μm, 0.61–0.68 μm, and 0.79–0.89 μm, respectively. This image has a higher spatial resolution of 20 m × 20 m. We have considered the same six classes for the land cover classification of the SPOT image as in case of IRS (Fig. 4a).

5.3. Classification of IRS Calcutta image

The three remote sensing images as shown in Fig. 4 are used for performance assessment of the proposed WF based classification scheme. These images are classified using fuzzy classifiers with original and wavelet features. The classified (IRS Calcutta) images with FPARR and WF based FPARR with Bior3.3 wavelet are shown in Fig. 5a and b, respectively, as FPARR based method performed well with original features and still better with Bior3.3 wavelet. From the visualization point of view, it is clear from the figures that the proposed WF based classifiers performed better in classifying the land covers (i.e., segregating different land covers) compared to its corresponding fuzzy version. Various objects in the IRS Calcutta image are clearly identified in the classified image. For example, as shown in Fig. 5b, we see that the *Hoogle (Ganges)* river situated in the middle of the image separating the image approximately into two halves, belong to TW class. The *pure* or *fishery water* (PW class) is easily identified in the classified image. The other classes like CA, HAB, VEG and OS are also clearly visible. Objects like *Airport runways, Saltlake Area, Saltlake Stadium, Vivekananda Bridge, Howrah Bridge* are also distinctly separable in the classified image using the proposed WF based classification scheme.

The above mentioned objects are more or less visible in case of the classified images obtained with FPARR as shown in Fig. 5a. These differences can be seen more clearly if we examine them closely. For this purpose a small portion of the lower right side of the image (Fig. 5) is zoomed and presented in Fig. 6. It is observed that the regions like *Saltlake stadium* is more clear and well-shaped in Fig. 6b compared to Fig. 6a. Similarly, water bodies came out as pure class in Fig. 6b and the separating regions of the water bodies are more distinct compared to the image in Fig. 6a. Further, a concrete distinction between various classes obtained by different classifiers are justified with the estimation of quantitative index rather than only visualizing the regions.

Two quantitative indexes named as β and *XB*, as discussed in Section 4, have been used to justify these findings. With β index, as discussed in the previous section, for a fixed number of classes, more the homogeneity within the class, more is the β value. Table 1 depicts the results of β. As expected, the β value is the highest for the training data, i.e., 9.4212 for IRS Calcutta image. Its values are 8.1717, 7.1312, 7.0523 and 7.0121 for the four fuzzy classifiers, i.e., FPARR, FE, FML and F*k*-NN, respectively. From these values it is clear that the FPARR is yielding better classification results compared to others. The β value is increased from 8.1717 to 8.6348, 8.1913, 8.7413 and 8.2012 for the WF based FPARR classifier with Db3, Db6, Bior3.3, and Bior3.5 wavelets, respectively. These increments are also there for other fuzzy classifiers with these wavelets. From Table 1, it is clear that the FPARR classification with Bior3.3 wavelet is providing the highest β value compared to others. The improvement of the classification for other fuzzy classifiers with WT as the preprocessor is also obvious from Table 1. For the rest

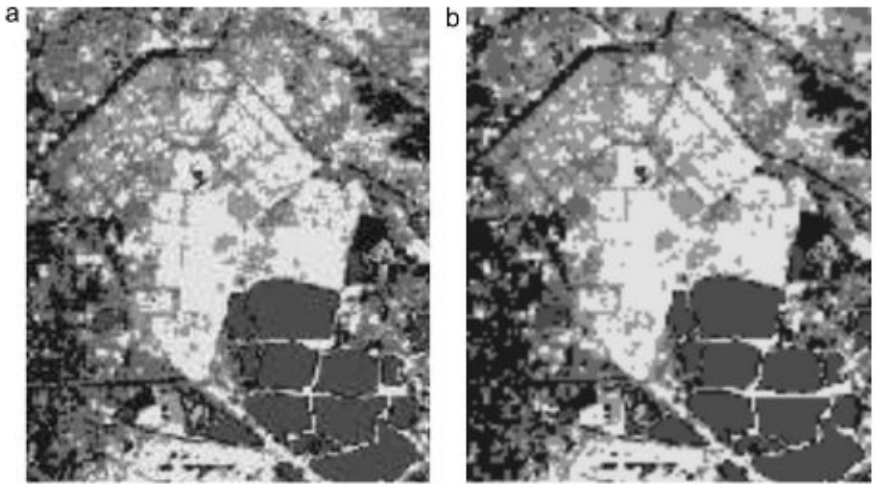

Fig. 6. Zoomed version of a selected region of Fig. 5 which show the *Saltlake stadium* and *Pure water* regions: (a) corresponds to Fig. 5a and (b) corresponds to Fig. 5b.

of the classifiers, the Bior3.3 wavelet provided better results than Db3, Db6 and Bior3.5. As a whole we can establish the following β relation in terms of the performance of the classifiers from the classification of IRS-Calcutta image, i.e.,

$$\beta_{training} > \beta_{FPARR} > \beta_{FE} > \beta_{FML} = \beta_{Fk-NN}.$$

From this relation we can get the performance quality of the different classification methods. Another relation

$$\beta_{Bior3.3} > \beta_{Db3} > \beta_{Bior3.5} = \beta_{Db6}$$

is also evident from Table 1. From these relations it is observed that the performance of the FPARR classifier is better than FE, FML and F*k*-NN. There is a similarity in the performance between FML and F*k*-NN classifiers.

Similar to this β index, the *XB* index also supported the superiority of the proposed WF based fuzzy classification scheme. The experimental values for *XB* index obtained with the three remote sensing images using all the classifiers are depicted in Table 2. It is seen that a better compaction and separation of different regions of the images are obtained with the FPARR classification method compared to FE, FML and F*k*-NN. The *XB* value for FPARR for IRS Calcutta image is 0.8310, and are 0.9012, 0.9201 and 0.9156 for FE, FML, F*k*-NN, respectively. Again, these values are still better for the Bior3.3 wavelet based features for all the classifiers. However, FPARR with Bior3.3 wavelet based features is the best among all.

Table 2
Comparison of *XB* value for different classification methods. Bold values indicate that they are the best results.

Classification method	IRS Cal	IRS Bom	SPOT Cal
FPARR	0.8310	0.7935	2.1021
FE	0.9012	0.8832	2.3031
FML	0.9201	0.9011	2.3779
F*k*-NN ($k=8$)	0.9156	0.8978	2.3415
FPARR + Bior3.3 (wavelet)	**0.7672**	**0.7366**	**1.9576**
FE + Bior3.3 (wavelet)	0.8367	0.8300	2.1166
FML + Bior3.3 (wavelet)	0.8489	0.8235	2.2210
F*k*-NN ($k=8$) + Bior3.3 (wavelet)	0.8521	0.8289	2.2001

To have another comparison of methods, we performed ten-fold cross validation for calculating the classification accuracies (with available training data) and results are shown in Table 3. All the methods are performing well with Bior3.3 wavelet, as observed from Table 1, and hence we have provided the results with this wavelet only. The results in Table 3 show the classification accuracies obtained from each fold of ten-fold cross validation, and the average over them. It is observed from Table 3 that FPARR classifier with Bior3.3 wavelet performs well compared to other classifiers.

Table 1
Comparison of β value for different classification methods. Bold values indicate that they are the best results.

Classification methods	Wavelets	IRS Cal	IRS Bom	SPOT Cal
Training patterns	–	9.4212	21.4783	9.3343
FPARR	–	**8.1717**	**19.4531**	**8.1078**
FE	–	7.1312	17.6283	7.0137
FML	–	7.0523	17.0887	6.9896
F*k*–NN (k=8)	–	7.0121	17.0013	6.9212
WF with FPARR	Db3	8.6348	20.0134	8.7315
WF with FPARR	Db6	8.1913	19.5012	8.1101
WF with FPARR	Bior3.3	**8.7413**	**20.1017**	**8.7411**
WF with FPARR	Bior3.5	8.2012	19.5102	8.1210
WF with FE	Db3	7.7017	18.3312	7.6513
WF with FE	Db6	7.1934	17.7131	7.1345
WF with FE	Bior3.3	**7.7918**	**18.4013**	**7.7022**
WF with FE	Bior3.5	7.1997	17.7343	7.2123
WF with FML	Db3	7.5786	17.7582	7.4123
WF with FML	Db6	7.1020	17.1634	7.0785
WF with FML	Bior3.3	**7.6901**	**17.7893**	**7.4078**
WF with FML	Bior3.5	7.2212	17.1456	7.3001
WF with F*k*-NN ($k=8$)	Db3	7.5412	17.6070	7.4012
WF with F*k*-NN ($k=8$)	Db6	7.1675	17.2010	7.0021
WF with F*k*-NN ($k=8$)	Bior3.3	**7.6589**	**17.6431**	**7.4902**
WF with F*k*-NN ($k=8$)	Bior3.5	7.2113	17.1341	7.1011

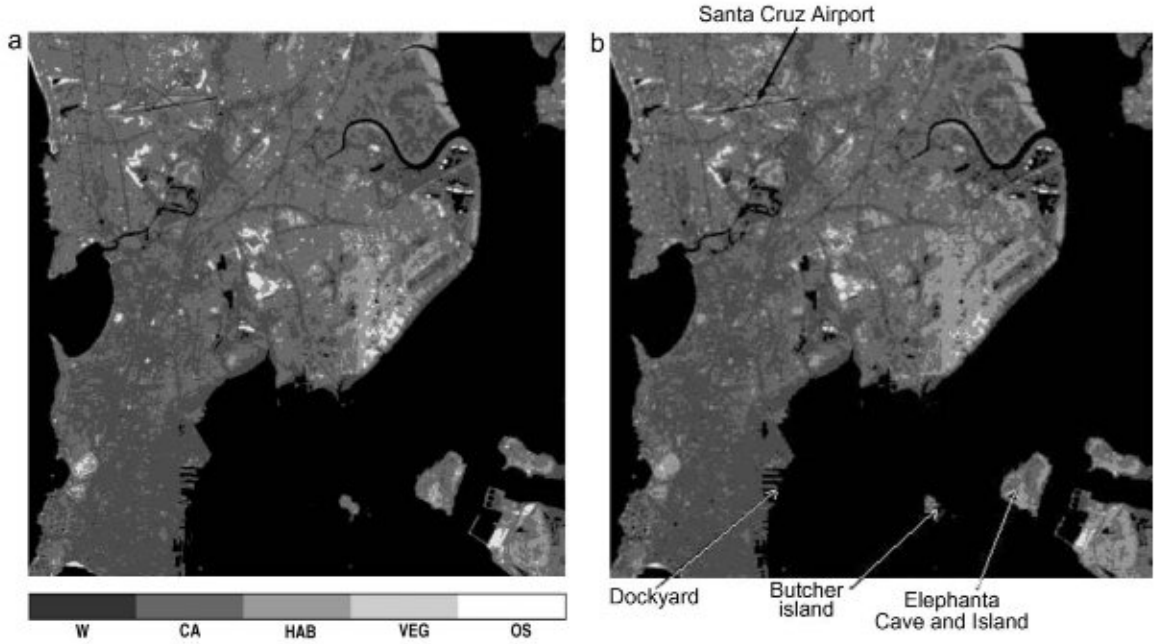

Fig. 7. Classified IRS Bombay image by (a) FPARR with spectral features and (b) FPARR with Bior3.3 wavelet.

This reveals the statistical significance of the proposed classification method (FPARR classifier with Bior3.3 wavelet). Considering all cases we can infer that the combination of the FPARR and Bior3.3 wavelet is outperforming the others.

5.4. Classification of IRS Bombay image

In case of IRS Bombay (Fig. 7b) image also the FPARR based wavelet-fuzzy method detected regions like *Dockyard*, *Butcher Island*, *Elephanta Cave and Island* and *Santa Cruz Airport* in crisp and homogeneous way compared to the FPARR method (original spectral feature only) (Fig. 7a). The classified images obtained from the FE, FML, F*k*-NN (original spectral feature only) and their corresponding WF based classifiers are not shown in figures, as their performances are not good compared to FPARR and its WF based method. However, their quantitative indexes are shown in the Tables 1 and 2. As mentioned, the two classified images in Fig. 7a and b show some visual differences. A proper visualization can be made with a zoomed version of some portion as shown in Fig. 8. From this figure, it is seen that the *Dockyard* region is more clearly and distinctly (well shaped) classified with the proposed scheme. The sharp edges and regions are properly visible from Fig. 8b compared to the region shown in Fig. 8a. Also just above the *Dockyard* region a water body came out properly in Fig. 8b, which is not that prominent in Fig. 8a. However, a better comparison can be made from the quantitative indexes like β and *XB*. Table 1 depicts the β values. β for the Bombay image with training data is found to be 21.4783, which are 19.4531, 17.6283, 17.0887 and 17.0013 for the FPARR, FE, FML and F*k*-NN methods, respectively. As in the classification with FPARR, the WF based other fuzzy classifiers provided better classification results compared to the original spectral feature based ones, and their improvement are still better with the Bior3.3 wavelet. Table 1 justifies this arguments. This finding is true for all other classifiers and for both the original and WF based methods. The same relations on β values are maintained for IRS Bombay image as in the case of Calcutta image.

Similarly, the *XB* index shown in Table 2 also indicates the superiority of the proposed scheme. The Bior3.3 wavelet based FPARR provided promising classification results over others. The *XB* value for this method is 0.7366, which is the lowest among all, and justifies the efficiency in classifying the land covers present in the IRS Bombay image.

5.5. Classification of SPOT Calcutta image

For SPOT Calcutta image, the classified regions of the images are shown in Fig. 9a for FPARR (original spectral feature only) and

Table 3
Classification accuracy (with ten-fold cross validation) for different classification methods for IRS-Calcutta image.

Classification method			
FPARR + Bior3.3 wavelet	FE + Bior3.3 wavelet	FML + Bior3.3 wavelet	F*k*-NN + Bior3.3 wavelet
81.43	74.06	76.82	75.06
77.37	73.03	69.37	66.15
79.62	80.61	68.89	72.68
77.56	79.28	75.02	75.11
78.22	77.72	74.18	75.14
81.23	73.45	70.28	72.23
80.03	79.32	73.04	67.07
82.51	78.03	67.08	73.01
76.84	75.24	74.62	69.66
76.78	80.01	67.56	70.51
Mean = **79.15**	Mean = **77.07**	Mean = **71.76**	Mean = **71.66**

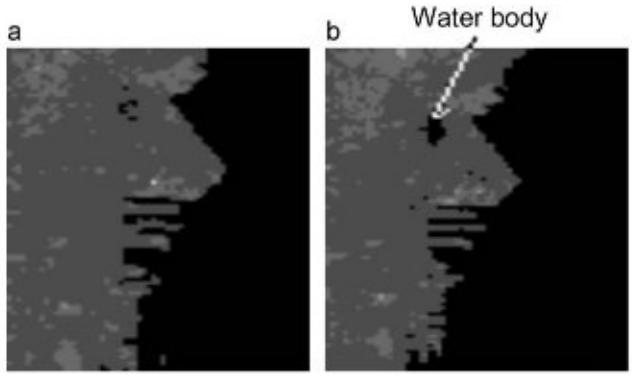

Fig. 8. Zoomed version of a selected region of Fig. 7 which show the *Dockyard* and *Water* regions: (a) corresponds to Fig. 7a and (b) corresponds to Fig. 7b.

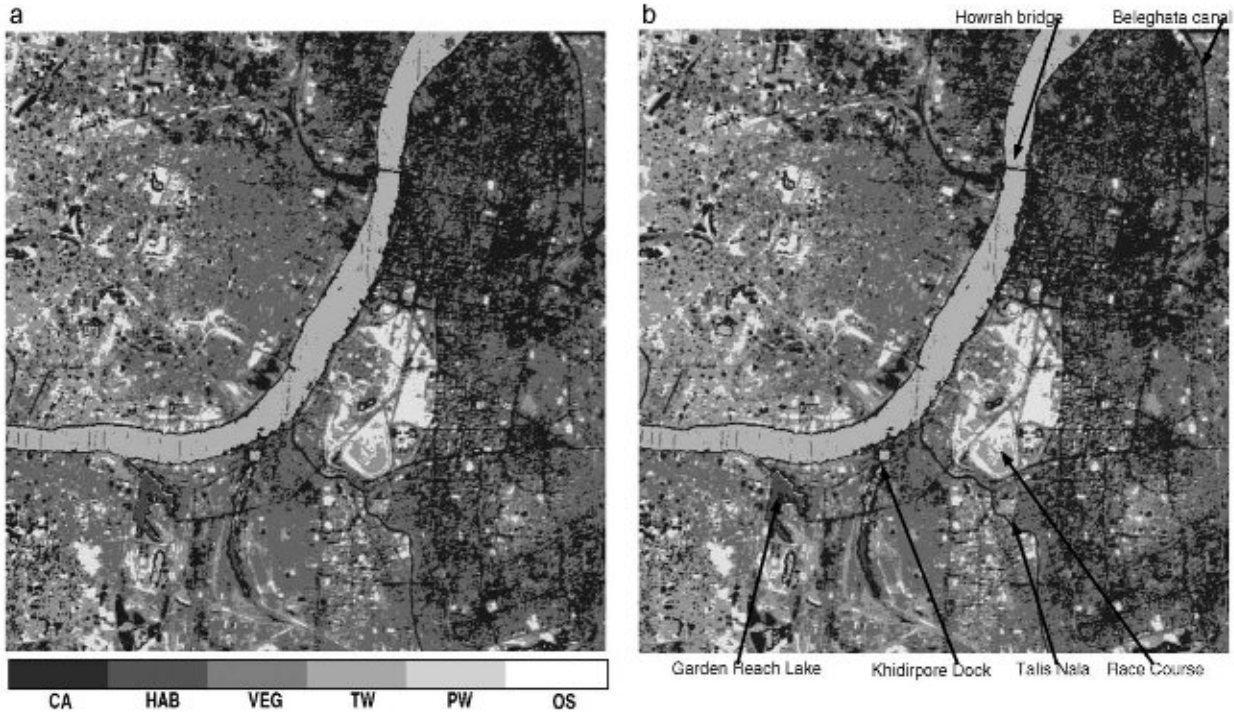

Fig. 9. Classified SPOT Calcutta image by (a) FPARR with spectral features and (b) FPARR with Bior3.3 wavelet.

Fig. 9b for the FPARR based wavelet classifiers with Bior3.3 wavelet only as these were better compared to others. From the figures it is observed that there is a clear separation of different classes and some known regions like *Race Course*, *Howrah Bridge (Setu)*, *Talis Nala(Canal)*, *Beleghata Canal*, *Khiderpore Dock* and *Garden Reach Lake* by the proposed scheme. For a proper visualization a zoomed image of some regions are shown in Fig. 10. It is seen that some of the structures (particularly curvilinear) are more prominent in Fig. 10b compared to Fig. 10a. For example, in Fig. 10b, the road structures (concrete area class) in the lower left-side of the image is properly connected whereas in Fig. 10a it is broken in few places. Water body (as shown in Fig. 10b) came out as pure water class and the boundary is also well connected. But in Fig. 10a the pure water class is mixed up with other classes and the boundary is broken in some places. From these comparisons it is evident that the WF based classification method produced well structured and proper shaped regions compared to original feature based method. However, a better performance comparison with the help of β value can

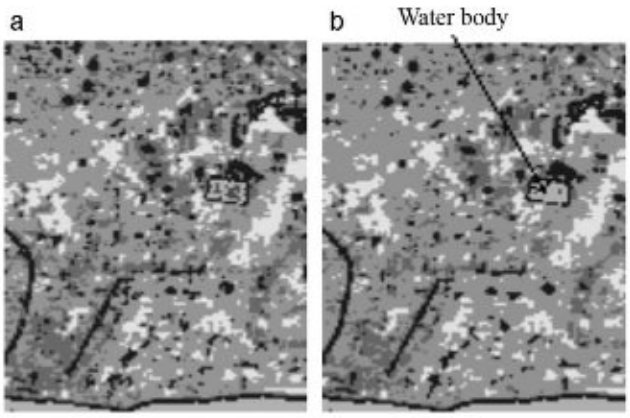

Fig. 10. Zoomed version of a selected region of Fig. 9 (showing a water region towards the right side of the central position): (a) corresponds to Fig. 9a and (b) corresponds to Fig. 9b.

be seen from Table 1. The β value for the training data set is 9.3343. Its values are 8.1078, 7.0137, 6.9896 and 6.9212 for the classified images using the FPARR, FE, FML and Fk-NN classifiers, respectively. In this case also the improved performance of the wavelet-fuzzy classifiers over their fuzzy versions are observed and is still better with FPARR based method. It is seen that β is the highest for the FPARR based wavelet fuzzy classification method using Bior3.3 wavelet and revealing its supremacy. The β values are depicted for all the classifiers in Table 1. The β relations, in the classification of SPOT Calcutta image, are also observed to be similar to the case of IRS Calcutta and IRS Bombay images. Like the β index, the *XB* values also corroborate the earlier findings.

From the land cover classification of three remote sensing images, we found that the proposed WF based fuzzy classification methods are superior compared to their corresponding original spectral feature based ones. Performance comparison among four fuzzy classifiers with their WF versions have been made using two quantitative indexes. The quantitative indexes supported the superiority of the proposed scheme. It is also seen that the Bior3.3 wavelet is outperforming the other wavelets. Visual inspections shows that the classified regions using the proposed scheme (Figs. 5b, 7b and 9b) are more crisp, homogeneous and compact compared to the corresponding original method (Figs. 5a, 7a and 9a).

6. Conclusion

We have proposed a wavelet feature based fuzzy (a soft computing approach) classification of multispectral remote sensing images. The proposed scheme tries to explore the possible advantages of using WT as a preprocessor for fuzzy classifiers in a hybrid framework. The WT is used to extract features from the original patterns. The extracted features acquire information of the pixel along with its neighbors both in spatial and spectral domains because of the inherent characteristic of the WT which makes the next stage of classification more efficient compared to that without preprocessing. We have used a two-level decomposition of WT, as

complexity increases proportionally with the level of decomposition with insignificant increase in performance.

The improvement in performance of the proposed hybrid framework (wavelet + fuzzy) is verified from the results obtained from classification of three remote sensing images. The β values (shown in Table 1) indicate the classification accuracy and support the visual separation in the identification of various known objects. The *Xie-Beni* (*XB*) index (Table 2) also supported the above findings. Further, different wavelets are being used in the preprocessing stage of the proposed scheme. Among all the methods, the FPARR classification method was seen to be superior compared to others and biorthogonal3.3 (Bior3.3) outperformed all other wavelets. Also it is observed that the different structures of the classified regions obtained with the proposed scheme are more crisp and well shaped.

In future we plan to use the concept of wavelet based feature extraction for non-fuzzy classifiers also, including support vector machine. In a companion study we are comparing the performance of wavelet based features with conventional statistical features.

Acknowledgements

The authors would like to thank the reviewers for their valuable and constructive suggestions. The authors like to thank the Department of Science and Technology, Government of India and the Ministry of Foreign Affair, Government of Italy, and the Autonomous Province of Trento, Italy, the sponsors of the ITPAR program, under which a project titled *Advanced Techniques for Remote Sensing Image Processing* is being carried out at the Machine Intelligence Unit, Indian Statistical Institute, Kolkata.

References

[1] J.A. Richards, X. Jia, Remote Sensing Digital Image Analysis: An Introduction, 4th ed., Springer, Germany, 2006.
[2] B. Tso, P.M. Mather, Classification Methods for Remotely Sensed Data, 2nd ed., CRC Press, Boca Raton, Florida, 2009.
[3] D. Lu, Q. Weng, A survey of image classification methods and techniques for improving classification performance, International Journal of Remote Sensing 28 (5) (2007) 823–870.
[4] L.A. Zadeh, Fuzzy sets, Information Control 8 (1965) 338–353.
[5] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Application, Prentice Hall, 1997.
[6] L.I. Kuncheva, Fuzzy Classifier Design, Springer-Verlag, 2000.
[7] F. Wang, Fuzzy supervised classification of remote sensing images, IEEE Transactions on Geoscience and Remote Sensing 28 (2) (1990) 194–201.
[8] J.M. Keller, M. Gray, J. Givens, A fuzzy *k*-nearest neighbor algorithm, IEEE Transactions on Systems, Man and Cybernetics 15 (4) (1985) 580–585.
[9] F. Melgani, B.A.R. Al Hashemy, S.M.R. Taha, An explicit fuzzy supervised classification method for multispectral remote sensing images, IEEE Transactions on Geoscience and Remote Sensing 38 (1) (2000) 287–295.
[10] A. Bardossy, L. Samaniego, Fuzzy rule-based classification of remotely sensed imagery, IEEE Transactions on Geoscience and Remote Sensing 40 (2) (2002) 362–374.
[11] Y. Wang, M. Jamshidi, Fuzzy logic applied in remote sensing image classification, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 6378–6382.
[12] J. Tang, L. Wang, S.W. Myint, Improving urban classification through fuzzy supervised classification and spectral mixture analysis, International Journal of Remote Sensing 28 (18) (2007) 4047–4063.
[13] A.K. Shackelford, C.H. Davis, A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas, IEEE Transactions on Geoscience and Remote Sensing 41 (9) (2003) 1920–1932.
[14] D. Dubois, H. Prade, On the use of aggregation operations in information fusion processes, Fuzzy Sets and Systems 142 (1) (2004) 143–161.
[15] A. Ghosh, S.K. Meher, B. Uma Shankar, A novel fuzzy classifier based on product aggregation operator, Pattern Recognition 41 (3) (2008) 961–971.
[16] R.A. Schowengerdt, Techniques for Image Processing and Classification in Remote Sensing, Academic Press, 1983.
[17] R.M. Haralick, K.S. Shanmugam, Combined spectral and spatial processing of ERTS imagery data, Remote Sensing of Environment 3 (1) (1974) 3–13.
[18] S. Tuominen, A. Pekkarinen, Performance of different spectral and textural aerial photograph features in multi-source forest inventory, Remote Sensing of Environment 94 (2) (2005) 256–268.
[19] M. Chica-Olmo, F. Abarca-Hernandez, Computing geostatistical image texture for remotely sensed data classification, Computers and Geosciences 26 (4) (2000) 373–383.
[20] F.S. Cohen, Z. Fan, M.A. Patel, Classification of rotation and scaled textured images using Gaussian Markov random field models, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (2) (1991) 192–202.
[21] H. Derin, H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, IEEE Transactions on Pattern Analysis and Machine Intelligence (1987) 39–59.
[22] M. Unser, Local linear transforms for texture measurements, Signal Processing 11 (1986) 61–79.
[23] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed., Academic Press, 1999.
[24] S. Zhang, X. Xue, X. Zhang, Feature extraction and classification with wavelet transform and support vector machines, in: Proceedings IEEE International Geoscience and Remote Sensing Symposium, IGARSS '05, vol. 6, 2005, pp. 3795–3798.
[25] M. Unser, Texture classification and segmentation using wavelet frames, IEEE Transactions on Image Processing 4 (11) (1995) 1549–1560.
[26] G.V. de Wouwer, P. Schenders, D.V. Dyek, Statistical texture characterization from discrete wavelet representation, IEEE Transactions on Image Processing 8 (4) (1999) 592–598.
[27] S. Arivazhagan, L. Ganesan, Texture classification using wavelet transform, Pattern Recognition Letters 24 (2003) 1513–1521.
[28] H.H. Szu, J.L. Moigne, N.S. Netanyahu, C.C. Hsu, Integration of local texture information in the automatic classification of landsat images, SPIE Procceding 3078 (1997) 116–127.
[29] J. Yu, M. Ekstrom, Multispectral image classification using wavelets: a simulation study, Pattern Recognition 36 (4) (2003) 889–898.
[30] V. Strela, P. Heller, G. Strang, P. Topiwala, C. Heil, The application of multiwavelet filter banks to image processing, IEEE Transactions on Image Processing 8 (4) (1999) 548–563.
[31] H. Soltanian-Zadeh, F. Rafiee-Rad, S.D. Pourabdollah-Nejad, Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms, Pattern Recognition 37 (10) (2004) 1973–1986.
[32] X. Huang, L. Zhang, P. Li, A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform, International Journal of Remote Sensing 29 (20) (2008) 5923–5941.
[33] W. Wong, C. Yuen, D. Fan, L. Chan, E. Fung, Stitching defect detection and classification using wavelet transform and BP neural network, Expert Systems with Applications 36 (2) (2009) 3845–3856 (Part 2).
[34] J.A.M. Zegarra, N.J. Leite, R. da Silva Torres, Wavelet-based fingerprint image retrieval, Journal of Computational and Applied Mathematics 227 (2) (2009) 294–307.
[35] E.-S.A. Dahshan, T. Hosny, A.-B.M. Salem, Hybrid intelligent techniques for MRI brain images classification, Digital Signal Processing 20 (2) (2010) 433–441.
[36] J. Daugman, Iris encoding and recognition using Gabor wavelets, in: Encyclopedia of Biometrics, 2009, pp. 787–797.
[37] K.W. Bowyer, K. Hollingsworth, P.J. Flynn, Image understanding for iris biometrics: a survey, Computer Vision and Image Understanding 110 (2) (2008) 281–307.
[38] A. Abhyankar, S. Schuckers, A novel biorthogonal wavelet network system for off-angle iris recognition, Pattern Recognition 43 (3) (2010) 987–1007.
[39] J.A.M. Zegarra, N.J. Leite, R. da Silva Torres, Wavelet-based fingerprint image retrieval, Journal of Computational and Applied Mathematics 227 (2) (2009) 294–307.
[40] M. Grzegorzek, S. Sav, E. Izquierdo, N. O'Connor, Local wavelet features for statistical object classification and localisation, IEEE Multimedia 17 (1) (2010) 56–66.
[41] E.A. Rashed, I.A. Ismail, S.I. Zaki, Multiresolution mammogram analysis in multilevel decomposition, Pattern Recognition Letters 28 (2) (2007) 286–292.
[42] F. Murtagh, J.-L. Starck, Wavelet and curvelet moments for image classification: application to aggregate mixture grading, Pattern Recognition Letters 29 (10) (2008) 1557–1564.
[43] R.A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, Academic Press, 2007.
[44] P.-H. Hsu, Y.-H. Tseng, P. Gong, Spectral feature extraction for hyperspectral images using wavelet transform, Journal of Photogrammetry and Remote Sensing 11 (1) (2006) 93–109.
[45] Y. Zhang, S.D. Backer, P. Scheunders, Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 47 (11) (2009) 3834–3843.
[46] NRSA IRS data users hand book, tech. rep., NRSA, (1989). Document No. IRS/NRSA/NDC/HB-02/89.
[47] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
[48] J. Mazzaferri, S. Ledesma, C. Iemmi, Multiple feature extraction by using simultaneous wavelet transforms, Journal of Optics, A: Pure and Applied Physics 5 (2003) 425–431.
[49] S. Soltani, P. Simard, D. Boichu, Estimation of the self-similarity parameter using the wavelet transform, Signal Processing 84 (1) (2004) 117–123.
[50] P.K. Topiwala, Wavelet Image and Video Compression, Springer, 2000.
[51] M. Vetterli, J. Kovacevic, Wavelets and Subband Coding, Prentice Hall, 1995.
[52] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 2nd ed., Prentice Hall, 2002.
[53] S.K. Pal, D.D. Majumder, Fuzzy Mathematical Approach to Pattern Recognition, Wiley Eastern Limited, 1986.
[54] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and Systems 52 (1) (1992) 21–32.

[55] S. Abe, M.S. Lan, A method for fuzzy rules extraction directly from numerical data and its application to pattern classification, IEEE Transactions on Fuzzy Systems 3 (1) (1995) 18–28.
[56] S.K. Meher, B. Uma Shankar, A. Ghosh, Wavelet-feature-based classifiers for multispectral remote-sensing images, IEEE Transactions on Geosciences and Remote Sensing 45 (6) (2007) 1881–1886.
[57] A. Ghosh, B. Uma Shankar, S.K. Meher, A novel approach to neuro-fuzzy classification, Neural Networks 22 (2009) 100–109.
[58] W. Pedrycz, Fuzzy sets in pattern recognition: Methodology and methods, Pattern Recognition 23 (1/2) (1990) 121–146.
[59] C.F. Chen, Fuzzy training data for fuzzy supervised classification of remotely sensed images, in: Asian Conference on Remote Sensing (ACRS 1999), 1999.
[60] J. Cohen, A coefficient of agreement for nominal scale, Education and psychological measurement 20 (1) (1960) 37–46.
[61] S.K. Pal, A. Ghosh, B. Uma Shankar, Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation, International Journal of Remote Sensing 21 (11) (2000) 2269–2300.
[62] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 841–847.
[63] K.-L. Wu, M.-S. Yang, A cluster validity index for fuzzy clustering, Pattern Recognition Letters 26 (9) (2005) 1275–1291.
[64] P. Mitra, B. Uma Shankar, S.K. Pal, Segmentation of multispectral remote sensing images using active support vector machines, Pattern Recognition Letters 25 (12) (2004) 1067–1074.