

Wavelet-based Texture Segmentation of Remotely Sensed Images

Mausumi Acharyya and Malay K. Kundu
Machine Intelligence Unit
Indian Statistical Institute
203 B.T. Road, Calcutta - 700 035, India
email: {res9522,malay}@isical.ac.in

Abstract

In this article a texture feature extraction scheme based on M -band wavelet packet frames is investigated. The features so extracted are used for segmentation of satellite images which usually have complex and overlapping boundaries. The underlying principle is based on the fact that different image regions exhibit different textures. Since most significant information of a texture often lies in the intermediate frequency bands, the present work employs an over-complete wavelet decomposition scheme called discrete M -band wavelet packet frame (DM-bWPF), which yields improved segmentation accuracies. Wavelet packets represent a generalization of the method of multiresolution decomposition and comprise of all possible combinations of sub-band tree decomposition. We propose a computationally efficient search procedure to find the optimal basis based on some maximum criterion of textural measures derived from the statistical parameters of each of the subbands, to locate dominant information in each subbands (frequency channels) and decide further decomposition.

1. Introduction

The segmentation of different landcover regions has been recognized as a difficult problem in the realm of satellite imagery. Remotely sensed images usually have poor illumination and are highly dependent on the environmental conditions. Spatial resolution of these images are also very low. The scene mostly contains many objects (landcovers), 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 landcovers present in the corresponding pixel area. Assigning unique class levels with certainty is thus a problem of remotely sensed images. Also these type of 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 resolution of the imagery may be different in many cases, and so it is important to understand how information changes through different scales of imagery. This problem leads naturally to multiresolution type analysis which are most effective using wavelets, also wavelet theory is well suited for the study of complex signals which are aperiodic. Wavelets are particularly good at describing a scene in terms of the scale of the textures in it.

Texture is an important property of reflective surface which human visual perception system uses to segment and classify different image objects in a digital image. In a remotely sensed image 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 remotely sensed images provide valuable information for segmentation of such images.

Segmentation is a process of partitioning an image space into some non-overlapping meaningful homogeneous regions. The term meaningful is ofcourse problem dependent and the success of an image analysis system depends on the quality of segmentation. So basically this is a multi-texture segmentation problem.

Several approaches have been considered in the last few decades and reported in [9]. Of the several approaches available for texture feature extraction we focus on the signal processing approach in the present work. Other approaches to segmentation of remotely sensed images have been reported in the literature. Various fuzzy thresholding techniques is demonstrated in remotely sensed images in [6] and genetic algorithm based pattern classifiers has been investigated in the domain of satellite imagery in [2].

Most of the texture segmentation algorithms based on signal processing techniques [4] apply the textured image to a filtering step followed by a nonlinear operation which gives an estimate of the energy. Recent development of wavelet theory has provided a promising tool for texture analysis.

The octave band (standard wavelet) decomposition gives a logarithmic frequency resolution and are not suitable

for the analysis of high frequency signals with relatively narrow bandwidth. So the main motivation of this work is to utilize the decomposition scheme based on M -band wavelets, which unlike the standard wavelet decomposition gives a mixture of logarithmic and linear frequency resolution. Since the most significant information of a texture often appears in the middle frequency channels, and also translational invariance is desirable for accomplishing texture analysis, an M -band wavelet packet frame transform is envisaged, which corresponds to a general tree-structured filter bank, and gives an overcomplete representation. But this decomposition scheme leads to a large number of independent basis. We propose a computationally efficient and adaptive technique for finding out the optimal basis based on some maximum criterion of textural measures derived from the statistical parameters extracted from each of the subbands.

Section 2 presents the analysis of the filtering technique used in the proposed work, while section 3 discusses about the extraction of features and integration of these features. Finally section 4 gives experimental results and concludes our study.

2. Multiscale M -band wavelet representation

The feature extraction scheme that we have used has a filtering stage and a subsequent nonlinear stage followed by a smoothing filter as shown in figure 1. The objectives of the filtering and that of the local energy estimator, are to transform the edges between textures into detectable discontinuities.

2.1. M -band wavelet filter

The *wavelet transform* maps a function $f(x) \in L^2(\mathbb{R})$ onto a scale-space plane, and are obtained from a single prototype function $\psi(x)$ by scalings a and shifts b [3] [5]. The M -band orthonormal wavelet is a direct generalization of the two band case [7], and are able to zoom in onto narrowband high frequency components of a signal and gives better energy compaction than 2-band wavelets [7].

An M -band wavelet system consist of $M - 1$ wavelets, $\psi_i(x)$, $i = 2, \dots, M$ associated with the scaling function $\psi_1(x)$. The scaling function $\psi_1(x)$ satisfies,

$$\psi_1(x) = \sum_k h_1(k) \sqrt{M} \psi_1(Mx - k) \quad (1)$$

where the sequence h_1 is the scaling filter, and satisfies the following linear and quadratic constraints.

$$\sum_{k=0}^{k=N-1} h_1(k) = \sqrt{M}, \quad \sum_{k=0}^{k=N-1} h_1(k) h_1(k + Ml) = \delta l \quad (2)$$

And there are $M - 1$ wavelets which also satisfy,

$$\psi_i(x) = \sum_k \sqrt{M} h_i(k) \psi(Mx - k) \quad (3)$$

The $(M - 1)h_i$ ($i = 2, \dots, M$) vectors called the wavelet filters satisfy the equation,

$$\sum_{k=0}^{k=N-1} h_i(k) h_j(k + Ml) = \delta(l) \delta(i - j) \quad (4)$$

The filter bank in essence is a set of bandpass filters with frequency and orientation selective properties. In the filtering stage we make use of orthogonal and linear phase M -band ($M=4$) wavelet following [1]. The 1-D $M(=4)$ -band wavelet filter responses are given by ψ_i and their corresponding transfer functions are denoted by H_i , where $i = 1, \dots, M$.

The filters $H_{ik}(\omega)$, at level k are generated as follows,

$$H_{i,k} = H_{i,0}(2^k \omega) \quad \text{for } i = 1, \dots, 4 \quad (5)$$

Suppose $\hat{I}_s^k(\omega)$ be the Fourier transform of the input signal $I(x)$ for subband (frequency channel) s at decomposition level k . For $0 \leq s \leq 4^k - 1$ we have,

$$\hat{I}_{i,4s+(i-1)}^{k+1}(\omega) = H_{i,k}(\omega) \hat{I}_{i,s}^k(\omega) \quad (6)$$

This corresponds to a filter bank with channel filters,

$$filt_{i,s}^k(\omega) \quad |i = 1, \dots, 4$$

from the filter bank theoretic point. The filters are given by the following recursive relation as follows,

$$\begin{aligned} filt_{i,0}^0(\omega) &= H_i(\omega), \\ filt_{i,4s+(i-1)}^{k+1}(\omega) &= H_{i,k+1}(\omega) filt_{i,s}^k(\omega) \\ &= H_{1,0}(2^{k+1} \omega) filt_{i,s}^k(\omega) \end{aligned} \quad (7)$$

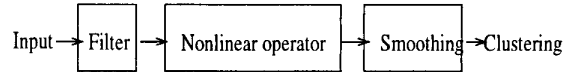


Figure 1. Experimental setup.

For 2-D signals, the transform is obtained by the tensor product of the 1-D channel filters.

2.2. Adaptive basis selection

An appropriate way to perform the wavelet transform for texture feature extraction is to detect the most significant frequency channels and then decompose them further.

This leads naturally to a tree structured wavelet transform. It is also usually redundant to decompose all the subbands in each scale to achieve the full tree of decomposition. Also for a decomposition depth of k , M^{M^k} number of basis are possible. It is quite evident that an exhaustive search to determine the optimal basis from this large set is computationally expensive and difficult procedure.

In order to avoid a full decomposition and simultaneously to find out the optimal basis, we propose an adaptive decomposition algorithm using a maximum criterion of textural measures based on the statistics extracted from each of the subbands, and identify the most significant subbands and then decide whether further decomposition of the particular channel would generate more information or not. This search is computationally efficient and enables us to zoom into any desired frequency channel for further decomposition. Energy is used as the textural measure in our work. Energy measures textural uniformity, meaning pixel pair repetitions.

Figure 2 shows a general tree structure of discrete M -band wavelet packet frame decomposition. The image is first decomposed into $M \times M$ channels using the 2-D M -band wavelet transform, but without downsampling. Energy of each subband is then evaluated. From the M^2 subbands only those that contain appreciable energy are considered and decomposed further. In our simulation we have considered only those bands that contain more than 2% of the total energy of the parent band. And we have further decomposed a subband if it atleast have more than 10% of the total energy of all the subbands at the current scale. This step results in a set of feature images $Feat_k(x, y)$, from which a set of feature vectors are derived.

3. Texture feature extraction

3.1. Local energy estimator

The objectives of the filtering and that of the local energy estimator (nonlinear operator and smoothing filter), are to transform the edges between textures into detectable discontinuities. The local energy estimator, estimates the energy of the filter output in a local region around each pixel. Since the magnitude of the correlation between the wavelet and the image is all that is important, we have used *absolute values* of the wavelet coefficients as a generalized energy definition. The local energy $Eng_k(i, j)$ is formally given as,

$$Eng_k(i, j) = |F_k(i, j)| \quad (8)$$

where, $F_k(i, j)$ is the filtered image at different scales k . The nonlinear transform is succeeded by a Gaussian low pass (smoothing) filter $h_G(x, y)$. Formally, the feature image $Feat_k(x, y)$ corresponding to filtered image $F_k(x, y)$ is

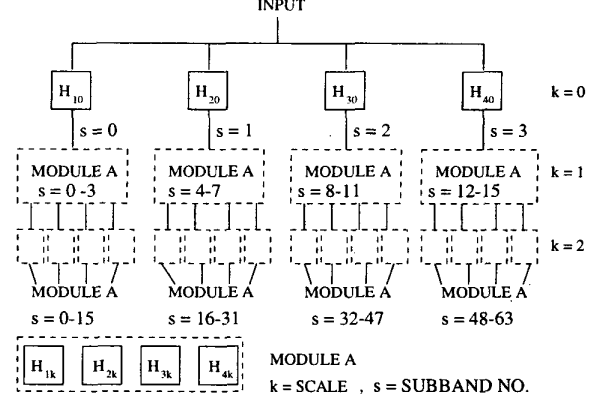


Figure 2. Structure of DM-bWPF transform.

given by,

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

where k correspond to different scales, $\Gamma(\cdot)$ is the nonlinear function and G_{xy} is a $G \times G$ window centered at pixel with coordinates (x, y) . The size G of the smoothing or the averaging window is an important parameter. More reliable measurement of texture feature demands larger window sizes. On the other hand, more accurate localization of region boundaries requires smaller windows.

3.2. Unsupervised classifier

Having obtained the feature images, the main task is to integrate these feature images to produce a segmentation. We define a scale - space signature as the vector of features at different scales taken at a single pixel in an image,

$$\bar{F}(i, j) = [F_0(i, j), F_1(i, j), \dots, F_N(i, j)] \quad (9)$$

Suppose these scale-space signatures are considered as feature vectors in a feature space. If the signatures of the text part are distinct from the signatures of non-text part a pattern recognition system can be used to identify those feature vectors that represent text part in the scale space.

Let us assume that there are M texture categories, C_1, \dots, C_M , present in the image. If the texture features obtained are capable of discriminating these categories then the patterns belonging to each category will form a cluster in the feature space which is compact and isolated from clusters corresponding to other texture categories. Pattern clustering algorithms are ideal modes for forming such clusters in the feature space. Segmentation algorithm accept as input a set of features and put a class labeling to each pixel. Fundamentally this can be considered a multidimensional

data clustering problem. Our aim in this work has been to make the segmentation scheme independent of any *a priori* knowledge about the input, thus we need an unsupervised classifier. We have used a traditional K-means clustering algorithm [8] for this purpose.

4. Results

We have implemented our algorithm on several remotely sensed images. We present our results on two IRS-1A (band-4) images and one SPOT (band-3) image. The IRS-1A images were obtained from Indian Remote Sensing Satellite. This is a circular sun-synchronous satellite and is equipped with two different sensors LISS-I and LISS-II. These IRS-1A images were taken using the scanner LISS-II (Linear Imaging Self Scanner) in the wavelength range $0.77\mu m - 0.86\mu m$ which has a spatial resolution of $36.25m \times 36.25m$. The image in fig. 3 covers an area around the city of Calcutta, while fig. 4 shows a part of the city of Bombay, both these cities are located in India.

The SPOT image fig. 5 was taken by the French satellites Systems Probatoire d'Observation de la Terre, they carry two imaging devices that consist of a linear array of charge coupled device (CCD). This image was taken in the near infrared band of wavelength $0.79\mu m - 0.89\mu m$ and have a spatial resolution of $20.0m \times 20.0m$, and covers another portion of the city of Calcutta. All the images considered in this work are of size 512×512 .

In fig. 3 the prominent black stretch across the figure is the river *Hoogly*. There is a prominent light patch on the bottom right corner, this the *Salt Lake stadium* and the black patches nearby are the *fisheries*. In the upper part right part of the image there is a distinct line structure, this the *airport runway*. In total there are five major classes in which the image can be classified. These are *water bodies* (WB), *vegetation* (VEG), *habitation* (HAB), *city area* (CA) and *open spaces* (OS).

Fig. 4 shows a part of Bombay, 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 crisscrossed structure, this is the *Santa Cruz airport*. In total the image can be classified into six major classes, these are *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 (as can be seen in fig. 4).

In fig. 5 the prominent black stretch across the figure is the river *Hoogly*. There are two distinct black, elongated patches below the river, on the left side of the image. These are water bodies, the one to the left being *Garden Reach Lake* and the one to the right being the *Khiderpore dockyard*. Just to the right of these water bodies, there is a thin line starting from the right bank of the river and going to the bottom edge of the image, this is a canal called *Talis nala*. On the right side of the image there is a triangular patch which is the *race course*. On the top right hand side of the image, there is a thin line, stretching from the top edge and ending in the middle of the image, this is the *Belehata canal*. This image has seven major classes, these are *turbid water* (TW), *pond water* (PW), *concrete* (CONCR), *vegetation* (VEG), *habitation* (HAB), *open space* (OS) and *roads/bridges* (B/R).

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 figures 3, 4 and 5, which highlight the different object regions. But the algorithms were implemented on the actual inputs.

Throughout the experiment we have used a window size of 3×3 . The test images have several fine (line) structures (roads and bridges), to detect those structures the window size has to be very small.

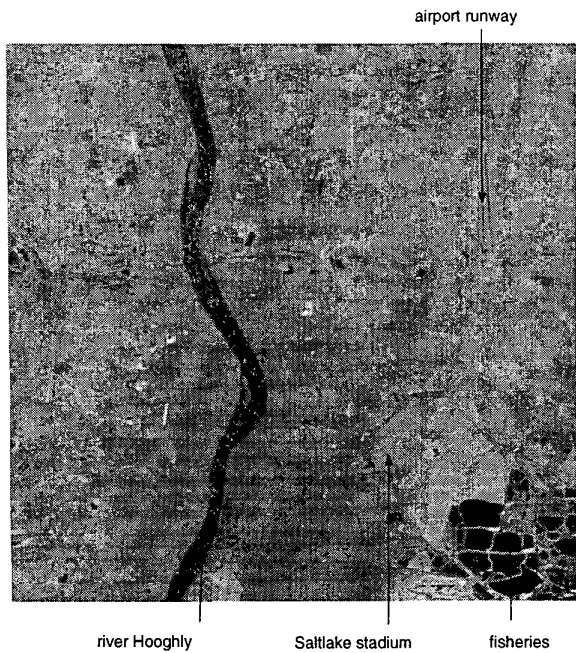
A quantitative index β [6] is being used to evaluate the segmentation results. β is defined as the ratio of the total variation and within class variation. Let n_i be the number of pixels in the i th ($i = 1, 2, \dots, c$) region, let X_{ij} be the gray value of the j th pixel ($j = 1, 2, \dots, n_i$) in the i th region, and \bar{X}_i be the mean of n_i gray values in the i th region. It is formally defined as,

$$\beta = \frac{\sum_{i=1}^c \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2}{\sum_{i=1}^c \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2} \quad (10)$$

where, n is the size of the image and \bar{X} is the mean gray value of the image.

The numerator is constant for a given image and number of class, so the value of β is dependent only on the denominator. The denominator on the otherhand decreases with homogeneity of a region. So higher the value of β better is the segmentation.

For fig. 3a the segmentation result is given in fig. 3b, with value of $c=5$ and for which β index is 3.65887. The *stadium* and *fisheries* as well as the *airport runway* are quite distinctly discernible. The segmentation result of fig. 4a. is given in 4b. with $c=6$ and β index 8.1046. The *dockyard*, the *concrete structure* connecting Bombay and New Bombay and *Santa Cruz airport* are detected in the segmented output. The segmentation result of fig. 5a. is shown in 5b. with $c=7$ and β index 3.27832. The *race course* and a triangular outline which is the *track* of the *race course* is also detected.



(a)



(b)

Figure 3. a) IRS-1A image of Calcutta b) Segmented output with $c = 5$ and $\beta = 3.65887$

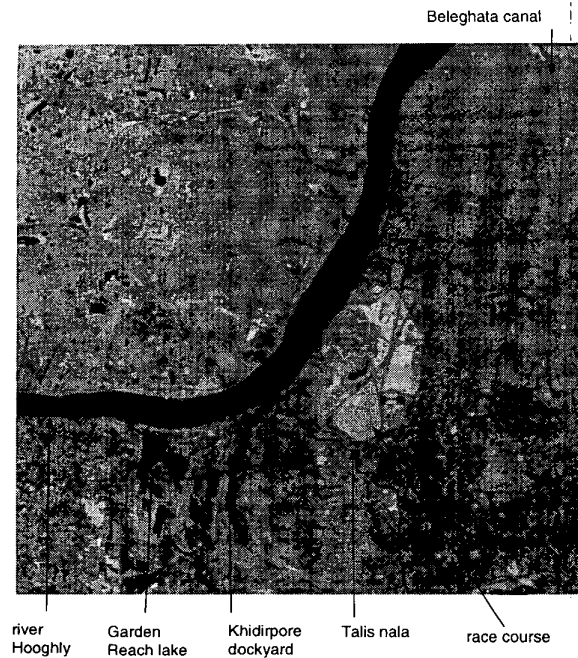


(a)

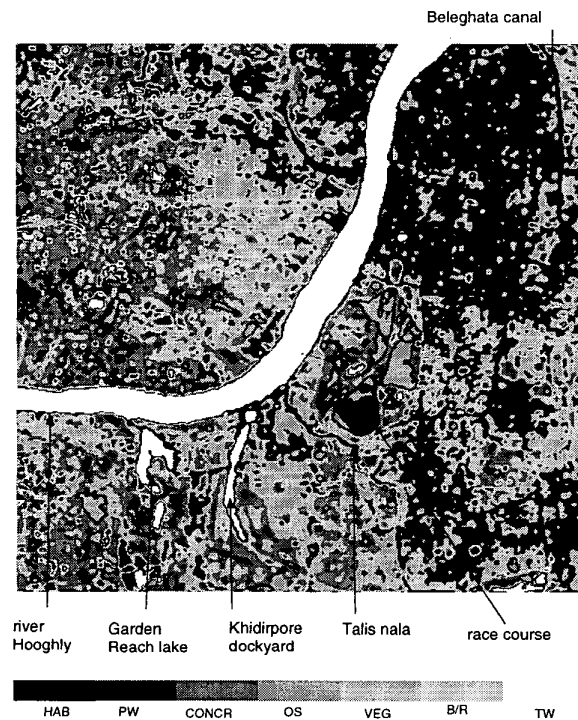


(b)

Figure 4. a) IRS -1A image of Bombay b) Segmented output with $c = 6$ and $\beta = 8.1046$



(a)



(b)

Figure 5. a) SPOT image of Calcutta b) Segmented output with $c = 7$ and $\beta = 3.27832$

5. Conclusion

One major advantage of our scheme over other methods is that even though we have made use of overcomplete wavelet representation of images, which imply large feature space (i.e. feature images of the same dimension as the input image), we have found that by intelligent selection of basis as discussed in subsection 2.2, only 4 to 6 features were sufficient for the desired segmentation of the images used in our study.

This implicates that in our method, dimensionality of the feature space can be greatly reduced, while still maintaining a high segmentation quality.

Another point which is mention worthy is that except for the knowledge of the number of classes present in a composite image we do not have any *a priori* knowledge about the test images, that is our scheme is completely unsupervised. The algorithm is computationally simple and less expensive.

It has been discussed in section 1 that assigning unique class levels with certainty is a problem of remotely sensed images. This problem can be handled using Fuzzy set theoretic approaches by associating certainty factors with each class labels. A work based on this is already in progress by us.

References

- [1] O. Alkin and H. Caglar. Design of efficient m-band coders with linear phase and perfect reconstruction properties. *IEEE Trans. Signal Processing*, 43(7):1579–1590, 1995.
- [2] S. Bandyopadhyay and S. K. Pal. Pixel classification using variable string genetic algorithms with chromosome differentiation. *IEEE Trans. on Geoscience and Remote Sensing*, (to appear).
- [3] I. Daubechies. *Ten Lectures on Wavelets*. Soc. Ind. Applied Math, Philadelphia, 1992.
- [4] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.
- [5] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(7):674–693, 1989.
- [6] S. K. Pal, A. Ghosh, and B. U. Shankar. Segmentation with remotely sensed images with fuzzy thresholding, and quantitative evaluation. *Int. Journal of Remote Sensing*, 21(11):2269–2300, 2000.
- [7] P. Steffen, P. N. Heller, R. A. Gopinath, and C. S. Burrus. Theory of regular m-band wavelet bases. *IEEE Trans. Signal Processing*, 41(12):3497–3510, 1993.
- [8] J. T. Tou and R. C. Gonzales. *Pattern Recognition Principles*. Adision-Wesley, Reading, 1974.
- [9] M. Tuceryan and A. K. Jain. Texture analysis. In *Handbook of Pattern Recognition and Computer vision*, pages 235–276. World Scientific, 1993.