Content-Based Image Retrieval Using Wavelet Packets and Fuzzy Spatial Relations

Minakshi Banerjee and Malay K. Kundu

Machine Intelligence Unit, Indian Statistical Institute
203, B.T. Road, Kolkata 700 108, India
{minakshi_r, malay}@isical.ac.in

Abstract. This paper proposes a region based approach for image retrieval. We develop an algorithm to segment an image into fuzzy regions based on coefficients of multiscale wavelet packet transform. The wavelet based features are clustered using fuzzy C-means algorithm. The final cluster centroids which are the representative points, signify the color and texture properties of the preassigned number of classes. Fuzzy Topological relationships are computed from the final fuzzy partition matrix. The color and texture properties as indicated by centroids and spatial relations between the segmented regions are used together to provide overall characterization of an image. The closeness between two images are estimated from these properties. The performance of the system is demonstrated using different set of examples from general purpose image database to prove that, our algorithm can be used to generate meaningful descriptions about the contents of the images.

1 Introduction

Effective searching of relevant images using information derived from the visual contents (color, texture, shape etc.) is the focus of interest for most research on image databases. In the recent past, Content-Based Image Retrieval (CBIR) techniques became popular [1], [2], [3], [4],[5], [6], [7], [8] for retrieving relevant images from an image database by measuring similarity between the automatically derived features (color, texture, shape etc.) of the query image and the images stored in the database [5]. But even with the most sophisticated design, it is often not possible to achieve satisfactory results, because the image features may convey different meanings under different context (semantic gap). There are ongoing efforts by the researchers for bridging the gap between human intuitive understanding and information derived from image features.

Apparently different set of features are suitable for different purposes. For example, shape related features are not suitable to extract textured pictures only. The features like color, texture, shape as well as the spatial relationship between the individual regions play a significant role in depicting the overall meaning associated with the scene. The properties of individual regions can be coupled with spatial relationships between them, to provide more realistic matching between two images. Methodologies to compute spatial relationships of the objects have

been proposed by [9], [10], [11]. However the methods mostly deal with the geometric attributes of a region like (area, shape adjacency, surroundedness etc.). Fuzzy similarity measure between regions have been introduced in FIRM [12] where the properties of all regions are integrated by a family of fuzzy features. The geometric relationships between the regions are not considered.

Extracting significant features using wavelets have been proved to be promising [13]. Wavelet transform provides a suitable framework for analysis and characterization of images at different scales [14], [15]. As significant texture information requires over complete decomposition, wavelet packet frames which comprise of all possible combinations of subband tree decomposition, can serve better representation of textural analysis than standard dyadic wavelets. Not all the bases are equally important. Hence finding out computationally efficient optimal basis based on the some statistical criteria becomes important [13].

In this paper, we propose a simple technique to identify uniform color textured regions. The approximate boundary regions are also extracted for detecting the differences of textures in adjacent regions. We use Fuzzy C- means algorithm for assigning multiclass membership values to each pixel, for approximate segmentation into homogeneous regions. We use wavelet packet frames to extract features for segmentation, and obtain the best basis based on entropy measure. The centroids of the segmented regions depict the color and textural properties. To estimate the intersection between the fuzzy partitions a property namely Index of fuzziness [16] is computed. Fuzzy Topological relation known as, shape distance [17] is computed between the fuzzy regions of the segmented image. The proposed algorithm shows fairly promising performance in case of retrieving perceptually similar images from benchmark databases. The performance is compared with other approaches like (a) Color, texture histogram similar to [18] (b) gabor texture features, to prove the efficiency of the proposed scheme.

The remaining sections are organized as follows. The segmentation algorithm and the feature extraction process is described in section 2. The Experimental results and comparative studies are made in section 3. The paper is concluded in section 4.

2   Integrating Wavelet Features for Extraction of Colored Textures

Features extracted from different frequency bands of wavelet coefficients are shown to be effective for representing texture properties [15]. This can be explained from the fact that coefficients in different frequency band show variation in different scales and directions. An input color image can be looked as a 3-D energy function $E(x, y, \lambda)$ where (x,y) denotes the spatial coordinates and λ denotes the wavelength of light energy. The local spatial frequency characteristics of $E(u, v, \lambda)$ can be used to capture texture characteristics. The properties can be combined in the wavelet domain by convolution with a wavelet filter in the spatial domain with the independent color channels [19].

$$\hat{M}(u,v,\lambda) = h(u,v) * \int E(u,v,\lambda)d\lambda \tag{1}$$

The filter bank $h(u,v)$ is a set of band pass filters with frequency selective properties. The original image is available in RGB format. The RGB values encode the color information. Each of the individual components are transformed separately. The discrete normalized scaling and wavelet basis functions are defined as [13],

$$\phi_{i,k}(l) = 2^{i/2}h_i(2^il - k)$$
$$\psi_{i,k}(l) = 2^{i/2}g_i(2^il - k) \tag{2}$$

where i and k are the dilation and translation parameters and h_i and g_i are respectively the sequence of lowpass and bandpass filters of increasing width indexed by i. The full discrete wavelet expansion of a signal that forms an orthonormal basis for $L^2(R^2)$ is given as,

$$x(l) = \sum_{k\in z} c_{(d_0)}(k)\phi_{d_0,k} + \sum_{i=1}^{d_0}\sum_{k\in z} d_{(i)}(k)\psi_{i,k} \tag{3}$$

$d_{(i)}$ are the wavelet coefficients and $c_{(d_0)}$ are the expansion coefficients of the coarser signal approximation $x_{(d_0)}$. The $c_{(d_0)}$ and $d_{(i)}$ can be interpreted in terms of simple filtering and down sampling operations. The 2-D DWT is computed by applying a separable filter bank to the image where $c_i(x,y)$ corresponds to low frequencies (LL) $d_i^1(x,y)$ corresponds to the vertical high frequencies (horizontal edges, LH), $d_i^2(x,y)$ horizontal high frequencies(vertical edges, HL), $d_i^3(x,y)$ the high frequencies in both directions (the corners, HH). The image $I(x,y)$ is represented at several scales by { cd_0, $d_i^n(x,y)$, n= 1,2,3, i=1,..., d_0 }. The 2D-DWT transform of an $2^i \times 2^i$ image is represented as a set of shifted and dilated wavelet function resulting into subbabnd images,each of size $2^{i-1} \times 2^{i-1}$. The standard dyadic transform are not suitable for analysis of high frequency signals with relatively narrow band because the subband decomposition is applied on the low pass component plane of the input image. To obtain efficient texture properties, multiscale overcomplete packet transform is necessary. Wavelet packets comprises all possible combinations of subband decomposition applied recursively to the lowpass and high pass filter results of the previous wavelet transform step. As a result, it is possible to create arbitrary tree structures that generates various combinations of orthonormal bases [20]. As an example, a 2-level decomposition on the orthogonal subspaces generates bases, $P_{2,i,k}(l)$, i= 0,1,2,3 $p_{1,1,k}(l)$=1/2$\psi(l/2 - k)$, $p_{2,0,k}(l)$=1/4$\psi(l/4 - k)$, $p_{2,1,k}(l)$=1/4$\psi(l/4 - k)$ forms orthogonal sets. The significant nodes are computed and the optimal basis is obtained based on entropy minimized tree, where the entropy function is computed between a parent and quad child. Only those nodes are considered for which the entropy exceeds the predefined threshold. The entropy minimized tree forms a nested sequence of subspaces, $V_1 \subset V_2 \subset ...V_n$.

2.1   Multiresolution Feature Extraction

Having obtained the entropy minimized tree, multiresolution analysis is performed considering the spatial similarities across subbands. In general there are spatial similarities across subbands [20]. The pixels in each subband are linked to the pixels of the adjacent subband at the next lower level. Each pixel (x, y) from the former set of subbands acts as the root of pixels $(2x, 2y)$, $(2x + 1, 2y)$, $(2x, 2y + 1)$, $(2x + 1, 2y + 1)$. For a (n) level decomposition the same rule links the pixels of the adjacent subbabnds starting from LHn, HLn, HHn, respectively. The selected bases V_n are interpolated by factor (n), to ensure the identity to the sampling spaces. This approach is computationally effective. It helps in reducing blockyness arising in case of the block based methods and the computational cost involved in pixel wise segmentation. The local features of the filter output (wavelet domain) around each (x,y) th pixel is estimated from the selected bases. The moments of the wavelet coefficients of various frequency bands have proven effective for discriminating textures [21]. To obtain the moments a Daubechies-4 wavelet transform is applied. Around the (x,y) pixels of a certain subband a $W \times W$ window is centered. The raw moments m00,m10,m01,m20,m02,m11 are computed. From which the centralized moments $\mu20,\mu02,\mu11$ are computed on each window, [22], as follows.

$$
\begin{aligned}
\mu20 &= m20 - \frac{m10^2}{m00} \\
\mu02 &= m02 - \frac{m01^2}{m00} \\
\mu11 &= m11 - \frac{m10*m01}{m00}
\end{aligned}
\tag{4}
$$

For a color input image, the moments computed along the individual channels are represented as, $F_k\lambda = [\mu20,\mu02,\mu11]$ The feature computed along the three channels are considered as linearly independent and represented as,

$$
F_k(x, y) = \sum F_{k\lambda}
\tag{5}
$$

The segmentation obtained with the moments are shown in Figs.1(b), 2(b), 3(b).

Apart from $\mu20,\mu02,\mu11$, the variance of the coefficients of the selected window is computed from the (R) plane across the dimension of rows. Classifying these features we approximately segment the boundaries between the homogeneous region as shown in Figs.1(c), 2(c).

$$
F_k(x, y)var = (\sum_{x=1}^{W} \sum_{y=1}^{W} w_k(x, y) - \bar{F}_k(x, y))^2
\tag{6}
$$

where $\bar{F}_k(x, y)$ is the local mean and $w_k(x, y)$ is the coefficients at different scales k. The size of the window is an important parameter. Accurate texture measurement demands larger window size and better localization of region boundaries require smaller window size. The extracted features are integrated in the feature space to produce approximate segmentation into homogeneous regions. The features need to be clustered into different categories. We have used the standard fuzzy C- means algorithm [23]for pattern classification. Let $X = \{x_1, x_2, ..,x_N\}$

be a finite subset of n dimensional vector space R^n where x_i is a feature vector for a pixel in the image. For an integer $C, 2 \le C \le N$, a $C \times N$ matrix $U=[u_{ik}]$ is called the fuzzy C partition of X whenever the entries U satisfies the following conditions. $u_{ij} \in [0,1]$ for all i and j, u_{ij} is the degree of membership of x_i in the cluster j, $\sum_{i=1}^{C} u_{ij} =1$ for all j and $0 \le \sum_{j=1}^{N} u_{ij} \le N$ for all i.

2.2   Attributes of Fuzzy Sets on Segmented Image

Index of fuzziness : From fuzzy C-means clustering of $F_k(x,y)$, or $F_k(x,y)var$ a fuzzy partition matrix U of size C by N is obtained, where C is number of clusters and N is number of data points. The feature vector usually belongs to multiple regions with different degrees of memberships as opposed to classical region representation in which the feature vector belongs to exactly one region. The measure of fuzziness between the C fuzzy partitions of X, as obtained from the fuzzy C-means clustering is considered. Let the fuzzy partitions $(u_1, u_2, ..$ $u_C)$ on X be defined by the set of C menbership functions, $\mu u_j(x)$ for all $x \in X$ and $j = 1, 2, ...C$. The intersection between two fuzzy partitions u_m and u_j , $(m \ne j$) from the C collections is defined from [16]

$$I(u_j \cap u_m) = \frac{1}{|X|} \sum_{x \in X} [min(\mu u_j(x), \mu u_m(x))] \tag{7}$$

Shape distance : The pixels of the fuzzy partition are assigned distinct class label for which the membership is maximum. As a result, we obtain the segmented image. Each segmented region can be looked upon a bounded fuzzy set where the membership is zero outside the region. We consider the shape properties of the fuzzy regions of the segmented image, as proposed in [17],

If $\mu(x,y)$ is the membership of the fuzzy set μ at the point (x,y) in R^2, the center of gravity of μ is the point (x_0, y_0) where $x_0 = \frac{\int_s x\mu(x,y)\,ds}{A(S)}$ and y_0 $= \frac{\int_s y\mu(x,y)\,ds}{A(S)}$ where A(S) $= \int_s u(x,y)ds$. The dissimilarity between two fuzzy sets u and v is defined as,

$$D_1(u,v) = \int_s |u-v| ds \tag{8}$$

The shape distance between two fuzzy sets u and v (here u and v corresponds to individual segmented regions) are defined as,

$$D_3 = min[D_1(u, v_{\alpha1-\alpha2}), D_1(u, v_{\alpha1-\alpha2+\pi})] \tag{9}$$

where $\alpha1$ and $\alpha2$ are the orientations defined as the angles that the major axis of u and v make with the x axis. $D_1(u, v_\alpha)$ is the dissimilarity with respect to rotation of v by angle α. The major and minor axes are perpendicular to each other and passes through the center of gravity. D_3 is the smaller of the D_1 values in these two orientations. The orientation θ of a fuzzy set can be obtained from the relation as follows,

$$tan2\theta = \frac{2 \int_s (x - x_0)(y - y_0)\mu(x, y)\,ds}{\int_s (x - x_0)^2 \mu(x, y)\,ds - \int_s (y - y_0)^2 \mu(x, y)\,ds} \tag{10}$$

It can be proved that D_3 is a metric. This metric can be used in pattern recognition and matching problems in a fuzzy framework. We have not considered the relations like left to , above etc. as the segmented regions may not represent the image at object levels. The feature representation of the segmented image can be looked as a collection of features $[F_1, F_2, ...F_C]$ representing the variance of the centroids where C represents the number of classes. The index of fuzziness I_{mj} for $m \neq j$ between any two partitions of the C separable classes are considered. Also the shape distance S_{m1m2} between any two segmented fuzzy regions (m1, m2) are computed to generate the effective characterization of an image. In order to compute feature similarities between two images, Euclidean distance metric is used to compute the dissimilarity value between the centroids. If $X = [x_1, x_2, \cdots, x_i]$ and $Y = [y_1, y_2, \cdots, y_i]$ are two feature vectors then the Euclidean distance metric between X,Y is given by,

$$E_d = \sqrt{\sum (x_i - y_i)^2} \tag{11}$$

The intuitive similarity between two regions are computed from shape distance. The rank is computed as follows. The similarity is performed on each type of features separately. Finally the rank obtained from each individual set of features are combined to get the final result. The rank obtained from matching the centroids and index of fuzziness features are decided by measuring Euclidean distance between the features of the query and the target images. The total similarity is calculated as follows :

$$d_{qt} = \sum_{k=1}^{n} (r_{tk})/n \tag{12}$$

where d_{qt} is the similarity distance of the query image q to image t. n is the total number of features r_{tk} is the rank of the image t in feature k.

3   Experimental Steps

We test our algorithm on a database consisting of 1000 images with 100 images from ten different categories down loaded http://wang.ist.psu.edu/docs/related. The experimental results are shown from Figs. 1 to Fig. 6. We grossly partition the image into three meaningful classes to analyze the region attributes and spatial relations between these regions. Fixing the number of classes may not effectively partition all database images at individual object level. It is expected that similar partitions will be generated for images with similar semantics for use in computing overall image to image similarity. If the segmentation becomes
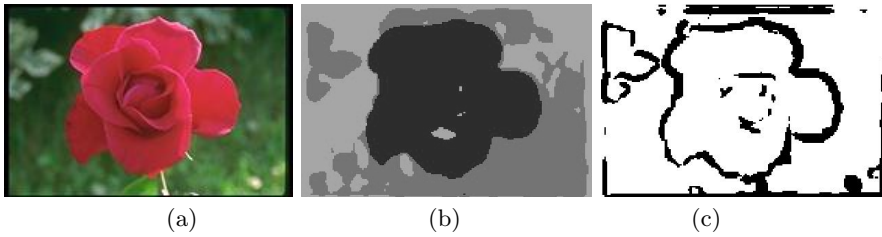
Fig. 1. (a) original image (b) segmented output (3 regions) (c) segmented output (3 class, boundaries assigned a common black label)
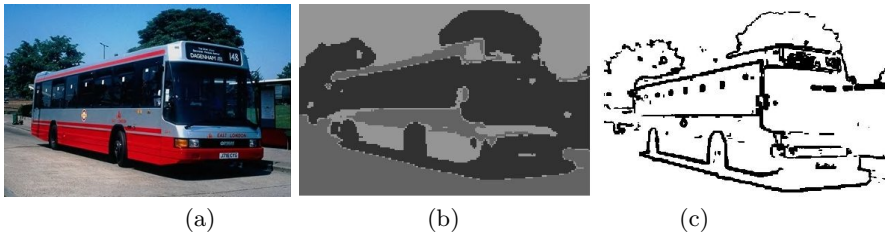

Fig. 2. (a) original image (b) segmented output (3 regions) (c) segmented output (3 class, boundaries assigned a common black label)
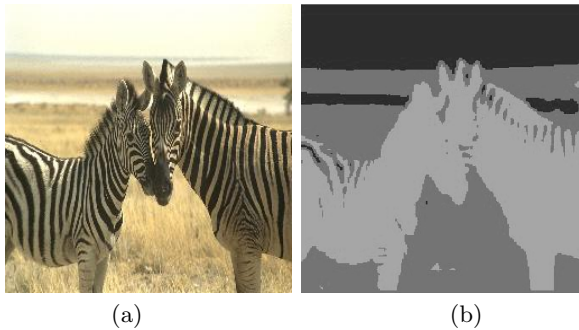

Fig. 3. (a) original image (b) segmented output (3 regions)

finer, the uncertainty in characterizing the perceptual contents increases due to detailed classifications within regions. This degrades the overall impression of the object nature. The segmentation results are shown from Figs. 1 to Fig. 3. The image shown as Fig. 1(b) is the segmented output of Fig.1(a) into three regions. Individual regions are characterized with different gray values. Fig.1(c) is the segmented output which broadly classifies the boundary between the regions using the features obtained from (6). The regions boundaries are assigned a common class label (plotted as dark boundary pixels). The segmented output of Fig. 2(a) is shown in Fig. 2(b) using the features computed from (4). The classified

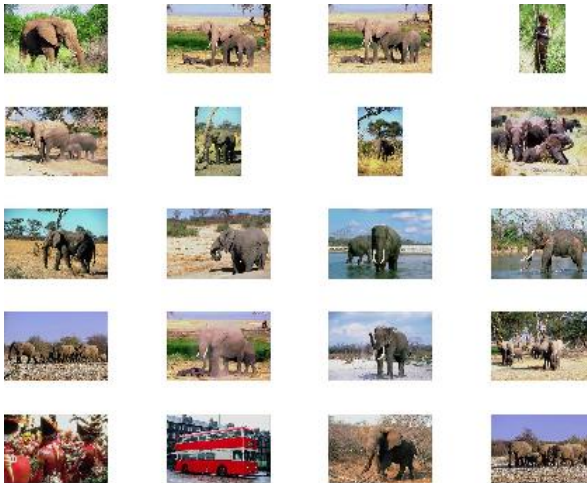**Fig. 4.** Retrieval results. With top left image as the query image.

Fig. 5. Retrieval results. With top left image as the query image.

regions boundaries plotted as dark pixels are shown in Fig. 2(c). Similarly the segmented output for Fig. 3(a) is shown in Fig. 3(b). The segmentation results shown are quite satisfactory in classifying the images into the assigned regions. By intelligent selection of the basis the number of features have been reduced for the desired segmentation. In our problem the description of the entire image is used in querying rather than taking the attribute of individual regions. We have taken query examples from all categories of images and tested the performance with the proposed features. We try two explain the results obtained on
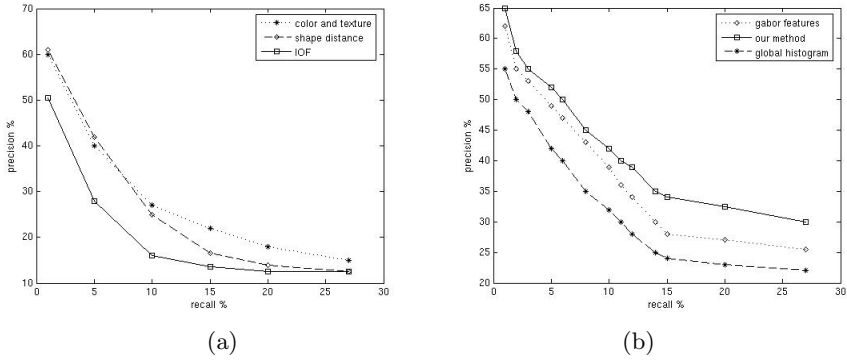
Fig. 6. (a) Recall, Precision curve from results with individual features (b) Comparative results

images with some distinctive properties. Images in Fig. 4 are constituting of a significant object with some background object. The main object almost occupies the major portion of the scene. We have tested the results considering the features (centroids) and other two spatial features separately. The combined result is shown in Fig. 4 with the top left image as as the query. The results obtained from the combined features outer performed the the individual feature retrieval. It has been found that shape distance and index of fuzziness provides a better attribute in such case. Intuitively the shape of the object itself is more meaningful than the color or texture information in the example. This provides evidence that our approach for computing relations between the regions is effective for such cases. In the second example, both the background and object are important. The object region has quite a common color and texture. The centroid features which carry the region attributes in terms of color and texture provide a better index than the shape distance. In this case the color and texture carried most of the useful information. Combining shape distance did not make significant difference to the query results. The retrieval results after combining all the features together is shown in Fig. 5 with top left image as the query. The final similarity of the combined features is found better than each of the individual feature ranking.

With each test set image as the query, the average precision and recall for individual feature retrieval is shown in the Fig. 6(a). If the system retrieves r images that belongs to the same class $C1$ as the query ($r \leq n$). If there are N_{c1} images in the class $C1$ of the query, then $P=r/n$ is the precision and $R = r/N_{c1}$ the recall for this query. We found that the similarity as obtained from index of fuzziness alone is not satisfactory. Intuitively this feature measures the fuzziness in clustering and indirectly captures dissimilarity between the regions. However the combined results improves the precision. We benchmark our results with retrieval algorithms using global color texture histograms as used in [18]. Color is represented using a 2D histogram over the HS coordinates of the HSV space. Texture is represented by two histograms (coarseness and directionality), of the

image. The similarity distance between two color histograms is computed by histogram intersection. The similarity between two textures of the whole image is determined by a weighted sum of the Euclidean distance. We also provide a comparison in retrieval results with Gabor features computed around randomly selected points. We tested the results in almost all categories of images as shown in Fig.6(b). Histogram gave satisfactory results for images objects having significant objects with quite a common color and texture. In case case of general scenes our algorithm generated better results. This proves that the proposed features captures additional spatial information of the segmented regions.

The experiment is performed in (SUN microsytems Ultra 60) system using MATLAB package. The average cputime time required for computing the feature vector is 10 seconds.

4   Conclusion

In this work we have developed an image segmentation algorithm using wavelet packet based features and the fuzzy C- means statistical clustering algorithm. The color and texture attributes of the fuzzy regions are unified with the topological features effective in capturing spatial relation between the regions. However we have not incorporated the topological properties extracted from the region boundaries in retrieval. Using these features we intend to incorporate some more spatial relation properties in order to capture crucial semantic details.

References

1. Smith, J.R., Chang, S.F.: VisualSEEK a fully automated content-based image query system. Proc. ACM Multimedia (1996) 87–98
2. Carson, C., Thomas, M., .Belongie, S., Hellerstein, J., J.Malik: Blobworld a system for region - based image indexing and retrieval. Proceedings of Visual Information Systems (1999) 509–516
3. Ma, W.Y., Manjunath, M.: NeTra :a toolbox for navigating large image databases. Proc. IEEE International Conference on Image Processing (1997) 568–571
4. Laaksonen, J., Koskela, M., Oja, E.: PicSOM :self organising maps for content based image retrieval. Proc. on International joint Conference on Neural Networks **75** (1999) 150–164
5. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence **22(12)** (2000) 1349–1380
6. Gevers, T., Smeulders., A.W.M.: Combining color and shape invariant features for image retrieval. Image and Vision computing. **17(7)** (1999) 475–488
7. Krishnapuram, R., Medasani, S., Jung, S.H., Choi, Y.S., Balasubramaniam, R.: Content-based image retrieval based on a fuzzy approach. IEEE Transactions on Knowledge and Data Engineering **16(10)** (2004) 1185–1199

8. Bracamonte, J., Ansorge, M., Pellandini, F., Farine, P.A.: Efficient compressed domain target image search and retrieval. 4th International Conference on Image and Video Retrieval, Singapore (2005) 154–163
9. Freeman, J.: The modeling of spatial relations. Computer Graphics and Image Processing **4** (1975) 156–171
10. Pal, S.K.: Fuzziness, image information and scene analysis. In Yager, R.R., zadeh, L.A., eds.: An Introduction to Fuzzy Logic Applications in Intelligent Systems. Kluwer Academic, Norwell (1992) 147–184
11. Krishnapuram, R., Keller, J.M., Ma, Y.: Quantitative analysis of properties and spatial relations of fuzzy image regions. IEEE Transactions on Fuzzy Systems **1(3)** (1993) 222–233
12. Chen, Y., Z.Wang, J., Li, J.: A region-based fuzzy feature approach to content-based image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **24(9)** (2002) 1–16
13. Mallat, S.: A theory for multiresolution signal decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence **11(7)** (1989) 674–693
14. M.Acharya, kundu, M.K.: Robust texture calssification using wavelet frames. Image Processing and communications **5(2)** (1999) 19–37
15. M.Acharya, kundu, M.K.: An adaptive approach to unsupervised texture segmentation using m-band wavelet transform. Signal Processing **81** (2001) 1337–1356
16. Pal, S.K., Majumder, D.D.: Fuzzy mathematical Approach to Pattern Recognition. Willey Eastern Limited, New York (1985)
17. Chaudhuri, B.B.: Fuzzy geometry and shape relations in image spaces. IETE Journal of Research **44(4-5)** (1998) 161–175
18. Ortega, M., Rui, Y., Chakrabarti, K., Mehrotra, S., Huang, T.S.: Supporting similarity queries in mars. In Proceedings of the 5th ACM International Multimedia Conference, Seattle, Washington **403-413** (1997)
19. Hoang, M.A., Geusebroek, J.M., Smeulders, A.W.M.: Color texture measurement and segmentation. Signal Processing **85** (2005) 265–275
20. Rao, R.M., Bopardikar, A.S.: Wavelet Transforms Introduction to Theory and Applications. Pearson Education, Asia (Singapore) (2002)
21. Unser, M.: Texture classification and segmentation using wavelet frames. IEEE Transactions on Image Processing **4(11)** (1995) 1549–1560
22. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Wiley, New York (1985)
23. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. New York, Plenium Press (1981)