Multispectral Image Segmentation Using the Rough-Set-Initialized EM Algorithm

Sankar K. Pal, *Fellow, IEEE* and Pabitra Mitra, *Student Member, IEEE,*

Abstract—The problem of segmentation of multispectral satellite images is addressed. An integration of rough-set-theoretic knowledge extraction, the Expectation Maximization (EM) algorithm, and minimal spanning tree (MST) clustering is described. EM provides the statistical model of the data and handles the associated measurement and representation uncertainties. Rough-set theory helps in faster convergence and in avoiding the local minima problem, thereby enhancing the performance of EM. For rough-set-theoretic rule generation, each band is discretized using fuzzy-correlation-based gray-level thresholding. MST enables determination of nonconvex clusters. Since this is applied on Gaussians, determined by granules, rather than on the original data points, time required is very low. These features are demonstrated on two IRS-1A four-band images. Comparison with related methods is made in terms of computation time and a cluster quality measure.

Index Terms—Clustering, granular computing, minimal spanning tree, mixture modeling, rough knowledge encoding.

I. INTRODUCTION

SEGMENTATION is a process of partitioning an image space into some nonoverlapping meaningful homogeneous regions. The success of an image analysis system depends on the quality of segmentation. Two broad approaches to segmentation of remotely sensed images are gray-level thresholding and pixel classification. In thresholding [1], one tries to get a set of thresholds $\{T_1, T_2, \ldots, T_k\}$ such that all pixels with gray values in the range $[T_i, T_{i+1})$ constitute the ith region type. On the other hand, in pixel classification, homogeneous regions are determined by clustering the feature space of multiple image bands. Both thresholding and pixel classification algorithms may be either local (i.e., context dependent) or global (i.e., blind to the position of a pixel). The multispectral nature of most remote sensing images makes pixel classification the natural choice for segmentation.

Statistical methods are widely used in unsupervised pixel classification framework because of their capability of handling uncertainties arising from both measurement error and the presence of mixed pixels. In most statistical approaches, an image is modeled as a "random field" [2] consisting of collections of two random variables $Y = (Y_s)_{s \in S}$ and $X = (X_s)_{s \in S}$. The first one takes values in the field of "classes," while the second one deals with the field of "measurements" or "observations." The problem of segmentation is to estimate Y from X. A general method of statistical clustering is to represent the probability density function of the data as a *mixture model*, which asserts that the data are a combination of k individual component densities (commonly Gaussians), corresponding to k clusters. The task is to identify, given the data, a set of k populations in it and provide a model (density distribution) for each of the populations. The Expectation Maximization (EM) algorithm is an effective and popular technique for estimating the mixture model parameters. It iteratively refines an initial cluster model to better fit the data and terminates at a solution that is locally optimal for the underlying clustering criterion [3]. An advantage of EM is that it is capable for handling uncertainties due to mixed pixels and helps in designing multivalued recognition systems.

The EM algorithm has the following limitations.

- Number of clusters needs to be known.
- Solution depends strongly on initial conditions.
- It can only model convex clusters.

The first limitation is a serious handicap in satellite image processing, since in real images the number of classes is frequently difficult to determine *a priori*. To overcome the second, several methods for determining "good" initial parameters for EM have been suggested, mainly based on subsampling, voting, and two-stage clustering [4]. However, most of these methods have high computational requirement and/or are sensitive to noise. The stochastic EM (SEM) algorithm [5] for segmentation of images is another attempt in this direction that provides an upper bound on the number of classes, robustness to initialization, and fast convergence.

Rough-set theory [6] provides an effective means for analysis of data by synthesizing or constructing approximations (upper and lower) of set concepts from the acquired data. The key notions here are those of "information granule" and "reducts." The information granule formalizes the concept of finite-precision representation of objects in real-life situations, and reducts represent the *core* of an information system (both in terms of objects and features) in a granular universe. An important use of rough-set theory and granular computing has been in generating logical rules for classification and association [7]. These logical rules correspond to different important regions of the feature space, which represent data clusters.

In this paper, we exploit the above characteristics of the rough-set-theoretic logical rules to obtain an initial approximation of Gaussian mixture model parameters. The crude mixture model, after refinement through EM, leads to accurate clusters. Here, rough-set theory offers a fast and robust (noise-insensitive) solution to the initialization, besides reducing the local minima problem of iterative refinement clustering. Also, the problem of choosing the number of mixtures is circumvented,
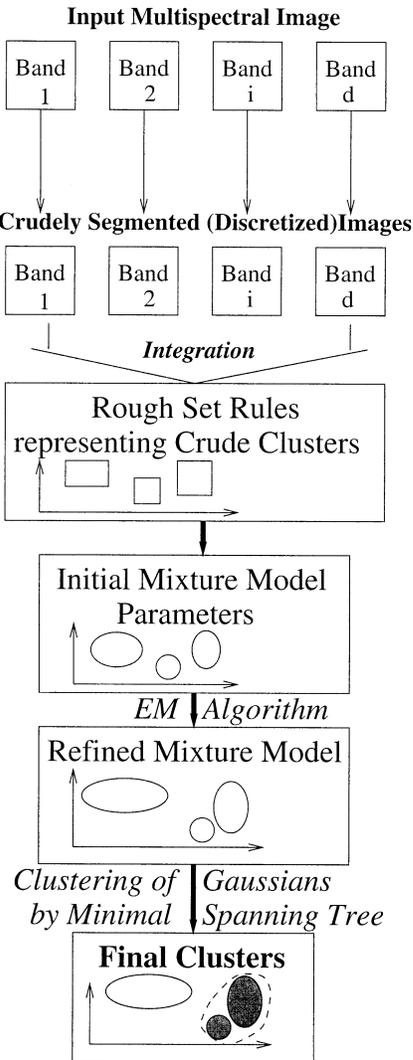
Fig. 1.   Block diagram of the proposed clustering algorithm.

since the number of Gaussian components to be used is automatically decided by rough-set theory.

The problem of modeling nonconvex clusters is addressed by constructing a minimal spanning tree (MST) with each Gaussian component as nodes and Mahalanobis distance between them as edge weights. Since MST clustering is performed on the Gaussian models rather than the individual data points and since the number of models is much less than the data points, the computational time requirement is significantly small. A block diagram of the integrated segmentation methodology is shown in Fig. 1. Discretization of the feature space, for the purpose of rough-set rule generation, is performed by gray-level thresholding of the image bands individually.

Experiments were performed on two four-band IRS-1A satellite images. Comparison is made both in terms of a cluster quality index [1] and computational time, in order to demonstrate the effect of the individual components.

II. MIXTURE MODEL AND EM ALGORITHM

The mixture model approximates the data distribution by fitting k component density functions f_h, $h = 1, \ldots, k$, to a dataset D having m patterns and d features. Let $x \in D$ be a pattern; the mixture model probability density function evaluated at x is

$$p(x) = \sum_{h=1}^{k} w_h f_h(x|\phi_h). \qquad (1)$$

The weights w_h represent the fraction of data points belonging to model h, and they sum to one ($\sum_{h=1}^{k} w_h = 1$). The functions $f_h(x|\phi_h)$, $h = 1, \ldots, k$, are the component density functions modeling the points of the hth cluster. ϕ_h represents the specific parameters used to compute the value of f_h. We use Gaussian distribution as the choice for component density function. The quality of a given set of parameters $\boldsymbol{\Phi} = \{(w_h, \mu_h, \Sigma_h), h = 1, \ldots, k\}$, is determined by the log-likelihood $L(\boldsymbol{\Phi})$ of the data, given the mixture model. The EM begins with an initial estimation of $\boldsymbol{\Phi}$ and iteratively updates it such that $L(\boldsymbol{\Phi})$ is nondecreasing. We outline the EM algorithm in the Appendix.

III. ROUGH SETS

We present some preliminaries of rough-set theory that are relevant to this paper. For details one may refer to [6] and [7].

An *information system* is a pair $\mathcal{S} = \langle U, A \rangle$, where U is a nonempty finite set called the *universe*, and A is a nonempty finite set of *attributes*. An attribute a can be regarded as a function from the domain U to some value set V_a.

With every subset of attributes $B \subseteq A$, one can easily associate an equivalence relation I_B on U: $I_B = \{(x, y) \in U$: for every $a \in B, a(x) = a(y)\}$. Then $I_B = \bigcap_{a \in B} I_a$.

We now define the notions relevant to knowledge reduction. The aim is to obtain irreducible but essential parts of the knowledge encoded by the given information system; these would constitute *reducts* of the system. Reducts have been nicely characterized in [7] by *discernibility matrices* and *discernibility functions*. Consider $U = \{x_1, \ldots, x_n\}$ and $A = \{a_1, \ldots, a_m\}$ in the information system $\mathcal{S} = \langle U, A \rangle$. By the discernibility matrix, $\mathbf{M}(\mathcal{S})$ of $\mathcal{S}$ means an $n \times n$ matrix such that

$$c_{ij} = \{a \in A\colon a(x_i) \neq a(x_j)\}. \qquad (2)$$

A discernibility function $f_{\mathcal{S}}$ is a function of m Boolean variables $\overline{a}_1, \ldots, \overline{a}_m$ corresponding to the attributes $a_1, \ldots, a_m$, respectively, and defined as follows:

$$f_{\mathcal{S}}(\overline{a}_1, \ldots, \overline{a}_m) = \wedge\{\vee(c_{ij})\colon 1 \leq i, j \leq n, j < i, c_{ij} \neq \emptyset\} \qquad (3)$$

where $\vee(c_{ij})$ is the disjunction of all variables $\overline{a}$ with $a \in c_{ij}$. It is seen in [7] that $\{a_{i_1}, \ldots, a_{i_p}\}$ is a reduct in $\mathcal{S}$ if and only if $a_{i_1} \wedge \cdots \wedge a_{i_p}$ is a prime implicant (constituent of the disjunctive normal form) of $f_{\mathcal{S}}$.

IV. ROUGH SET INITIALIZATION OF EM PARAMETERS

A. Discretization of Feature Space

Discretization of the feature space is performed by gray-level thresholding of the individual band images. Thus, each attribute (band) now takes on values in $\{1, 2, \ldots, N+1\}$, where N is the number of threshold levels for that band. The fuzzy correlation

$(C(\mu_1, \mu_2)$, defined in the Appendix) between a fuzzy representation of an image (μ_1) and its nearest two-tone version (μ_2) is used. For details of the above method, one may refer to [8]. We have considered correlation as a measure of thresholding, since it is found recently to provide good segmentation in less computational time compared to similar methods [1]. However, any other gray-level thresholding technique may be used.

B. Generation of Rough-Set Reducts

Here we discuss the methodology for generating rough-set reducts, which represents crude clusters in the feature space. Let there be m sets of discretized objects in the attribute-value table having identical attribute values, and let their cardinalities be n_{k_i}, $i = 1, \ldots, m$. Let $n_{k'_1}, n_{k'_2}, \ldots, n_{k'_m}$ denote the distinct elements among $n_{k_1}, \ldots, n_{k_m}$ such that $n_{k'_1} > n_{k'_2} > \cdots > n_{k'_m}$. Let a heuristic threshold function be defined as [9]

$$\text{Tr} = \left\lceil \frac{\sum_{i=1}^{m} \frac{1}{n_{k'_i} - n_{k'_{i+1}}}}{\text{Th}} \right\rceil \tag{4}$$

where Th is a constant $(=0.5$, say$)$, so that all entries having frequency less than it are eliminated from the table, resulting in the reduced attribute-value table $\mathcal{S}$. The value of Tr is high if most of the $n_{k'_i}$s are large and close to each other. The above condition occurs when a small number of large clusters are present. On the other hand, if the $n_{k'_i}$s have wide variation among them, then the number of clusters with smaller size increases. Accordingly, Tr attains a lower value automatically.

From the reduced attribute-value table obtained, reducts are obtained using the methodology described in Section III. From the reducts, one obtains a rule r_i, viz. $P_i \rightarrow \text{cluster}_i$, where P_i is the disjunctive normal form (d.n.f) of the discernibility function.

Also, define the support factor sf_i for a rule r_i as

$$\text{sf}_i = \frac{n_{k_i}}{\sum_{i=1}^{p} n_{k_i}} \tag{5}$$

where n_{k_i}, $i = 1, \ldots, p$, are the cardinality of the sets O_i of identical objects belonging to the reduced attribute-value table.

C. Mapping Reducts to Mixture Parameters

We describe below the methodology for obtaining the mixture model parameters, namely, the number of component Gaussian density functions (k) and weights (w_h), means (μ_h), and variances (Σ_h) of the components from the rough-set rules generated.

1) *Number of Gaussians* (k): Consider the antecedent part of a rule r_i. For each such conjunctive rule, assign a component Gaussian. Let the number of such formulae be k; then we consider k Gaussians.
2) *Component weights* (w_h): Weight of a each Gaussian is set equal to the normalized support factor sf_i [obtained using (5)] of the rule (r_i) from which it is derived, $w_h = \text{sf}_i / \sum_{i=1}^{k} \text{sf}_i$.
3) *Means* (μ_h): A rule consists of conjunction of a number of literals. The literals are interval variables of pixel values of a feature (band). The component of the mean vector

along that feature is set equal to the center (c) of the corresponding interval. Note that all features do not appear in a formulae, implying those features are not necessary to characterize the corresponding cluster. The component of the mean vector along those features that do not appear are set to the mean of the entire data along those features.
4) *Variances* (Σ_h): A diagonal covariance matrix is considered for each component Gaussian. As in means, the variance for feature j is set equal to half the width of the interval corresponding to that feature appearing in the rule. For those features not appearing in a formulae, the variance is set to a small random value.

V. CLUSTERING OF GAUSSIAN COMPONENTS USING MST

In this section, we describe the methodology for obtaining the final clusters from the Gaussian components used to represent the data. An MST-based approach is adopted for this purpose. The MST is a graph that connects a set of N points so that a complete "tree" of $N-1$ edges is built. (A tree is a connected graph without cycles.) The tree is "minimal" when the total length of the edges is the minimum necessary to connect all the points. An MST may be constructed using either Kruskal's or Prim's algorithm. The desired number of clusters may be obtained from an MST by deleting the edges having weights above a threshold. The threshold is selected from maxima of the derivative of the edge weights.

Instead of using individual points, we construct an MST whose vertices are the Gaussian components of the mixture model, and the edge weights are the Mahalanobis distance (D) between them. D is defined as

$$D^2 = (\mu_1 - \mu_2)^T 0.5(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2) \tag{6}$$

where μ_1, μ_2 and Σ_1, Σ_2 are the means and variances of the pair of Gaussians.

Note that each cluster obtained as above is a mixture model in itself. The number of its component Gaussians is equal to the number of vertices of the corresponding subgraph. For assigning a point (x) to a cluster, the probability of belongingness of x to each cluster (submixture models) is computed using (1), and the one with the highest probability $p(x)$ is assigned to x, i.e., we follow the Bayesian classification rule.

VI. EXPERIMENTAL RESULTS

Results are presented on two IRS-1A (four-band) images. The images were taken usingthe LISS-II scanner in the wavelength range 0.77–0.86 μm, and it has a spatial resolution of 36.25 m $\times$ 36.25 m. The images are of size 512×512. They cover areas around the city of Calcutta and Bombay, respectively.

For the Calcutta image, the gray-level thresholds obtained using the correlation-based methodology (described in Section IV-A) are band 1: {34, 47}, band 2: {20, 29}, band 3: {24, 30}, and band 4: {31, 36}. For the Bombay image, the corresponding values are {36, 60}, {22, 51}, {23, 68}, and {11, 25}. After discretization, the attribute-value table is constructed. Eight rough-set rules (for the Calcutta image) and seven rules (for the Bombay image), each representing a crude cluster, is obtained. The rules are then mapped to initial

TABLE I
COMPARATIVE PERFORMANCE OF DIFFERENT CLUSTERING METHODS
FOR THE CALCUTTA IMAGE

Algorithm	No. of clusters	Index β	Time (sec)
EM	5	5.91	1720
KM	5	5.25	801
REM	8	6.97	470
RKM	8	5.41	301
KMEM	8	6.21	1040
EMMST	5	6.44	1915
FKM	5	5.90	2011
Proposed	5	7.37	505

TABLE II
COMPARATIVE PERFORMANCE OF DIFFERENT CLUSTERING METHODS
FOR THE BOMBAY IMAGE

Algorithm	No. of clusters	Index β	Time (sec)
EM	5	9.11	1455
KM	5	8.45	701
REM	7	10.12	381
RKM	7	10.00	277
KMEM	7	12.71	908
EMMST	5	14.04	1750
FKM	5	9.20	1970
Proposed	5	17.10	395

parameters of the component Gaussians and refined using the EM algorithm. The Gaussians are then merged using the MST-based technique discussed in Section V; thereby resulting in five clusters (from original eight and seven Gaussians). For both images, progressive improvement was observed from the initial gray-level thresholding of the individual bands, clustering using crude mixture model obtained from rough-set rules, clustering using the refined mixture model obtained by EM, and finally to graph-theoretic clustering of the component Gaussians.

The performance of the proposed hybrid method is compared extensively with various other related ones. These involve different combinations of the individual components of the proposed scheme, namely, rough-set initialization, EM and MST, with other related schemes, e.g., random initialization and k-means algorithm. The algorithms compared are

1) randomly initialized EM and k-means algorithm (EM, KM) (best of five independent random initializations)
2) rough-set-initialized EM and k-means (centers) algorithm (REM, RKM)
3) EM initialized with the output of k-means algorithm (KMEM)
4) EM with random initialization and MST clustering (EMMST)
5) fuzzy k-means (FKM) algorithm.

For the purpose of qualitative comparison of the segmentation results, we have considered an index β [1], which measures the ratio of total variation and within-cluster variation. The higher the β value is the better is the segmentation. The detailed definition of the index β is provided in the Appendix. We also present the total CPU time required by these algorithms on a DEC Alpha 400-MHz workstation. It may be noted that except for the algorithms involving rough sets, the number of clusters is not automatically determined.

Comparative results are presented in Tables I and II. Segmented images of the city of Calcutta obtained by these algorithms are also presented in Fig. 3, for visual inspection. For the Bombay image, we show the segmented versions only for the proposed method and KM algorithm having the highest and lowest β values. The following conclusions can be arrived at from the results:

1) *EM versus KM:* It is observed that EM is superior to KM both with random and rough-set initialization. However, k-means requires considerably less time compared to EM.
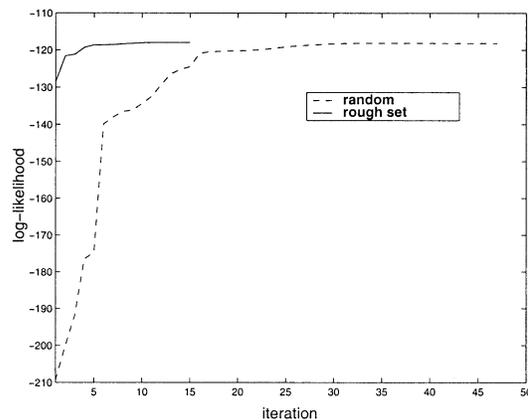

Fig. 2. Convergence of log-likelihood of EM with rough-set and random initialization.

The performance of fuzzy k-means (FKM) is intermediate between KM and EM, though its time requirement is more than EM.

2) *Effect of Rough-Set Initialization:* Rough-set-theoretic initialization (REM, RKM) is found to improve the β value as well as reduce the time requirement substantially for both EM and KM. Rough-set-initialized EM is seen to converge in much fewer steps compared to randomly initialized EM (Fig. 2). Rough-set initialization is also superior to KM initialization (KMEM).

3) *Contribution of MST:* Use of MST adds a small computational load to the EM algorithms (EM, REM); however, the corresponding integrated methods (EMMST and the proposed algorithm) show a definite increase in β value.

4) Integration of all the three components, (EM, rough set, and MST) in the proposed algorithm produces the best segmentation in terms of β value in the least computation time. This is also supported visually if we consider Figs. 5 and 6, which demonstrate the zoomed image of two man-made structures, *viz.,* river bridge and airport strips of the Calcutta image corresponding to the proposed method and KM algorithm providing the highest and lowest β values, respectively.

5) *Computation Time:* It is observed that the proposed algorithm requires significantly less time compared to other algorithms having comparable performance. Reduction in time is achieved due to two factors. Rough-set initialization reduces the convergence time of the EM al-
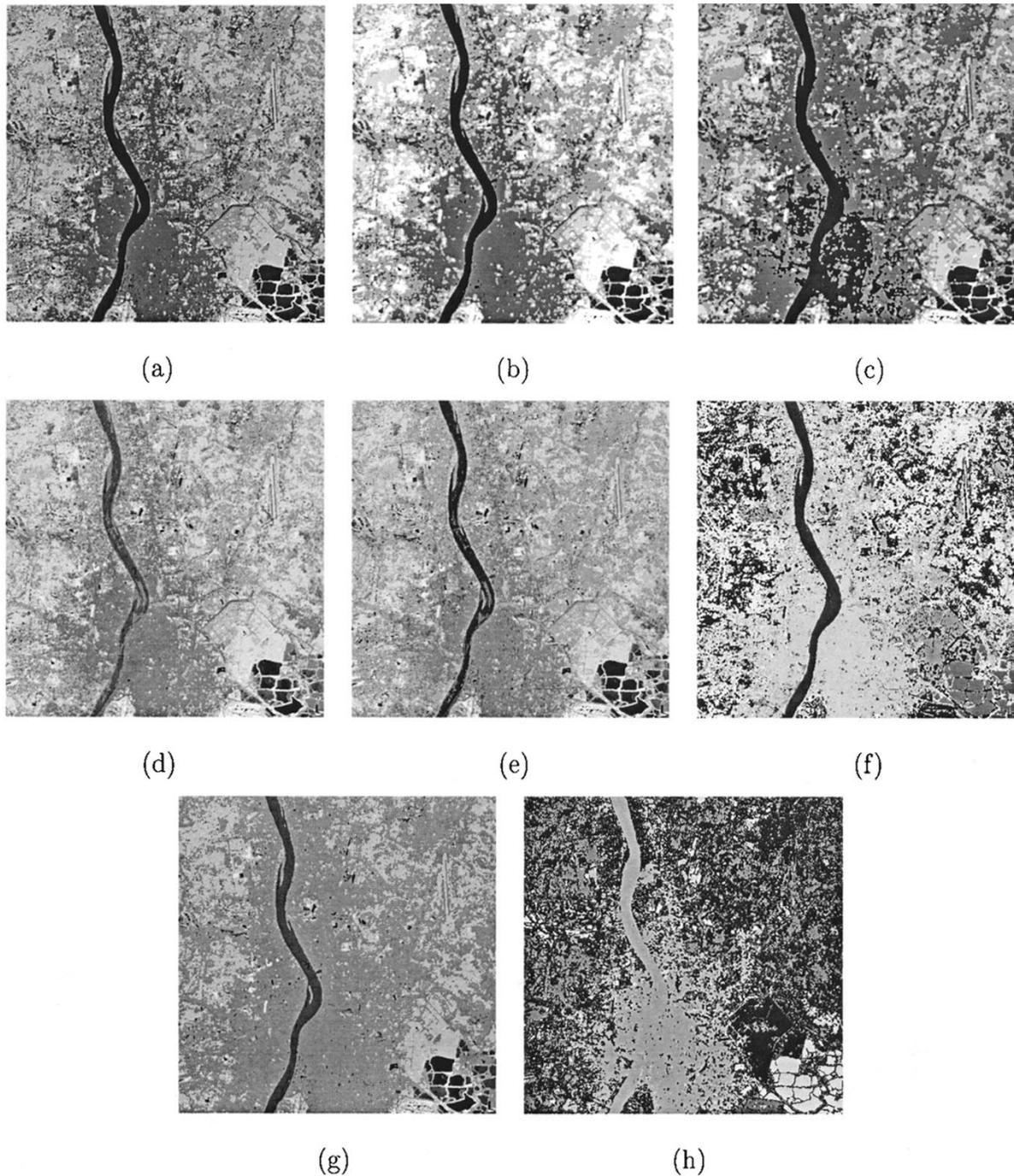
Fig. 3. Segmented IRS image of Calcutta using (a) proposed method, (b) EM with MST (EMMST), (c) fuzzy k-means algorithm (FKM), (d) rough-set-initialized EM (REM), (e) EM with k-means initialization (KMEM), (f) rough-set-initialized k-means (RKM), (g) EM with random initialization (EM), and (h) k-means with random initialization (KM).

gorithm considerably, compared to random initialization. Also, the MST, being designed on component Gaussians rather than individual data points, add very little load to the overall time requirement, while improving the performance significantly.

VII. CONCLUSION AND DISCUSSION

The contribution of the paper is twofold. First, rough-set theory is used to effectively circumvent the initialization and local minima problems of the EM algorithm. This also improves the clustering performance, as measured by the β value. Besides, the number of clusters is automatically determined.

The second contribution lies in the development of a methodology integrating the merits of graph-theoretic clustering (e.g., having the capability of generating nonconvex clusters) and iterative refinement clustering (e.g., having a low computational time requirement). At the local level, the data are modeled by Gaussians, i.e., asa combination of convex sets, while globally these Gaussians are partitioned using a graph-theoretic tech-
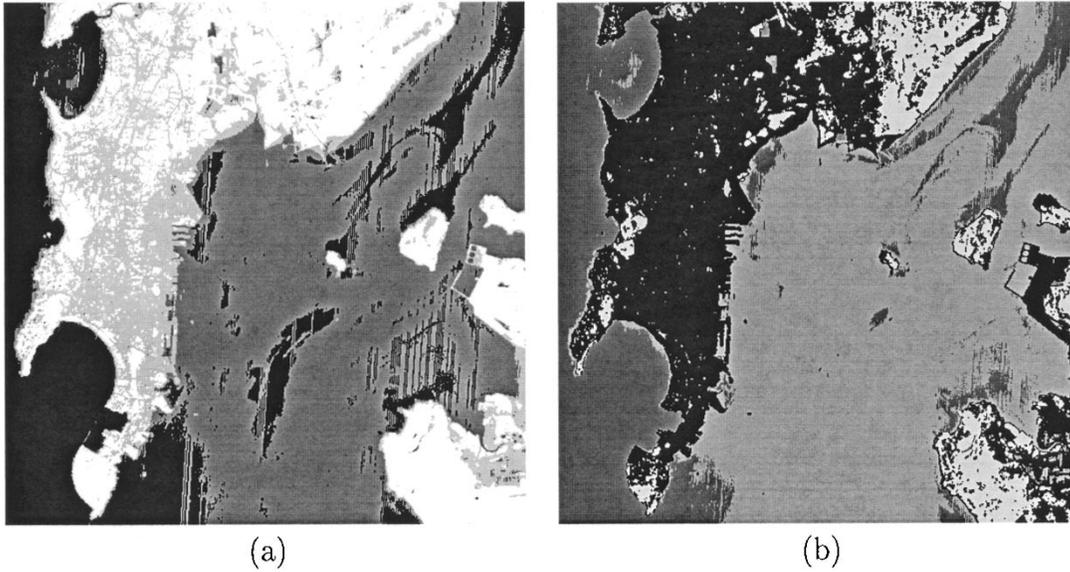
Fig. 4.   Segmented IRS image of Bombay using (a) proposed method and (b) k-means with random initialization (KM).
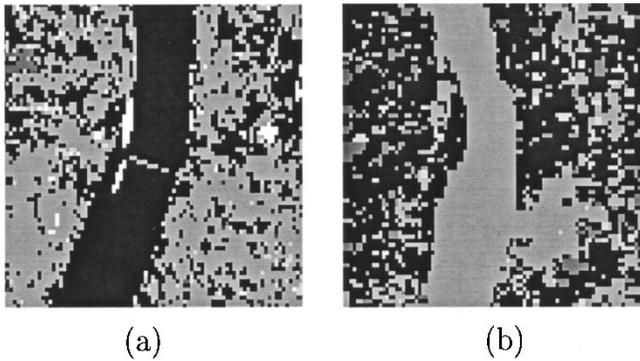

Fig. 5.   Zoomed images of a bridge on the river Ganges in Calcutta for (a) proposed method and (b) k-means with random initialization (KM).
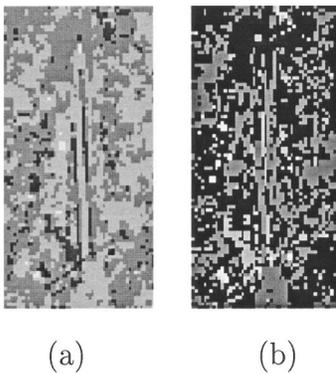

Fig. 6.   Zoomed images of two parallel airstrips of Calcutta airport for (a) proposed method and (b) k-means with random initialization (KM).

nique, thereby enabling fast and efficient detection of the non-convex clusters. The reduction in time is due to the merits of granular computing. Although the methodology of integrating rough sets, fuzzy sets, MST, and the EM algorithm has been efficiently demonstrated for segmenting remote sensing images, the concept can be applied to other unsupervised classification problems, even for mining large datasets.

It may be noted that the role of the threshold function of (4) is to reduce the size of the mixture model by eliminating the noisy pattern representatives (having lower values of n_{k_i}) from the reduced attribute-value table, thereby reducing the computational time. If no such reduction is performed, the computational time increases, but the final mixture model obtained remains almost the same, since the initial insignificant Gaussian components get merged with the larger ones when the EM algorithm converges.

APPENDIX
EM ALGORITHM

Given a dataset D with m patterns and d continuous features, a stopping tolerance $\epsilon > 0$, and mixture parameters $\mathbf{\Phi}^j$ at iteration j, compute $\mathbf{\Phi}^{j+1}$ at iteration $j + 1$ as follows.

Step 1)   *E-Step:* For pattern $x \in D$: Compute the membership probability of x in each cluster $h = 1, \ldots, k$

$$w_h^j(x) = \frac{w_h^j f_h(x|\mu_h^j, \Sigma_h^j)}{\sum_i w_i^j f_i(x|\mu_i^j, \Sigma_i^j)}.$$

Step 2)   *M-Step:* Update mixture model parameters.

$$w_h^{j+1} = \sum_{x \in D} w_h^j(x), \mu_h^{j+1} = \frac{\sum_{x \in D} w_h^j(x)x}{\sum_{x \in D} w_h^j(x)}$$

$$\Sigma_h^{j+1} = \frac{\sum_{x \in D} w_h^j(x)(x - \mu_h^{j+1})(x - \mu_h^{j+1})^T}{\sum_{x \in D} w_h^j(x)}, \qquad h = 1, \ldots, k.$$

Stopping Criterion: If $|L(\mathbf{\Phi}^j) - L(\mathbf{\Phi}^{j+1})| \leq \epsilon$, stop. Else set $j \leftarrow j + 1$ and go to Step 1). $L(\mathbf{\Phi})$ is given by

$$L(\mathbf{\Phi}) = \sum_{x \in D} \log\left(\sum_{h=1}^k w_h f_h(x|\mu_h, \Sigma_h)\right). \qquad (7)$$

Fuzzy Correlation: Fuzzy correlation $C(\mu_1, \mu_2)$ is defined as [8]

$$C(\mu_1, \mu_2) = 1 - \frac{4}{X_1 + X_2}$$
$$\cdot \left(\sum_{i=0}^{T} \{[\mu_1(i)]^2 h(i)\} + \sum_{i=T+1}^{L-1} \{[1 - \mu_1(i)]^2 h(i)\} \right) \quad (8)$$

with $X_1 = \sum_{i=0}^{L-1} [2\mu_1(i) - 1]^2 h(i)$ and $X_2 = \sum_{i=0}^{L-1} [2\mu_2(i) - 1]^2 h(i) = \text{constant}$; $L-1$ is the maximum gray level; and $h(i)$ is the frequency of the ith gray level. The *maxima* of the $C(\mu_1, \mu_2)$ represent the threshold levels.

Index β: β is defined as [1]

$$\beta = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X})^T (X_{ij} - \overline{X})}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^T (X_{ij} - \overline{X}_i)} \quad (9)$$

where n_i is the number of points in the ith $(i = 1, \ldots, k)$ cluster; X_{ij} is the feature vector of the jth pattern $(j = 1, \ldots, n_i)$ in cluster i; $\overline{X}_i$ is the mean of n_i patterns of the ith cluster; n is the total number of patterns; and $\overline{X}$ is the mean value of the entire set of patterns.

REFERENCES

[1] S. K. Pal, A. Ghosh, and B. U. Shankar, "Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation," *Int. J. Remote Sens.*, vol. 21, no. 11, pp. 2269–2300, 2000.

[2] H. Caillol, A. Hillion, and W. Pieczynski, "Fuzzy random fields and unsupervised image segmentation," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 801–810, July 1993.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, ser. B, vol. 39, pp. 1–38, 1977.

[4] M. Meila and D. Heckerman, "An experimental comparison of several clustering and initialization methods," Microsoft, Redmond, WA, Microsoft Res. Tech. Rep. MSR-TR-98-06. [Online]. Available ftp://ftp.research.microsoft.com/pub/tr/TR-98-06.PS, 1998.

[5] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 618–633, May 1993.

[6] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991.

[7] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowiński, Ed. Dordrecht, The Netherlands: Kluwer, 1992, pp. 331–362.

[8] S. K. Pal and A. Ghosh, "Image segmentation using fuzzy correlation," *Inform. Sci.*, vol. 62, pp. 223–250, 1992.

[9] S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*. New York: Wiley, 1999.

Sankar K. Pal (M'81–SM'84–F'93) received the M.Tech. and Ph.D. degrees in radio physics and electronics in 1974 and 1979, respectively, from the University of Calcutta, Calcutta, and the Ph.D. degree in electrical engineering along with the DIC degree from Imperial College, University of London, London, U.K.

From 1986 to 1987, he was with the University of California, Berkeley, and the University of Maryland, College Park, as a Fullbright Postdoctoral Visiting Fellow. From 1990 to 1992, he was with the NASA Johnson Space Center, Houston, TX, and in 1994, he was a Guest Investigator under the NRC–NASA Senior Research Associateship Program. He was also a Visiting Professor with the Hong Kong Polytechnic Univesity, Hong Kong, in 1999. He is a Distinguished Scientist and Founding Head of the Machine Intelligence Unit, Indian Statistical Institute, Calcutta, India. His research interests include pattern recognition and machine learning, image processing, data mining, soft computing, neural nets, genetic algorithms, fuzzy sets, and rough sets. He is coauthor of nine books and about three hundred research publications.

Dr. Pal received the 1990 S. S. Bhatnagar Prize (which is the most coveted award for a scientist in India), and many prestigious awards in India and abroad including the Jawaharlal Nehru Fellowship (1993), Vikram Sarabhai Research Award (1993), NASA Tech Brief Award (1993), IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award (1994), NASA Patent Application Award (1995), IETE—Ram Lal Wadhwa Gold Medal (1997), Om Bhasin Foundation Award (1998), G. D. Birla Award for Scientific Research (1999), Khwarizmi International Award (1st winner) from the Islamic Republic of Iran (2000), the INSA-Syed Husain Zaheer Medal (2001), and the FICCI Award (2000–2001) in engineering and technology. From 1997 to 1999, he served as a Distinguished Visitor of the IEEE Computer Society for the Asia–Pacific Region and held several visiting positions in Hong Kong and Australian universities during 1999–2002. He is a Fellow of the Third World Academy of Sciences, Italy, and all four National Academies for Science/Engineering in India. He is an Associate Editor and a Guest Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE COMPUTER, *Pattern Recognition Letters*, *Neurocomputing*, *Applied Intelligence*, *Information Sciences*, *Fuzzy Sets and Systems*, *Fundamenta Informaticae*, and the *International Journal of Computational Intelligence and Applications*. He is also a Member of the Executive Advisory Editorial Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *International Journal on Image and Graphics*, and *International Journal of Approximate Reasoning*.

Pabitra Mitra (S'02) received the B.Tech. in electrical engineering from Indian Institute of Technology, Kharagpur, India, in 1996.

He is currently a Senior Research Fellow at the Indian Statistical Institute, Calcutta, India. He has worked as a Scientist with the Institute for Robotics and Intelligent Systems, India. His research interests are in the area of data mining and knowledge discovery, pattern recognition, learning theory, and soft computing.