A MRF Based Segmentatiom Approach to Classification Using Dempster Shafer Fusion for Multisensor Imagery

A. Sarkar[1], N. Banerjee[2], P. Nair[1], A. Banerjee[1], S. Brahma[2], B. Kartikeyan[3], and K.L. Majumder[3]

[1] Department of Mathematics, IIT Kharagpur,
[2] Department of Computer Science and Engineering, IIT Kharagpur,
[3] Space Application Centre, Ahmedabad, India

Abstract. A technique has been suggested for multisensor data fusion to obtain landcover classification. It takes care of feature level fusion with Dempster-Shafer rule and data level fusion with Markov Random Field model based approach vis-a-vis for determining the optimal segmentation. Subsequently, segments are validated and classification accuracy for the test data is evaluated. Two illustrations of data fusion of optical images and a Synthetic Aperture Radar (SAR) image is presented and accuracy results are compared with those of some recent techniques in literature for the same image data.

Index Terms- Dempster-Shafer Theory, Hotelling's T^2, Markov Random Field(MRF), Fisher's discriminant.

1 Introduction

We address the problem of landcover classification for multisensor images that are similar in nature. Images acquired over the same site by different sensors are to be analyzed by combining the information from them. The role of feature level fusion using Dempster-Shafer(DS) rule and that of data level fusion in MRF context have been studied in this work to obtain an optimal segmented image. This segmented image is then labelled with groundtruth classes by a cluster validation scheme to obtain the classified image. Classification accuracy results of the method are evaluated with test set data and compared with that of some of recent works in literature.

A number of techniques are available in the literature [2,3,5,7] for analyzing data from different sensors or sources. An extensive review work is given in Abidi and Gonzales [1]. A very brief survey is also available in [7]. Among many approaches for data fusion Dempster theory of evidence has created a lot of interest although its isolated pixel by pixel use has not shown much encouraging results. A statistical approach with similar sources has been investigated in [5] under the assumption of multivariate Gaussian distribution incorporating a source reliability factor. This work [5] also demonstrates the use of the mathematical theory of evidence or DS rule for aggregating the recommendations

A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3212, pp. 421–428, 2004.

of the two sources. A methodological framework due to Solberg et al. [7], that considers the important element of spatial context as well as temporal context is the Markov Random Field model for multisource classification. An interesting work of Bendjebbour et al. [2] demonstrates the use of DS theory in Markovian context. In this work the MRF has been defined over pixel sites, and as such, the computation time for such an approach is expected to be very high when large number of groundtruth classes occur in a natural scene, which is usually the case. We attempt this investigation in a way similar to [6]. After an initial segmentation performed by a technique developed in the framework for tonal region image , we define a MRF on the sites comprising the initial oversegmented regions. Such oversegmented regions are expected to be merged, resulting in an optimal segmentation through an energy minimization process associated with the underlying MRF.

To consider evidences from different sensors, DS fusion is carried out pixel by pixel and is incorporated in the Markovian context while obtaining the optimal segmentation by the energy minimization scheme associated with the MRF. To incorporate the DS fusion we associate a binary variable with the energy function. This binary variable takes values depending upon some characteristics of DS labelling of the pixels of two adjacent regions in the clique potential function. If a specific DS label is found to be common to the majority of the pixels in each of these two adjacent regions then the binary variable takes the value one otherwise it is zero. Through this binary variable in the energy function, mixing of the feature level DS fusion is carried out in the data level fusion process for obtaining the optimal segmentation. The originality of the paper lies in underlining how the features of DS theory may be exploited in the MRF based segmentation approach to classification of natural scenes without much of intensive computation.

The paper is organized as follows. Section 2 describes an evidential approach for multisource data analysis with the derivation of mass functions that are used in DS fusion. Section 3 describes the MRF model based segmentation scheme. Section 4 discusses the experimental results and concludes the paper.

2 Evidential Approach for Multisource Data Analysis

We consider N separate data sensors(/sources), each providing a measurement $y_s, s = 1, 2, ..., V$, for a pixel of interest, where $V = n_R \times n_C$, n_R and n_C being the number of rows and columns of the image. Here y_s is a vector for a multidimensional source. Suppose there are K classes(true state of the nature),$\{\omega_j, j = 1, 2, ..., K\}$ into which the pixels are to be classified according to per pixel approach. The classification method involves labelling of pixels as belonging to one of these classes. We consider here pixel specific numerical data after appropriately co-aligning the pixels arising out of different sensors. As is well known, the mathematical theory of evidence or Dempster-Shafer (DS) theory [8,5] is a field in which the contributions from separate sources of data, numerical or non-numerical, can be combined to provide a joint inference concerning the labelling of the pixels. For the N sensors we thus have N mass functions $M_i, i = 1, 2, ..., N$. These functions have the following characteristics,

(i) $M_i(\phi) = 0$, where ϕ is an empty set meaning thereby a null proposition.

(ii) $\sum_{A \in 2^{\Omega}} M_i(A) = 1$ where Ω is the set of propositions for pixel labelling and 2^{Ω} represents all possible propositions. In DS theory of evidence, two more functions that are derived from this mass function, viz., plausibility(Pls) and belief(Bel) (see [8]).

The question now is how can we bring the evidences from each of the sources together to get a joint recommendation on the pixel's label with some confidence by increasing the amount of global information while decreasing its imprecision and uncertainty. The rule of aggregating evidences from different sources is called the Dempster's orthogonal sum or rule of combination [8] and is given by the combined mass function $M = M_1 \oplus M_2 \oplus .... \oplus M_N$ as follows:

$$M(\phi) = 0$$

$$M(A) = \frac{\sum_{B_1 \cap ... \cap B_N = A} \prod_{1 \leq i \leq N} M_i(B_i)}{1 - L} \tag{3}$$

where $L = \sum_{B_1 \cap ... \cap B_N = \phi} \prod_{1 \leq i \leq N} M_i(B_i)$ (4)

In remote sensing landcover classification the union of two or more labelling propositions is of little interest as the classes considered are usually distinct. Further, its mass is determined substantially by the masses for each of the simple propositions concerned. If our labelling propositions are $\omega_j, j = 1, 2, ..., K$, then the three functions viz., M, Bel and Pls are equal and give the same decision for labelling. Hence any of them may be adopted.

Mass Function Derivation.

Let us consider a set of classes $\Omega = \{\omega_1, \omega_2, ..., \omega_K\}$ into which pixels are to be classified from N different sensor image data. For each of the jth sensor, the ith class conditional density f_i^j is determined with the help of the ground truth samples. With equal prior probabilities these class conditional density functions become posterior probabilities. Now for each pixel $s \in S, S = \{1, 2, ..., V\}$, we calculate $f_i^j(y_s)$ and assign label $l_{sj} = i^*$ to the sth pixel, $i^* \in \Omega$, if

$$f_{i^*}^j(y_s) = max_{1 \leq i \leq K} \quad f_i^j(y_s)$$

At this stage each of the pixels in different images will have a label. For $s = 1, 2, ..., V$ if $l_{sj} = l_{sk} \forall j \neq k : j = 1, 2, ..., N$, then the sth pixel of the DS output image is labelled as i^*. The remaining pixels that have been assigned contradictory labels by different sensors (i.e., $l_{sj} \neq l_{sk}$ for some j, k) are kept unlabelled in the DS output image at this stage. For unlabelled pixels, we consider the power set $\Omega^* = \{\Omega_1, \Omega_2, ..., \Omega_{2^K - 1}\}$ of Ω where $\Omega_1 = \{\omega_1\}, ..., \Omega_{2^K - 1} = \{\omega_1 \cup \omega_2 \cup .... \cup \omega_K\}$ and determine normalized mass functions for each of the sensors as given over the set Ω^* by eqn(5) below. Such unlabelled pixels will be labelled by the maximum belief rule or the maximum plausibility rule for singleton classes using combined mass functions. Thus, for each sensor j we calculate the mass functions:

$$M_j^{(s)}(\Omega_i) = \frac{f_i^j(y_s^j)}{\sum_{q=1}^{t} f_q^j(y_s^j)} \tag{5}$$

where y_s^j is the intensity value of the sth pixel for the jth sensor and $t = 2^K - 1$. These mass functions $M_1, M_2, ..., M_N$ are the probabilities on Ω^*. While computing probabilities of the combined hypotheses we use the simple elementary rule of probability of union of events for singleton classes. In this way, a singleton class with a high probability reflects its dominance in the aggregate evidence and hence a pixel is more likely to be labelled with such a class which is possibly a better reflection of the true scene. Thus, the unlabelled *sth* pixel in the DS labelled output image is now labelled with the maximum belief rule, that is, labelled with i^* if

$$M^s(\{\omega_{i^*}\}) = max_{1 \leq i \leq K} M^s(\{\omega_i\}) \qquad (6)$$

where $M^s(\{\omega_i\})$ is given by eqn(3). In the next section we outline the procedure to determine the optimal segmentation based on maximum aposterior probability (MAP) estimate.

3 MRF Model Based Segmentation Scheme

We follow the scheme of Sarkar et.al. [6] in defining the MRF on a region adjacency graph(RAG) of initial oversegmented regions- the details are omitted here. Our discussion here is directed in formulating the energy function that takes the features of DS theory in MRF based segmentation approach. Minimizing this energy function will result in a MAP estimate of the optimal segmented image. After carrying out initial segmentation following the approach as in [6] on each of the selected channels of all the different sensors (say N in number) producing segments $\{t_1^\zeta, t_2^\zeta, ..., t_{u_\zeta}^\zeta\}, \zeta = 1, ..., N$, these segments are intersected among each other to give rise to a set of new segments $\{\cap_{\zeta=1}^N t_{i_\zeta}^\zeta | \quad 1 \leq i_\zeta \leq u_\zeta, \zeta = 1, ..., N\}$ comprising a merged initial segmented image which is then passed as an input to the MRF model. Since each of the sensor images are co-aligned pixel by pixel and the intensity values are all numerical we may consider all the sensor data together as if they were from a single source having multiple channels. It is assumed that the merged initially segmented image has Q number of regions $R_1, R_2, ..., R_Q$ and a set of labels $X = \{X_1, X_2, ..., X_Q\}$ each $X_i \in \omega = \{\omega_1, \omega_1, ..., \omega_q\}$, a set of discrete values or labels, corresponding to the spectral classes of the image. The objective we adopt is to assign the region labels satisfying the constraints of an optimal segmentation for $r_1 + r_2 + ... + r_N = P$ channels multisensor imagery. We impose two constraints as per our notion of optimal segmentation from multisensor image data.
(i) An optimal segmented image region R_i should be uniform with respect to the measured characteristics as obtained from all the sensors.
(ii) Two distinct adjacent regions R_i and R_j should be as dissimilar as possible with respect to the measured characteristic as evident from the combined evidence from all the sensors.

As per merged initial segmented image, the same regions R_i, $i = 1, 2, ..., Q$ are grown in each of the channels of the different sensors. Thus, the multichannel image is initially segmented into a set of Q disjoint regions denoted by $R_1 = R_1(p), R_2 = R_2(p), ..., R_Q = R_Q(p), p = 1, 2, ..., P$. Representing each region R_i as a node with multichannel information, a RAG, $\Gamma = (R, E)$ is defined, where $R = \{R_i; 1 \leq i \leq Q\}$ is a set of nodes and E is a set of edges connecting them. With appropriate neighborhood system a MRF is defined (see details in [6]). The posterior probability distribution is given by

$$P(\mathbf{X}=\mathbf{x}|\mathbf{Y}=\mathbf{y}) = \frac{e^{-U_{ps}(\mathbf{X}|\mathbf{y})}}{Z_{ps}},$$

where $Z_{ps} = \sum_{\mathbf{X}} e^{-U_{ps}(\mathbf{X}|\mathbf{y})}$. The events $\{\mathbf{X}=\mathbf{x}\}$ and $\{\mathbf{Y}=\mathbf{y}\}$ represent respectively a specfic labelling configuration and a specific realization. Since the energy function $U_{ps}(\mathbf{x}|\mathbf{y})$ is a sum of the clique potentials $V_c(\mathbf{x}|\mathbf{y})$, it is necessary to select appropriate cliques and clique potential functions to achieve the desired objective. For the cliques and clique potential functions only the set of adjacent two-region pairs, each of which is directly connected in the RAG are considered here. The two components of the energy fuction as per the two constraints are denoted as the *region process*(H) and the *edge process*(B) respectively.

Let $\mathbf{M}_1, \mathbf{M}_2, ..., \mathbf{M}_Q$ represent the mean intensity vectors of the initially segmented regions where each $\mathbf{M}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} \mathbf{Y}_{ik}$, is a $(P \times 1)$ vector, n_i the number of pixels in the region R_i and let $\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_Q$ represent the scatter matrices, that is, $\mathbf{S}_i = \sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)^{'}$ is a $(P \times P)$ matrix with elements of sum of squared deviations from the mean and sum of cross-product deviations.

Region process (H): A measure of the uniformity of the region with respect to its intensity values is given by the elements of the matrix of sum of squares deviations from mean and cross products, i.e., $\sum(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)^{'}$ or equivalently by the generalized covariance (see [6]) of the region R_i. With the above measure of uniformity, the evidences from different sensors may also be combined with the following scheme. This scheme takes into account the pattern of the DS labels of pixels of two regions R_i and R_j belonging to a clique c and thus examines whether the majority pixels of each of these two regions are having the same class labels. If R_i and R_j are regions belonging to a clique c, then corresponding to the first constraint a clique potential function [6] can be defined as

$V_c(\mathbf{x}|\mathrm{H}) = (\eta_{ij}/\nu_{ij})[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)^{'} + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j) * (\mathbf{Y}_{jk} - \mathbf{M}_j)^{'}]$ where $\nu_{ij} = n_i + n_j - 2$ and η_{ij} is a binary variable taking values 0 and 1. It takes the value 1 when the following two conditions are together satisfied. The first condition is on the pattern of DS labels for the regions R_i and R_j as mentioned above. The second condition is that regions are homogeneous with respect to the multisensor pixel intensity values. If any of the above two conditions is violated, η_{ij} takes the value 0. It may be noted that $\eta_{ij} = 1$ indicates that $x_i = x_j$. With this variable η_{ij}, the feature level fusion is coupled with data level fusion in the energy minimization process. However, with only the above definition, the dissimilarity between adjacent regions is not taken into account and the formulation of the energy function is not complete. Therefore, an edge process is introduced through the second constraint as given below.

Edge process (B): We note that merging the two distinct regions R_i and R_j results in a new scatter matrix of the merged region as given by

$$\mathbf{S}_{ij} = \mathbf{S}_i + \mathbf{S}_j + (\mathbf{M}_i - \mathbf{M}_j) * (\mathbf{M}_i - \mathbf{M}_j)^{'} \frac{n_i n_j}{n_i + n_j}.$$

The third term is also a $P \times P$ matrix whose elements exhibit a measure of dissimilarity existing between the regions R_i and R_j. Incorporating the edge process we re-define the clique potential function as

$V_c(\mathbf{x}|\mathrm{H,B}) = V_c(\mathbf{x}|\mathbf{y}) = (\eta_{ij}/\nu_{ij})[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)' + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j) * (\mathbf{Y}_{jk} - \mathbf{M}_j)'] + \theta_{ij}(1 - \eta_{ij})\frac{n_i n_j}{n_i + n_j}(\mathbf{M}_i - \mathbf{M}_j) * (\mathbf{M}_i - \mathbf{M}_j)'$
The parameter θ_{ij} controls the weight to be given to the two processes for regions involved in the clique c. For convenience we write the above equation as $V_c(\mathbf{x}|H,B) = \eta_{ij}\mathbf{W}_{i,j} + \theta_{ij}(1-\eta_{ij})\mathbf{B}_{ij}$. Here, $\mathbf{B}_{ij} = \frac{n_i n_j}{n_i+n_j}(\mathbf{M}_i - \mathbf{M}_j)*(\mathbf{M}_i - \mathbf{M}_j)'$ and $\mathbf{W}_{ij} = \frac{1}{\nu_{ij}}[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i)*(\mathbf{Y}_{ik} - \mathbf{M}_i)' + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j)*(\mathbf{Y}_{jk} - \mathbf{M}_j)']$. A suitable comparative criterion among the elements of these two matrices $\mathbf{B}_{ij}$ and $\mathbf{W}_{ij}$ is necessary for deciding the merging of two adjacent regions. Since the ratio of $\mathbf{B}_{ij}$ and $\mathbf{W}_{ij}$ can be expressed as

$$T^2 = (\mathbf{M}_i - \mathbf{M}_j)'[(1/n_i + 1/n_j)\mathbf{s}_{pooled}]^{-1}(\mathbf{M}_i - \mathbf{M}_j),$$

where $\mathbf{s}_{pooled} = \frac{\mathbf{S}_i + \mathbf{S}_j}{\nu_{ij}} = \mathbf{W}_{ij}$, the comparative criterion needed here is based on Hotelling's T^2 statistics.

Therefore, given that the Dempster Shafer labelling is same (according to the region labelling scheme followed) for the regions R_i and R_j in the clique, the regions should be merged if $T^2 < F_\alpha$ and the regions should not be merged if $T^2 \geq F_\alpha$, where $P[T^2 > F_\alpha] = \alpha$ [as in [6, p. 1106]]. It is also to be noted here that the minimization of the energy function has been investigated by first identifying the node having the maximum aggregate clique potential with its neighbor j, $\Delta V_{ci} = \sum_j \sum_{p=1}^{P} s_{pp(i)} + \sum_{p=1}^{P} s_{pp(j)} - \sum_{p=1}^{P} W_{pp(ij)}$.
The segmented image so obtained is by minimizing the energy function $U_{ps}(\mathbf{x}|\mathrm{H,B}) = \sum_{c\epsilon C} V_c(\mathbf{x}|\mathrm{H,B}) = \sum_{c\epsilon C} V_c(\mathbf{x}|\mathbf{y})$ as described above. The flowchart of the methodology is depicted in Fig.1.
Cluster Validation Scheme: In order to validate the segments of the optimal segmented image we follow the first stage of the cluster validation scheme of Sarkar et.al. [6] and label the unlabelled segments using Fisher's method for discriminating among K ground truth classes.

4 Experimental Results and Conclusion

The proposed methodology of Markov Random Field based segmentation approach to classification for multisensor data in the context of DS theory has been applied to two subscenes. Both of them are of one optical sensor with four channels and a SAR image of the same site. The first subscene is of size 442×649 and the second subscene is of size 707×908.

The date of acquisition of the subscenes for optical image was 19 January 2000 and for the SAR was 30 September 1999, thus having a time lag of 110 days. There are 12 and 16 different land cover classes involved in the first and the second subscenes respectively. The total available groundtruth samples which equals about 5.5% of the total number of pixels of the first subscene and 3.5% that of the second subscene are divided into two subsamples. For both these subscenes, the first subsample is used for labelling some of the clusters as in [6] and subsequently, the remaining clusters are labelled with the help of these labelled clusters using Fisher's discriminant scores. The second subsamples in each of the subscene are used for the quantitative evaluation of the classification accuracy after all clusters are validated. The measurements from

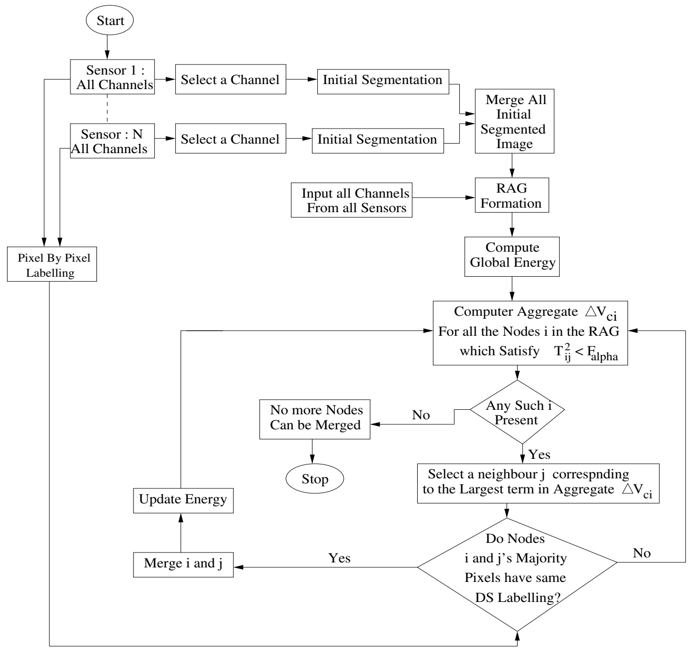

Fig. 1. Multisensor Image Segmentation Scheme

different sensors is assumed to be conditionally independent [7]. The probability density function(pdf) of the SAR intensity distribution after it is made speckle free has been considered to be Gaussian. The pdf of an optical image with 4 channels(bands) is considered to be multivariate Gaussian. We investigate the following approaches.

Case (i) (Proposed Method): Initial segmentation is first performed in each of the sensor's selected channel. These initial segmented images, one on channel-2 of the optical image and the other on the SAR image, are then merged as described in section 3. The aggregate evidences of the different sensors as obtained with eqn (3) are then incorporated into data level fusion in image space (spatial context) through the energy minimization process. Finally, a cluster validation scheme is applied to this segmented image. For the sake of comparison we have investigated the approach of Tupin et al.[9] as case(ii) and two other nonparametric multisensor fusion methods as case(iii) and case(iv) respectively.

Case (ii): In this approach [9] a direct classification is done on the initial segmented regions of the RAG using the DS rule. Unlike the proposed methodology where a separate set of labels is used for labelling the RAG, the regions are labelled here from the set of thematic (groundtruth) classes.

Case (iii): Multilayer Perceptron and Case (iv): Radial Basis functions[4].

A comparison of classification accuracies of the proposed methodology along with case (ii) through case (iv) for both subscenes are presented in Table-I. This table provides normalized classification accuracies and time durations in a

Pentium IV system with 1.86GHz and 512MB RAM. The (∗) in Table-I indicates the methods over which Case(i) is significantly better (by Kappa coefficients). The test results show that the proposed method has an edge over other methods.

Table 1. Comparison of Accuracies For All the Four Cases of the two Sub-scenes

	SubScene 1		SubScene 2	
Approaches	Normalized Accuracy	Time(hh:mm:ss)	Normalized Accuracy	Time(hh:mm:ss)
Case (i)	95.4	1:16:00	85.1	12:01:00
Case (ii)	91.4*	3:04:00	84.0*	14:02:00
Case (iii)	87.1*	0:02:35	83.2*	0:04:55
Case (iv)	83.7*	1:00:50	82.3*	2:20:00

Acknowledgment. (The work was supported by the ISRO Grant, Ref: 10/4/416, Dt. 27 Feb.2003)

References

1. M.A. Abidi and R.C. Gonzalez,"*Data Fusion in Robotics and Machine Intelligence*". New York, Academic, 1992.
2. A.Bendjebbour, Y.Delignon, L.Fouque,V.Samson and W Pieczynski,"Multisensor Image Segmentation Using Dempster-Shafer Fusion in Markov Fields Context",*IEEE Trans. Geosci. Remote Sensing*, vol.39, no.8, pp 1789-1798, Aug. 2001.
3. S. Le Hegarat-Mascle, I. Bloch and D. Vidal-Madjar,"Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing", *IEEE Trans. Geosci. Remote Sensing*, vol.35, pp 1018-1031, July 1997.
4. Y.S.Hwang and S.Y.Bang,"An efficient method to construct a radial basis function neural network classifier",*Neural Networks*,vol.10, pp. 1495- 1503, August 1997.
5. T.Lee, J.A.Richards and P.H.Swain,"Probabilistic and evidential approaches for multisource data analysis",*IEEE Trans. Geosci. Remote Sensing*, vol.GRS-25, pp 283-293, May 1987.
6. A. Sarkar, MK Biswas, B.Kartikeyan, V.Kumar, K.L.Majundar and D.K.Pal, "A MRF Model Based Segmentation Approach to Classification for Multispectral Imagery",*IEEE Trans. Geosci. Remote Sensing*, vol.40, pp 1102-1113, May 2002.
7. A.H.Schistad Solberg, T.Taxt and A.K.Jain, " A Markov random field model for classification of multisource satellite imagery",*IEEE Trans. Geosci. Remote Sensing*, vol.34, pp 100-113, Jan. 1996.
8. G. Shafer, "*A Mathematical Theory of evidence.*" Princeton, NJ: Princeton University Press, 1976.
9. F. Tupin, I.Bloch and H.Maitre," A First step towards automatic interpretation of SAR images using evidential fusion of several structure detectors",*IEEE Trans.of Geosci. Remote Sensing*, vol.37, pp 1327-1343, Mar. 1999