LANDCOVER CLASSIFICATION IN MRF CONTEXT USING DEMPSTER-SHAFER FUSION FOR MULTISENSOR IMAGERY

A.Sarkar***, A.Banerjee*****, N.Banerjee******, S.Brahma******,**

B.Kartikeyan $^{\bullet}$**, M.Chakraborty** $^{\bullet}$ **and K. L.Majumder** $^{\bullet}$

$\star$ Department of Mathematics IIT Kharagpur.(anjan@maths.iitkgp.ernet.in)

$\star\star$ Department of Computer Science and Engineering, IIT Kharagpur.

$\bullet$ 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 are 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.

(The work was supported by the ISRO under Grant for the Development of Landcover Classification Methodology with Fusion of Data from Different Sensors, Ref: 10/4/416, Dt. 27 Feb.2003)

1 Introduction.

In this work we address the problem of landcover classification for multisensor images that are similar in nature. Images of the same site by different sensors are to be analyzed by combining the information available in them. Such a combination of is necessary since data from individual sensors are usually incomprehensive - they may be partially complementary and partially redundant as sensors have different characteristics and physical interaction mechanisms. In principle, fusion of multi-source data for a purpose provides significant improvements over a single source data. Moreover, fusion can be carried out at different stages viz. data level, feature level, decision level, etc. An excellent framework for different stages of multisensor data fusion is available in Hall and Llinas [16]. Although it is advantageous to carry out fusion at one of the levels, yet there is a possibility of further improvement if we combine two different types of fusion. With this idea, the role of feature level fusion using Dempster-Shafer(DS) rule and that of data level fusion in MRF context is studied in this work to obtain an optimal segmented image. This segmented image is subsequently labelled with groundtruth classes by a cluster validation scheme to obtain the classified image.

There are numerous reports available in the literature [32, 27, 33, 19, 3, 6] which analyze data from different sensors or sources. An extensive review work is presented in Abidi and Gonzales [1]. A very brief survey is also available in [33]. In the past many approaches for data fusion has been attempted. However, Dempster theory of evidence has created a lot of interest among researchers although its isolated pixel by pixel use has not shown encouraging results. Among the statistical approaches, the simplest method adopted for fusion is to form an extended data vector for a pixel, comprising information from all the sources that are similar in nature. A statistical approach with similar sources has been investigated in [27] under the assumption of multivariate Gaussian distribution incorporating a source reliability factor. In [27] the authors also demonstrate the use of the mathematical theory of evidence or DS rule for aggregating the recommendations of the two sources. Although the latter approach has certain advantages, the statistical approach demonstrates better performance in overall classification for the numerical data investigated. The DS theory at pixel level [19] has also been used for unsupervised identification of landcover type with some degree of success. A methodological framework is presented in Solberg et al.[33] that considers the important element of spatial context as well as temporal context in the Markov Random Field model for multisource classification. The performances analysis of the proposed MRF fusion model was found to be quite encouraging. An interesting work of Bendjebbour et al.[3] demonstrates the use of DS theory in Markovian context. In this work, the MRF was 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.

The MRF model based image segmentation allows the spatial interactions between pixels and renders different Bayesian methods of segmentation very effective. Hence, such a model has been an area of interest for several researchers [10, 5, 15, 11, 13, 38]. However, Its main drawback is that it is computationally intensive. A few works in the literature aim at minimizing the running time of the above approach [7, 24, 25, 31]. As shown in [31] a substantial reduction of computational time is possible if the MRF is defined on a set of preliminary segmented regions instead of defining it on image pixels. In this work, we adopt a policy similar to [31]. After an initial segmentation performed by a technique developed for tonal region image [23], 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. In principle, every initial oversegmented image can be a valid input to the MRF modeling as long as the regions are sufficiently large to estimate the region characteristics.

In this work, the Dempster Shafer fusion is carried out pixel by pixel and is incorporated in the Markovian context when obtaining the optimal segmentation with the energy minimization scheme associated with the MRF. To incorporate the DS fusion, we associate a binary variable with the energy function. This value of the binary variable depends on the characteristics of the DS labelling of the pixels of two adjacent regions in the clique potential function. These DS characteristics of each region may be defined according to either of the two schemes described below. In the first scheme, it is the DS label of a certain prototype pixel specific to the region which is considered, and in the other, it is a specific DS label that is common to the majority of the pixels in the region. The latter DS characteristic is investigated in this work. The value taken by the binary variable can be either 0 or 1 as per the scheme defined. Through this binary variable in the energy function, mixing of feature level DS fusion is carried out in the data level fusion process for obtaining 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 in a computationally efficient manner.

The remaining paper is organized as follows. In Section II we present a framework of MRF on a region adjacency graph (RAG). Section III describes an evidential approach for multisource data analysis with the derivation of mass functions that are used in DS fusion. In section IV, the segmentation scheme with a flow chart has been discussed. Section V discusses the experimental results and Section VI concludes the paper.

2 The framework of MRF on RAG

In this section, we describe a Markov random field model which has been used as a framework for addressing data level fusion of multisensor imagery. We couple it with feature level fusion information described in the next section, in order to obtain an optimal segmented image. We follow the scheme of Sarkar et.al. [31] in defining the MRF on a region adjacency graph(RAG) of initial oversegmented regions. We also note here that our methodology for the problem of landcover classification with multisensor data is based on a segmentation approach to classification. In our approach, we first select a channel to be used for obtaining an initial oversegmented image for each sensor. For each sensor, among the available channels, the channel that accounts for maximum amount of spectral/intensity variations for the classes is selected for initial segmentation (see details in [35] and [30]).

The approach of Kartikeyan and Sarkar [23], which is based on tonal region characteristics of gray value images is adopted here for initial segmentation. Each region of the oversegmented image has only a few pixels, but just enough to measure the region characteristics. WIth such an initial segmentation, the MRF model based methodology adopted here is very close to a per-pixel based segmentation approach but has an added sdvantage of reducing computation time (see [29, p.811]). After carrying out initial segmentation 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 $\{\cap_{\zeta=1}^N t_{i_\zeta}^\zeta | \quad 1 \le i_\zeta \le u_\zeta, \zeta = 1, ..., N\}$ 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.

The observed multisensor image consists of $n_R \times n_C$ pixel vectors, n_R and n_C being the number of rows and columns of the image. The pixel vectors are

$y(1,1), y(1,2), ..., y(n_R, n_C)$ where $y_s = y(i,j) =$

$(y(i,j,1), ..y(i,j,r_1), y(i,j,r_1+1), ..., y(i,j,r_2), y(i,j,r_2+1), ..., y(i,j,r_3), ..., y(i,j,r_N))$, $r_1, r_2, ..., r_N$ being the respective number of channels for each sensor and $r_1 + r_2 + ... + r_N = P$, the total number of channels from all the sensors. 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, that depends only on the pixel intensity values. We note here that these are are quite different from the groundtruth classes.

We seek to obtain an optimal segmented image from this intersection of the initial oversegmented images by appropriately merging the oversegmented regions arising out of the tonal based segmentation technique used in initial segmentation. The number of possible label configurations for the merged initial segmented image is finite, namely q^Q but extremely large. Such a large number of possibilities can be reduced by using certain optimization criteria. Without such criteria all q^Q possible labellings must be tested to determine the best labelling configuration resulting in an optimal segmentation. The objective we adopt is to assign the region labels satisfying the criteria of an optimal segmentation for $r_1+r_2+...+r_N = P$ channels multisensor imagery. We impose two optimization criteria 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.

The first criterion involves the properties of each individual region, while the second criterion involves two adjacent regions. The region characteristics with respect to the above criteria are measured from the consensus interpretation of the sensors. The consensus interpretation of the sensors is available from the per pixel combined mass function as described in section II as well as spatial interactions via MRF modelling. Spatial adjacency is involved in the second criterion. Two regions are adjacent if they share a common boundary. Adjacency is a symmetric relationship that can be easily represented graphically with the help of a RAG. 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. In a RAG a node can be used to represent a region corresponding to $p = 1, 2, ..., P$ number of channels and an arc to represent adjacency between nodes. Through a Markov random field model one can incorporate the desired statistical properties among the nodes of the RAG by choosing appropriate cliques and clique potential functions. Representing each region as a node with $p = 1, 2, ..., P$, channel characteristics, a graph structure based on the regions is as follows.

Let $\Gamma = (R, E)$ be the RAG where $R = \{R_i; 1 \leq i \leq Q\}$ is a set of nodes representing regions of the multichannel image and E is a set of edges connecting them. Further, it is assumed that there exists a neighborhood system on Γ denoted by $\eta = \{\eta(R_i) : 1 \leq i \leq Q\}$, where $\eta(R_i)$ is a set of regions in R which are neighbors of R_i. Let $\mathbf{X} = \{X_i, 1 \leq i \leq Q\}$, where each X_i denotes the label for all the P channels and is a discrete valued random variable taking values from a finite set $\omega = \{\omega_1, \omega_2, ..., \omega_q\}$. Let Θ be the set of all configurations.

$$\Theta = \{(x_1, ..., x_Q) : x_i \in \omega, 1 \leq i \leq Q\} \tag{1}$$

The event $\{X_1 = x_1, ..., X_Q = x_Q\}$ is abbreviated as $\{\mathbf{X} = \mathbf{x}\}$ represents a specific configuration . Then X is a MRF with respect to the above neighborhood η [31]. Now by the equivalence of MRF and GRF (see [5]) a MRF defined on this RAG [31] can be represented by the Gibb's distribution defined by

$$P(\mathbf{X} = \mathbf{x}) = \frac{exp(-U(\mathbf{x}))}{Z} \tag{2}$$

where U($\mathbf{x}$), referred to as the energy function is defined as

$$U(\mathbf{x}) = \sum_{c \in C} V_c(\mathbf{x}) = \sum_{c \in C} V_c(x, p), p = 1, 2, ..., P \tag{3}$$

and Z is a normalizing constant given by

$$Z = \sum_{x \in \Theta} exp(-U(\mathbf{x})) \tag{4}$$

C is the set of all cliques in the graph. $V_c(x,p), p = 1,2,...,P$ is the potential function defined over clique c . Clique c is defined as a fully connected subgraph of Γ.

It is assumed that at each region $R_i = R_i(p), p = 1,2,...,P$, there are n_i number of pixels in a channel and thus Pn_i observations of a multivariate random variable $Y_i = [Y_i(1), Y_i(2), ..., Y_i(P)]^T$, describing pixel intensities. Then these observations can be used to provide a likelihood for the label X_i at region $R_i = R_i(p), p = 1,2,...,P$. We assume that these likelihoods of different regions are conditionally independent, that is, given the label x_i the likelihoods of Y_i are assumed to be independent. Using the Bayes theorem one can write the posterior probability distribution of $\mathbf{X}$ given the observation $\mathbf{y}$ as

$$P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}) = \frac{P(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x})P(\mathbf{X} = \mathbf{x})}{\sum_{x' \in \Theta} P(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x}')P(\mathbf{X} = \mathbf{x}')} \tag{5}$$

The Bayesian estimate of the configuration is the one which maximizes the posterior distribution. The r.h.s in (5) follows a Gibbs distribution under the assumption of conditional independence of the likelihoods [14],

$$P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}) = \frac{e^{-U_{ps}(\mathbf{x}|\mathbf{y})}}{Z_{ps}} \tag{6}$$

where

$$\begin{aligned} U_{ps}(\mathbf{x}|\mathbf{y}) &= -ln[\prod_{i=1}^{Q} P(Y_i = y_i|X_i = x_i)P(\mathbf{X} = \mathbf{x})] \qquad (7) \\ &= -ln[\prod_{i=1}^{Q} P(Y_i(1) = y_i(1), ..., Y_i(P) = y_i(P)|X_i = x_i)P(\mathbf{X} = \mathbf{x})] \qquad (8) \end{aligned}$$

using the relation $x = e^{lnx}$. This enables us to express the posterior energy function as a sum of the potentials at different points in the image lattice [13, 14]. The term $Z_{ps} = \sum_{x \in \Theta} e^{(-U_{ps}(\mathbf{x}|\mathbf{y}))}$ is the normalizing constant. The maximum posterior estimate $\widehat{x}$ of $\mathbf{x}$ is obtained by minimizing the posterior energy function $U_{ps}(\mathbf{x}|\mathbf{y})$. 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. As said earlier, clique potential functions are defined by incorporating the criteria for optimal segmentation. The energy function is so designed that it takes into account the initial oversegmented characteristics yielding a very high-energy state. Subsequently, depending upon the measure of uniformity of the two adjacent regions in a clique potential function as well as the pattern of Dempster-Shafer labelling of the pixels in these two regions as described in next section, it is decided whether to merge them or not.

3 Evidential Approach for Multisource Data Analysis.

Problems and approach proposed

We consider N separate data sensors(/sources), each providing a measurement $y_s, s = 1, 2, ..., V$, for a pixel of interest, where $V = n_R \times n_C$. 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. It is also important to note that this label set is exclusive and exhaustive (possibly with ω_{k+1} as unclassified class). In order to devise any methodology for multisensor or multisource data analysis we first note that there will be a variety in the range of data types that might be encountered. We consider here pixel specific numerical data after appropriately co-aligning the pixels arising out of different sensors. Other non-numerical pixel specific data (such as a map of labels) may also be considered with this methodology after suitable conversion in an appropriate stage. For example, a suitable mass function may be determined for non-numerical data and may be combined as described later in this section. As is well known the mathematical theory of evidence or Dempster-Shafer (DS) theory [34, 27] 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. We consider the following general situations that are usually encountered in practice for multisensor data analysis and where DS fusion is adopted to provide a joint inference concerning the labelling of the pixel.

First, one of the sensors is very noisy. Probabilistic analysis of such data may yield sensible results if parameters of the models are estimated appropriately. Several methods have been proposed in the literature to obtain parameter estimation with imperfect observations. In such situations, one may follow methods such as the Chalmond's Gibbsian EM algorithm (a pseudolikelihood based algorithm) and the Younes algorithm (a stochastic gradient based algorithm) (see [18]). When increased noise makes parameter estimation more difficult, it may be fruitful to use the theory of evidence. Secondly, some sensor's probabilistic model is appropriately considered but in spots the gray values are hidden by clouds and thus there exists multiple options for labelling of that pixel contrary to the basic principle of mutually exclusive labelling. Thirdly, when the source of prior knowledge (if any) provides information from which the landcover class may be derived, the information is not single valued but in a range, i.e. the prior information has an upper and lower specification.

Consider a simple situation where two images, a SAR and an optical (visible/ infrared) image are acquired by two different sensors on a cloudy day over the same site comprising three classes (groundtruth) only, viz., $\omega_j, j = 1, 2, 3$. As a consequence of the first phenomenon, we observe an overlap in the SAR data distributions of adjacent objects(regions) belonging to certain specific classes [12], viz., different vegetated areas or different types of fallow due to their high standard deviations. Distinct true state of nature (clusters) at times remains indistinguishable by ambiguous response of the sensor. As a result, adjacent regions (clusters), belonging to distinct classes ω_1, ω_2 and ω_3 are likely to exhibit indistinguishable radar response. Due to the second phenomenon, we have no information about the classes on that cloud covered pixel and thus encounter an uncertainty of $\omega_1 \cup \omega_2 \cup \omega_3$ for the optical image.

With regard to the third phenomenon, if our prior estimates for ω_1 is between 20% and 30%, for ω_2 between 30% and 40% and for ω_3 is 40% then all these prior knowledge lead us to consider the prior estimate of the class $\omega_1 \cup \omega_2$ to be 10% which amounts to be an ignorance in the labelling of the pixel.

In evidential theory, a preliminary labelling or impression is carried out for each of the sensors (sources) based on the available data vector (whether it has numerical, nominal or mixed entries). The preliminary labelling or impression for each of the sensors (sources) is then quantified by assigning a mass of evidence to each of the likely labelling propositions for the pixel vis-a-vis accommodating uncertainty in what the pixel label might be. The prime use of DS theory is in modelling imprecisions in each of the sensors and combining their evidences to compute an appropiate label for the pixels. Imprecisions are modelled by defining suitable mass functions for each of the labelling propositions and the idea is to assign a larger mass to the propositions that have lesser imprecision. This may depend on the source itself. For example, Le Hegarat-Mascle et al. [20] assume a global ignorance factor for radar images and a total ignorance for cloud covered pixels in optical images. We propose to model the ignorance using a uniform statistical framework, as explained below.

Basis of DS theory

Let us assume that there are N sensors. For each of the N sensors we thus have N mass functions $M_i, i = 1, 2, ..., N$. These functions have the following characteristics,

$$(i) M_i(\phi) = 0 \quad (ii) \sum_{A \in 2^\Omega} M_i(A) = 1 \tag{9}$$

where ϕ is an empty set meaning thereby a null proposition, Ω is the set of proposition 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) are as follows. For sensor i,

$$Pls(A) = \sum_{B \cap A \neq \phi} M_i(B) \tag{10}$$

$$Bel(A) = \sum_{B \subseteq A} M_i(B) \tag{11}$$

These two functions are the essence of the DS theory, with the help of which both imprecision and uncertainty can be modelled. $Pls(A)$ and $Bel(A)$ are also called the upper and the lower probability for the proposition A respectively. Thus the Bel and Pls functions for the 3rd situation cited above are as follows

$Bel(\omega_1) = 0.2$, $Pls(\omega_1) = 0.3$; $Bel(\omega_2) = 0.3$, $Pls(\omega_2) = 0.4$; $Bel(\omega_3) = 0.4$, $Pls(\omega_3) = 0.4$

This implies that $M(\omega_1) = 0.2$, $M(\omega_2) = 0.3$, $M(\omega_3) = 0.4$, $M(\omega_1 \cup \omega_2) = 0.1$, $M(\omega_1 \cup \omega_3) = M(\omega_2 \cup \omega_3) = M(\omega_1 \cup \omega_2 \cup \omega_3) = 0$.

For each of the sensors we can have a mass function, a Bel function and a Pls function. 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 [34] and is given by the combined mass function $M = M_1 \oplus M_2 \oplus .... \oplus M_N$ as follows:

$$M(\phi) = 0 \tag{12}$$

$$M(A) = \frac{\sum_{\mathrm{B_1 \cap \ldots \cap B_N = A}} \left(\prod_{\mathrm{1 \leq i \leq N}} \mathrm{M_i(B_i)} \right)}{\mathrm{1 - L}} \tag{13}$$

where

$$L = \sum_{B_1 \cap \ldots \cap B_N = \phi} \left(\prod_{1 \leq i \leq N} M_i(B_i) \right) \tag{14}$$

The factor L ($\neq 1$) measures the extent of conflict among various sensors. This orthogonal sum treats the data from different sensors independently. Having computed the combined mass function, we can easily compute the plausibility and belief values with the help of eqn(10) and eqn(11). The latter two functions thus also give the measure of uncertainty values from the combined evidence of the different sensors.

Usually, for making a decision on pixel labelling, an evidential interval bounded by Bel and Pls is attached to every candidate proposition(class label) by the following options

(i) a maximum belief rule, that is the proposition for which the highest belief is obtained,

(ii) a maximum plausibility rule, that is the proposition for which the highest plausibility is obtained and

(iii) an absolute rule, the maximum of belief over all other plausibilities(for all other labels), that is, $Bel(A) \geq max_{A' \in \Omega, A' \neq A} Pls(A')$. It may be noted that the theory of evidence is a kind of extension of the classical theory of probability in a continuous manner.

In remote sensing landcover classification the union of two or more labelling propositions is of little interest as the classes considered are usually distinct. Considering the compound labelling propositions leads to better ignorance modelling. However, removing ambiguity from the compound labelling propositions usually requires the use of spatial information. Since our aim is to investigate the role of feature level fusion in MRF context, compound labelling propositions are ignored as that would perhaps not lead to significant increase in classification accuracy as spatial information is already being incorporated using MRF theory. 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. However, our decision making process will be governed by either rule (i) or rule (ii).

Mass Function Derivation.

In image processing the widely used mass functions are usually derived from the probabilities at pixel level [27]. As we consider pixel specific numerical data from different sensors that are partially redundant and partially complementary, the approach of computing mass function from class conditional probabilities is adopted here. For the case of partially redundant sensors, which means that they are conditionally dependent on the classes, one may also follow the methods due to Pieczynski [28]. Now assuming that the information classes (true state of the nature) for each of the sensors has a correspondence with spectral(data) classes, derivation of our mass functions is based on class conditional probabilities. Thus, the mass functions are derived from the source specific labelling function, that is with posterior probabilities used in conventional statistical classification techniques. For simplicity, we consider equal prior probabilities for each of the K classes and thus the class conditional density functions become posterior probabilities for different classes. The density functions are estimated from the ground truth when it is available and by a mixture estimation algorithm as given in [9], when it is not. The mass functions considered here ensure that the sources which have a greater imprecision have relatively low masses for the singleton classes. We do not assume any *apriori* ignorance factor and model the ignorance using non-null mass functions for the compound hypotheses.

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. Thus, after assuming an appropriate probability law the parameters are estimated from the ground truth sample intensity values.

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 sth pixel, $i^* \in \Omega$, if

$$f_{i^*}^j(y_s) = max_{1 \le i \le K} \quad f_i^j(y_s) \tag{15}$$

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(16) below. We note here that the different members in the power set denote the compound labelling proposition for the pixel. Such a labelling proposition may arise in various situations, for example, when in an optical image some clusters of pixels are hidden by clouds. The unlabelled pixels will now 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:

$$M_j^{(s)}(\Omega_i) = \frac{f_i^j(y_s^j)}{\sum_{q=1}^{t} f_q^j(y_s^j)} \text{ needs to be changed} \tag{16}$$

where y_s^j is the intensity value of the sth pixel for jth sensor and $t = 2^K - 1$. These mass functions $M_1, M_2, ..., M_N$ are the probabilities on Ω^*. The non-null mass functions for compound hypotheses entails uncertainty that the classes are not distinguishable for some cluster of pixels. While computing probabilities of the combined hypotheses we use the simple elementary rule of probability of union of events for singleton classes. Such a probability assignment (see [19]) for the combined hypotheses is perhaps better than the zero assignment of such events as is usually followed. 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. Note that this approach is different from the classical Bayesian approach of multiplying the independent source mass functions for each class and assigning the pixel label the class with the maximum product. If the mass functions of two independent sources are such that one of them has a high value and the other a very low value for a particular class, then their product becomes relatively small as compared to the situation where both mass functions have a moderate value for another class. In such a situation DS theory as applied here gives more weightage to the formaer case, contrary to the Bayesian approach. For example consider a situation when there are 2 sources (indicated by subscripts 1 and 2) and 3 classes (a, b, c). Let the mass values be as follows :-

$$m_1(a) = 0.1, m_1(b) = 0.5, m_1(c) = 0.4$$
$$m_2(a) = 0.8, m_2(b) = 0.2, m_2(c) = 0$$

In this case the Bayesian approach will clearly select the class b as the label. However, using the above definition of masses of compound hypothesis, DS theory selects the class a as the label. As shown above, eqn(5) enables one to aggregate different pieces of information as outlined by the Dempster-Shafer combination rule. 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

$$M^s(\{\omega_{i^*}\}) = max_{1 \leq i \leq K} M^s(\{\omega_i\}) \quad (17)$$

where $M^s(\{\omega_i\})$ is given by eqn(13).

The DS labelled output image thus obtained has been labelled with propositions ω_j ,$j = 1, 2, ..., K$ exploiting the DS theory based on a per pixel approach. Eqn(17) supposes that no ambiguous classes exist, that is, the mass of any compound hypothesis is considered zero during the labelling process. We note that in such a labelling scheme the spatial context (image space) is not considered and clustering (labelling) is performed in the feature space to result in a direct classification. In order to accommodate spatial interactions between pixels we adopt the hidden Markov random field model in the Bayesian context (described in section II) to segment the images of the different sensors.

Thus, our objective being a study of the role of feature level fusion using DS rule and that of data level fusion in the MRF context, we have used here a preliminary per pixel classification method based on a maximum likelihood classification approach and aggregated evidences as described above. Subsequently, in the following sections we attempt to determine the optimal segmentation based on a maximum a posterior probability(MAP) estimate, using an energy function that takes both the above aspects of fusion into account. Finally, the segments are labelled on the basis of groundtruth classes and the classification accuracy is examined as described below.

4 Segmentation scheme based on MRF model

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 energy function has two components, one corresponding to within-region sum of squares representing uniformity or homogeneity of regions, say H and the other corresponding to between-region sum of squares representing dissimilarity between regions, say B. The two components of the energy fuction as per the two optimization criteria are denoted as the *region process*(H) and the *edge process*(B) respectively. The formulation of the energy function with these two processes is described below. The observed data of intensities, that is, $\mathbf{y}$ values are used for incorporating the constraints. Usually constraints are expressed in terms of the prior knowledge by defining suitable parameters for the prior distribution. But in the case of natural scenes one rarely has any such detailed prior knowledge and as such the observed data on intensities are used for incorporating constraints.

Let $\mathbf{M}_1, \mathbf{M}_2, ..., \mathbf{M}_Q$ represent the mean intensity vectors of the initially segmented regions where each $\mathbf{M}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{Y}_{ik}$,$\mathbf{Y}_{ik} = Y_{ik}(1), Y_{ik}(2), ..., Y_{ik}(P)$ 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,

$$\mathbf{S}_i = \sum_{k=1}^{n_i} (\mathbf{Y}_{ik} - \mathbf{M}_i)(\mathbf{Y}_{ik} - \mathbf{M}_i)^{'} \tag{18}$$

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 [22] 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 [31] can be defined as

$$V_c(\mathbf{x}|\mathrm{H}) = (\eta_{ij}/\nu_{ij})[\sum_{k=1}^{n_i} (\mathbf{Y}_{ik} - \mathbf{M}_i)(\mathbf{Y}_{ik} - \mathbf{M}_i)^{'} + \sum_{k=1}^{n_j} (\mathbf{Y}_{jk} - \mathbf{M}_j)(\mathbf{Y}_{jk} - \mathbf{M}_j)^{'}] \tag{19}$$

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. That is , 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 pattern of DS labelling is considered to be same for these two regions. The second condition is that regions are homogeneous with respect to the multisensor pixel intensity values. The test of homogeneity of regions in the clique has been described below. 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

$$\mathbf{S}_{ij} = \mathbf{S}_i + \mathbf{S}_j + (\mathbf{M}_i - \mathbf{M}_j)(\mathbf{M}_i - \mathbf{M}_j)' \frac{n_i n_j}{n_i + n_j} \tag{20}$$

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 (see [31])

$$V_c(\mathbf{x}|\mathrm{H,B}) = V_c(\mathbf{x}|\mathbf{y}) = (\eta_{ij}/\nu_{ij})[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)' + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j)(\mathbf{Y}_{jk} - \mathbf{M}_j)'] \tag{21}$$

$$+\theta_{ij}(1 - \eta_{ij})\frac{n_i n_j}{n_i + n_j}(\mathbf{M}_i - \mathbf{M}_j)(\mathbf{M}_i - \mathbf{M}_j)'$$

The parameter θ_{ij} controls the weight to be given to the two processes for regions R_i and R_j involved in the clique c. For convenience let us rewrite the above equation as

$$V_c(\mathbf{x}|H, B) = \eta_{ij}\mathbf{W}_{i,j} + \theta_{ij}(1 - \eta_{ij})\mathbf{B}_{ij} \tag{22}$$

where

$$\mathbf{B}_{ij} = \frac{n_i n_j}{n_i + n_j}(\mathbf{M}_i - \mathbf{M}_j) * (\mathbf{M}_i - \mathbf{M}_j)' \tag{23}$$

$$\mathbf{W}_{ij} = \frac{1}{\nu_{ij}}[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)' + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j) * (\mathbf{Y}_{jk} - \mathbf{M}_j)'] \tag{24}$$

Merging Regions Criteria

The clique potential will be equal to $\mathbf{W}_{ij}$ if the two regions R_i and R_j have been assigned the same region label and will be equal to $\mathbf{B}_{ij}$ otherwise. 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

$$T^2 = (\mathbf{M}_i - \mathbf{M}_j)'[(1/n_i + 1/n_j)\mathbf{s}_{pooled}]^{-1}(\mathbf{M}_i - \mathbf{M}_j) \tag{25}$$

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 [22].

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 $P[T^2 > F_\alpha] = \alpha$ [see details in [31, p. 1106]]. This decision criterion describes the test of homogeneity that was alluded to previously. Here the probability level or the level of significance α is taken to be 0.0001 as the upper percentage points of F distribution, although any $\alpha \leq 0.05$ may be selected for this purpose as per usual statistical test procedures. The implicit meaning for this α, is that if the two regions R_i and R_j in the clique come from the same multivariate population (implying that they should not have been kept separate), then the chance that the appropriate characteristics based on the intensity values of the pixels of these two regions will be greater than the statistical threshold value, is α. This statistical threshold value is called F_α. Smaller the value of α, lesser is the error (type-I) in the decision of statistical test procedure.

Optimization Procedure

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

$\Delta V_{ci} = \sum_j (\Delta V_{cij})$ with its neighbours j, where

$$\Delta V_{cij} = \sum_{p=1}^{P} s_{pp(i)} + \sum_{p=1}^{P} s_{pp(j)} - \sum_{p=1}^{P} W_{pp(ij)} \tag{26}$$

$W_{pp(ij)}$ are the diagonal elements of $\mathbf{W}_{ij}$, and $\sum_{p=1}^{P} s_{pp(i)}$ and $\sum_{p=1}^{P} s_{pp(j)}$ are the generalized variances of the regions R_i and R_j respectively. For a given region R_i, an aggregate clique potential ΔV_{ci}, is computed by summing ΔV_{cij} over all those neighbours j which satisfy the condition $T_{ij}^2 < F_\alpha$. With such a restriction, the possible global energy reduction for a node i is now limited to few of its neighbors. The process then selects that region R_j for which ΔV_{cij} is maximum for a given ΔV_{ci} and reduce the global energy by merging region R_j with region R_i. Such a merging reduces the global energy by an amount that equals ΔV_{cij}. This amount is the maximum possible reduction in the global energy for that node i, resulting in a new configuration. Eventually, we attain a stage where no nodes can be merged, indicated by the condition that no two adjacent nodes satisfy $T_{ij}^2 < F_\alpha$.

The ultimate aim of a MRF based approach would be to reach a configuration for which the global energy is minimum. The approach of minimization that is considered here is similar to a "greedy approach for minimization" combined with a statistical test on the clique potential. It is difficult to verify whether the approach proposed in this paper produces the global minimum but it yields good results [31] in practice. This technique of labelling the regions has been carried out by considering both data level fusion in image space and feature level fusion through pixel based DS combination rule. The segmented image so obtained is by minimizing the energy function $U_{ps}(\mathbf{x}|\text{H,B}) = \sum_{c\epsilon C} V_c(\mathbf{x}|\text{H,B}) = \sum_{c\epsilon C} V_c(\mathbf{x}|\mathbf{y})$ as described above. The flowchart of the methodology is depicted in Fig.1. We also note that unlike the segmentation algorithm of Beaulieu and Goldberg [4] which also deals with merging of similar clusters in a hierarchical structure starting with one pixel each, the present

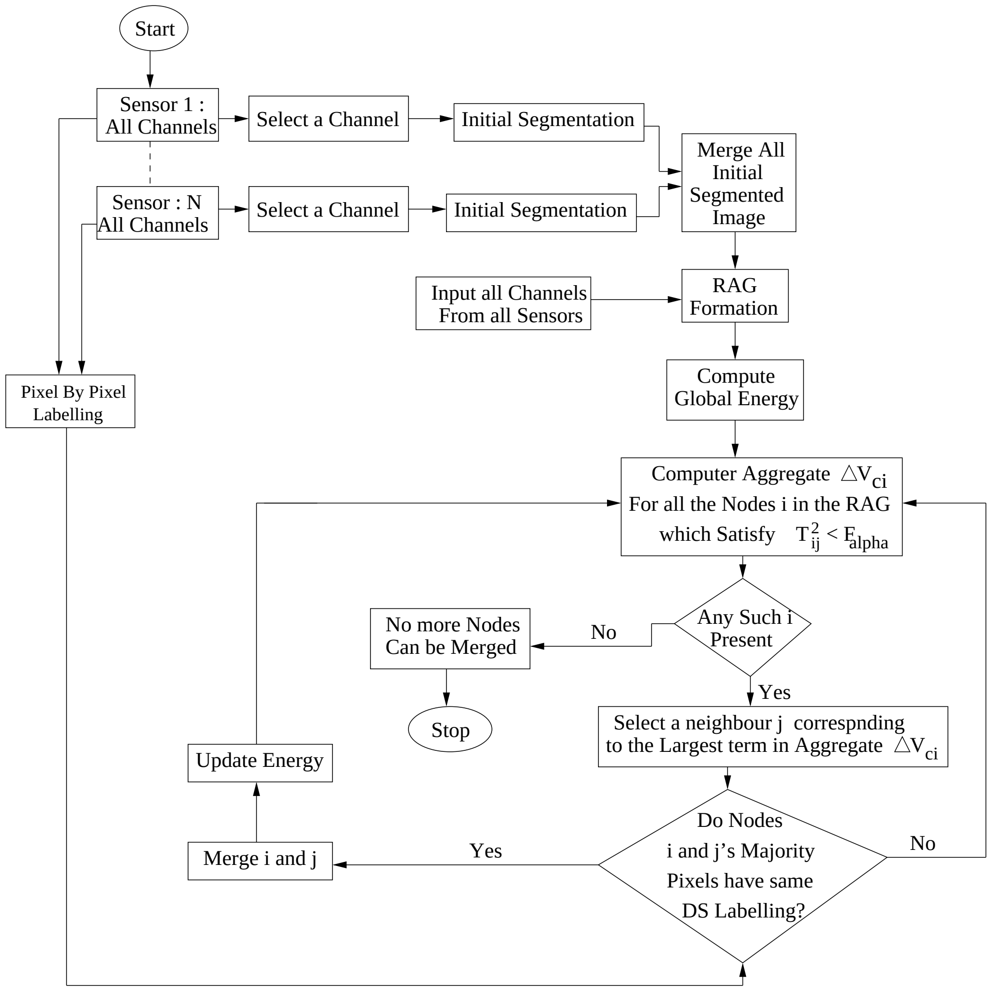


Figure 1: Multisensor Image Segmentation Scheme

work is defined in a MRF framework with initial segmented regions as input to MRF so as to capture the intrinsic characters of tonal regions (see details [31]). The merging of adjacent regions in the RAG that arises here is consistent with the energy minimization process. The use of the statistical test for similarity of the regions expedites the minimization process.

Cluster Validation Scheme

In order to validate the segments of the optimal segmented image we first identify at least one cluster(segment) for each of the groundtruth (training set) classes by following the segment labelling technique described in the first stage of the cluster validation scheme of Sarkar et.al. [31] and label the unlabelled segments accordingly as per the groundtruth class. To do the labelling, we use Fisher's method [22] for discriminating among K ground truth classes. With this procedure, we first compute

$$\mathbf{B_0} = \sum_{\mathbf{i=1}}^{\mathbf{K}} (\mu_{\mathbf{i}}{}^{\mathbf{g}} - \overline{\mu})(\mu_{\mathbf{i}}{}^{\mathbf{g}} - \overline{\mu})' \text{ where } \overline{\mu} = \frac{\mathbf{1}}{\mathbf{K}} \sum_{\mathbf{i=1}}^{\mathbf{K}} \mu_{\mathbf{i}}{}^{\mathbf{g}} \tag{27}$$

$\mu_{\mathbf{i}}{}^{\mathbf{g}}$ being the mean vector of the groundtruth class ω_i, and $\mathbf{W} = \sum_{i=1}^{K}(n_i - 1)\mathbf{s}_i^g$.
The product of the matrices $\mathbf{W}^{-1}$ and $\mathbf{B}_0$ is computed and the eigen vectors of $\mathbf{W}^{-1}\mathbf{B}_0$ are evaluated. We select only those eigen vectors for which the eigen values of $\mathbf{W}^{-1}\mathbf{B}_0$ are positive. These eigen vectors are then normalized that is, the eigen vectors $\mathbf{l}$ are obtained such that $\mathbf{l}\frac{\mathbf{W}}{\sum_{i=1}^{K}(n_i-1)}\mathbf{l}'$=1. Now for each unlabelled cluster R_j we compute the squared distance

$$\sum_{i=1}^{z} l_i^{'}(\mu_j^u - \mu_k^g)^2 \quad \text{for } k = 1, 2, ..., K$$

where $z \leq P$, is the number of positive eigen values of $\mathbf{W}^{-1}\mathbf{B}_0$. The cluster R_j is classified to that class k for which the squared distance or the score is minimum. We note that the conventional approaches for cluster validation viz,. the classical Bayesian method requires an assumption for the joint distribution of SAR and optical images. In the absence of any such assumption it is sensible to use a non-parametric method such as above.

5 Experimental Results

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 viz., channel-1: (0.52-0.59), channel-2: (0.62-0.68), channel-3: (0.77-0.86) and channel-4: (1.55-1.70) in (μm) satellite image and a SAR(Synthetic Aperture Radar) image of the same site. For Subscene 1 the specifications for the optical images (acquired from IRS-P6, ISRO) are - Pixel spacing : 12.5m, Resolution : 23.5m and for the SAR image (acquired from RADARSAT) are - Pixel spacing : 8m, Resolution : 30m; C-Band 5.3 GigaHertz,and HH Polarization. For Subscene 2, the specifications for the optical images (acquired from IRS-1D, ISRO) are : Pixel spacing : 24m, Resolution : 23.5m and for the SAR image (acquired from RADARSAT) are - Pixel : spacing 25m, Resolution : 50m; C-Band 5.3 GigaHertz,and HH Polarization. The optical and the SAR images considered here for Subscene 1, have been acquired on Janury 13, 2004 and Februray 12, 2004 respectively with a time lag of 30 days. For Subscene 2, the dates of acquisition of the were 19 January 2000 and 30 September 1999 for the optical and the SAR image respectively, thus having a time lag of 110 days.

Example-1

This refers to Subscene 1 and is of size 1016×1485 where the former number denotes the number of columns and the later the number of rows. The number of classes identified in groundtruth collection is found to be 14. Most of the classes are in level-2 details (of Anderson et.al classification system [2]) and, in particular, agriculture classes are in level-3.

Example-2

This refers to Subscene 2 and is of size 707×908. We consider these images as one of our test cases to see the effect of lower resolution SAR image in this methodology with respect to the classified images. The number of groundtruth classes is found to be 16. In both the examples the SAR image was registered with the geocoded optical image.

Both the examples cited here are from the same area and having some common points of lattitude and longitude. The different landcover classes involved in the two subscenes are as follows. The total number of groundtruth pixels for each class are given in brackets. For Example 1, the classes are 1. urban (9422), 2. trees with close canopy (5639), 3. deep water (1287), 4. shallow water (347), 5. fallow with paddy stubble (18280), 6. fallow with no crop (4974), 7. irrigated fallow (8571), 8. sand (1606), 9. homestead (1566), 10. laterite (12190), 11. fallow with grass (8796), 12. mustard (614), 13. potato (206) and 14. trees with open canopy (1763). For Example 2, the 16 different classes are 1. deep water (150), 2. shallow water (110), 3. sand (444), 4. urban concrete (4843), 5. homestead (3667), 6. fallow with paddy stubble (2432), 7. irrigated fallow (1296), 8. moist fallow (524), 9. fallow with pet grass (204), 10. fallow with different type of grass (4624), 11. current fallow with no crop (2354), 12. crop-1 (134), 13. crop-2 (293), 14. open land (1584), 15. trees with close canopy (165) and 16. trees with open canopy (720). The classes were identified by experts in the field after the optical sensor data was collected. The total available groundtruth samples (which equals about 5 % of the Example 1's total pixels for 14 groundtruth classes and about 3.7% of that of Example 2 for 16 classes) are divided into two subsamples. The first subsample (about 1.5% for Example 1 and about 1% for Example 2) is used as a training set. That is, with these groundtruth observations, parameters of the class conditional densities are estimated and some of the clusters are labelled as in [31] after obtaining optimal segmentation. Subsequently, the remaining clusters are labelled with the help of these labelled clusters using Fisher's discriminant scores. The second subsample (of 3.5% for the first subscene and of 2.8% for the second) is used as a test set. That is, these are used for the quantitative evaluation of the classification accuracy after all clusters are validated.

The measurements from different sensors is assumed to be conditionally independent [33]. The SAR data provided to us is preprocessed and is made speckle free by the use of enhanced Lee filter of 7×7 window size due to Lopes et al.[26]. The probability density function(pdf) of the SAR intensity distribution after it is made speckle free has been considered to be Gaussian [12]. Infact, the assumption has been investigated from crudely collected

SAR image data for each of the groundtruth classes and is found to be satisfactory for most of the classes except the classes viz., urban, deep water and shallow water. The classwise crude data of SAR sensor have been obtained with the help of optical sensor images i.e. first a classified image has been obtained with the help of 4 channels of optical images following [31]. The same regions are then grown in the SAR image corresponding to the classwise segmented output and are then used for the above purpose. The pdf of an optical image with 4 channels(bands) is considered to be multivariate Gaussian, that is $N(\mu, \mathbf{\Sigma})$ where μ is a 4-component vector and $\mathbf{\Sigma}$ is a 4$\times$4 covariance matrix. We investigate the proposed methodology in three different ways as follows.

Case (i): Initial segmentation is first performed in each of the sensor's selected channel. Since channel-2 accounts for maximum amount of spectral variation for the classes among the four channels of the optical image, it is selected for initial segmentation along with the SAR image. These initial segmented images, one on channel-2 of optical image and the other on SAR image, are then merged as described in section II. On the other hand as described in section III, the mass functions of the respective sensors are computed by eqn(16) and the aggregate evidence of the different sensors are obtained with eqn(13). These aggregate evidences of the two sensors are then incorporated into data level fusion in image space (spatial context) through the energy minimization process associated with the underlying MRF model to obtain an optimal segmentation, as described in section IV. Finally, a cluster validation scheme is applied to this segmented image.

Case (ii): In this case optimal segmentation is performed in the same way as in case(i) but segments are labelled with thematic(DS) labels directly, contrary to the separate set of labels used in optimal clusters(segments) for case(i). Thus, unlike case(i) no cluster validation of segments is required after obtaining optimal segmentation.

Case (iii): As in case(i), MRF based segmentation is performed but no DS features are incorporated in the energy function. That is to say, only data-level fusion is performed in this case.

In each of the above cases, corresponding to these subscenes, the classification accuracies are then obtained with test sets. For the sake of comparison, we have investigated the approach of Tupin et al.[37] as the following case.

Case (iv): In this approach a direct classification is done on the initial segmented regions of the RAG using the DS rule. Unlike the proposed methodology(case(i)) where a separate set of labels is used for labelling the RAG, the regions are labelled here from the set of thematic (groundtruth) classes like case(ii). A MRF is defined on the RAG. The energy function associated with this MRF has two components, viz., a region-wise mass function and a clique-wise potential function. For a particular class assignment, the region mass is the orthogonal sum of DS mass functions of a region calculated for each sensor for the particular class assignment, based on the groundtruth data. The clique potential is a measure of the favorability of the class assignment on the two adjacent regions belonging to a clique based on supervised knowledge of the classes. The global energy of the MRF is minimized by using the simulated annealing approach.

We have also compared the proposed method with two other nonparametric multisensor fusion methods (both of which are based on neural network approaches). These are

Case (v): Multilayer Perceptron [17]

Case (vi): Radial Basis functions (RBF) [21].

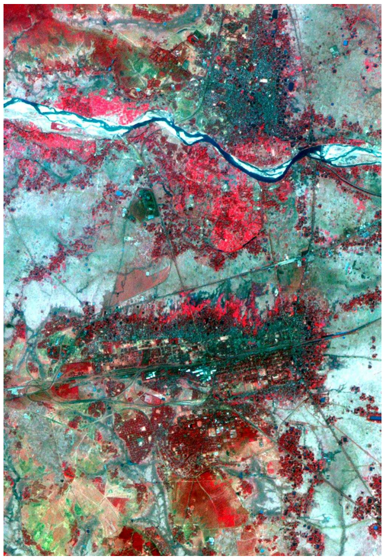
A comparison of classification accuracies of the various cases for both the examples are presented in Table I. This table provides normalized classification accuracies and identifies the cases (indicated by $*$) over which case(i) exhibits significant differences in accuracy determination (by Kappa coefficient). Table I also exhibits CPU time for the different cases considered here in a Pentium IV system with 1.7 GHz and 1024MB RAM. Table II exhibits intermediate results of the proposed methodology (Case(i)). The original images and the validated segments of these examples are shown respectively in Fig.2 and Fig.3. Fig 2(a) and Fig.2(b) exhibit the orginal optical images and the SAR image respectively for Example 1. In Fig.2(c), different shades of colours represent different groundtruth classes for Example 1. Fig.3(a) and Fig.3(b) exhibit respectively one channel optical image and the SAR image for Example 2.

Our indices of comparison of various cases are normalized accuracy and Kappa coefficients. With the former measure, a direct comparison is made and with the latter significance tests are performed for case(i) with other cases. Although the normalization process of the confusion(error) matrix provides a convenient way of comparing different error matrices, the Kappa coefficient is another discrete multivariate technique in use for accuracy assessment. The performance of the proposed method has been compared with that of the other cases with respect to Kappa coefficient to determine whether these are statistically significant [8, 36]. The test results show that the proposed method(case(i)) has been found to be significantly better then the other methods considered here. In Table I, the value of Kappa coefficients are not exhibited, instead we use a symbol * on normalized accuracies to indicate the methods over which the proposed method is found to be significantly better with respect to the Kappa coefficient. These two performance measurements are in agreement with the performance findings for all the cases. Although as stated in[8, p. 40] "... depending on the amount of error included in the matrix(confusion), these measures may not agree".

Further, we should take note of the fact that the image data of the two sensors for Example 1 and Example 2 have a time lag of 30 and 110 days respectively. During this time, not many changes have taken place in the subscene of Example 1 but many temporal changes might have taken place in some classes (groundtruth), particularly in different types of fallow and crop classes for the subscene of Example 2. In spite of these temporal changes we note from classification accuracies presented in Table I, that the proposed methodology (Case(i)) has an edge over other cases. We note here that the normalized accuracies for Example 2 are approximately the same, but that Kappa coefficients indicate that there is a significant difference. This is due to the fact that Kappa coefficients are obtained from the original error (confusion) matrices for each of these cases. The overall accuracies from these error matrices are found to be 84.6% and 78.0% respectively. In most cases, due to the probabilistic labelling of pixels in Case(ii), the accuracy obtained is lower than Case(i). The accuracies in Case(iii) is lower than that of Case(i) because of the fact that feature level fusion with the help of DS theory has not been incorporated. The energy functions used in Case(iv) are selected in a somewhat adhoc manner and the fact that the labelling in this case is also probabilistic, may be the reasons for its inferior performance than Case(i). In fact, a better choice of energy functions in Case(iv) may lead to higher accuracies. Case(v) and Case(vi) which are based on Neural Networks and Radial Basis Function Networks respectively, fail to achieve high accuracies compared to Case(i) as they are unable to model the spatial interactions inherent in the image.

The computation time for the proposed method is on the higher side as it uses DS rule with eqn(13) for aggregate evidences. We have used the fast Mobius transform [**?**] to implement the DS rule of combination, but the complexity depends exponentially on the number of classes and linearly on the number of pixels leading to a large difference in the running times of the two exmaples. Further, case(iv) uses simulated annealing for the energy minimization scheme along with DS rule for aggregate evidences and thus its computation time is the highest. Computation time for case(v) and case(vi) are in the lower side and so are the classification accuracies of these nonparametric approaches.

The above analysis corroborates the fact that the proposed methodology (Case(i)) has a potential.



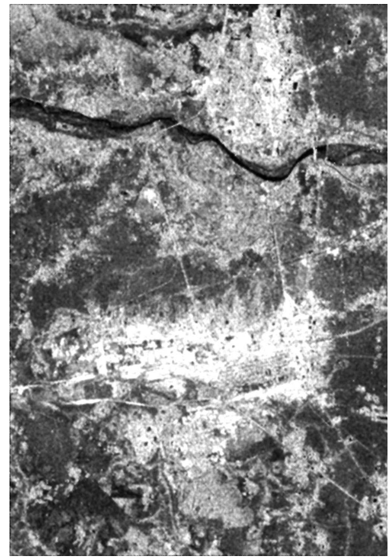


Fig.2(a).Original Image for Optical Sensor(FSS) for Subscene 1

Fig.2(b). : Registered Image of SAR Sensor for Subscene 1

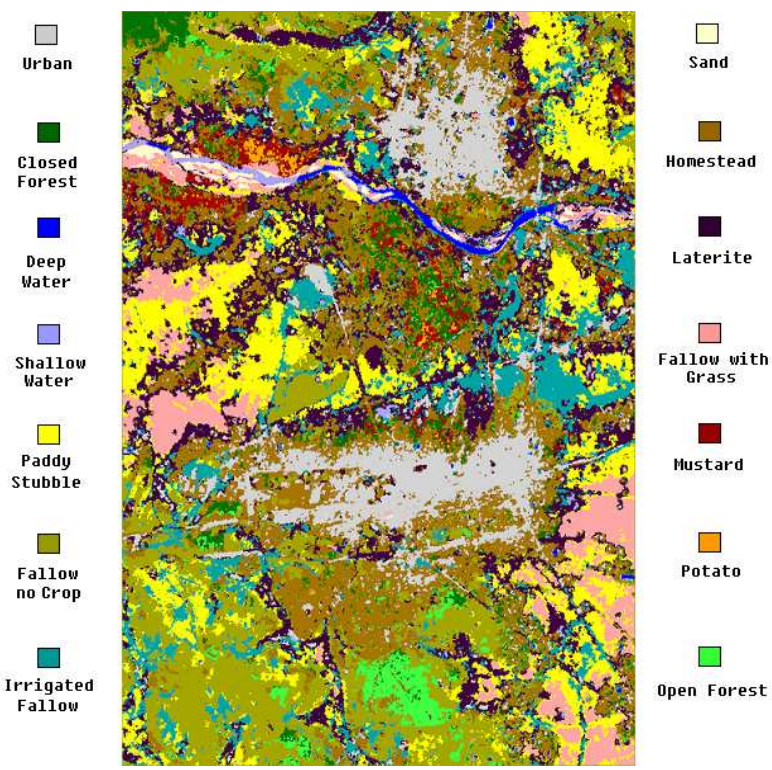


Fig.3: Validated Segments of Fig.2

Fig 4(a).Original Image for
Optical Sensor of Channel 2 for Subscene 2

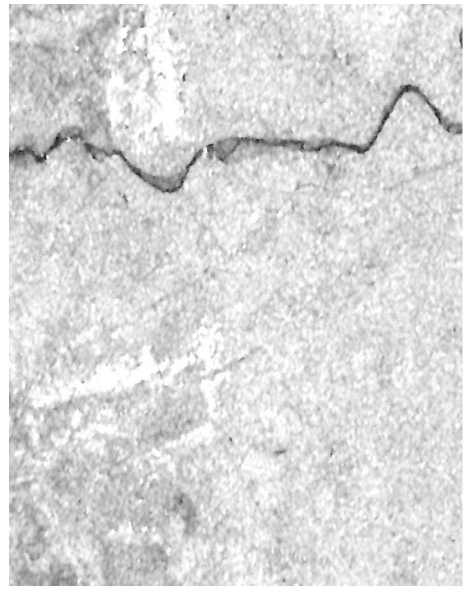


Fig 4(b). : Registered Image of
SAR Sensor for Subscene 2

Table I: Comparison of Accuracy and Time Duration For All the Six Cases of the Sub-scenes

Approaches	Subscene 1		Subscene 2	
	Norm. Accuracy	Time(hh:mm:ss)	Norm. Accuracy	Time(hh:mm:ss)
Case (i)	95.8	06:30:35	88.0	10:17:00
Case (ii)	92.8*	06:30:35	88.1*	10:17:00
Case (iii)	94.2*	00:17:24	83.1*	00:06:00
Case (iv)	90.9*	23:25:14	79.8*	15:21:00
Case (v)	86.1*	00:10:00	80.4*	00:02:30
Case (vi)	86.6*	00:51:00	81.4*	00:07:13

Table II: Intermediate Results for Case-(i) for both Examples

Subscene	Initial Regions	Class Segment	Initial Disagreement	DS Conflict (Value of L in eq.(14))
Subscene 1	435370	128549	75.4 %	0.006
Subscene 2	180768	65715	91.3 %	0.005

6 Conclusion.

The proposed methodology of landcover classification with multisensor data uses DS rule for feature level fusion and exploits it in the MRF context to consider the important element of spatial interactions between pixels. In this paper we have analysed the fusion of data coming from similar sources. The experimental results obtained with the proposed methodology as shown in Table I are encouraging and corroborates the fact that it has a potential This methodology can also be extended for combining data from dissimilar sources by defining suitable mass functions.

References

[1] M. A. Abidi and R. C. Gonzalez, "*Data Fusion in Robotics and Machine Intelligence*". New York, Academic,1992.

[2] J. R. Anderson, E. H. Ernest, J. T. Roach and R. E. Witmer, "A Land Use and Land Cover Classification System for Use with Remote Sensor Data", *USGS Professional Paper # 964*, 1976.

[3] 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. and Remote Sensing*, vol. 39, no. 8, pp:1789 - 1798, Aug. 2001.

[4] J. M. Beaulieu and M.Goldberg, " Hierarchy in Picture Segmentation : A Stepwise Optimization Approach ", *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-11, no.2, 150 - 163, 1989.

[5] J. Besag, "On the Statistical Analysis of Dirty Pictures ",*J.R.Statist.Soc.*, vol.48, pp. 259 - 302, 1986.

[6] I.Bloch," Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account ",*Pattern Recognit. lett.*, vol.17, no. 8, pp. 905-916, 1996.

[7] C.Bouman and B.Liu," Multiple resolution segmentation of texture images ", *IEEE Trans. Pattern Anal. Machine Intell.*, vol.13, 99-113, 1991.

[8] R. G. Congalton," A review of assessing the accuracy of classification of remotely sensed data ",*Remote Sensing Environment*, vol.37, 35-46, 1991.

[9] Y.Delignon, A. Marzouki and W.Pieczynski,“ Estimation of generalized mixtures and its application in image segmentation ”, *IEEE Trans.Image Process.* vol. 6, no. 10, pp.1364-1375, Oct 1997.

[10] H. Derin and W.S. Cole, “ Segmentation of Texture Image using Gibbs Random Fields ”,*Computer Vision Graphics Image Process.*, Vol-35, 72-98, 1986.

[11] H. Derin and H.Elliott, “ Modeling and segmentation of noisy and textured images using Gibbs Random Fields ”, *IEEE Trans.Pattern Analysis Mach.Intell.* vol.9, no.1, 39-55, 1987.

[12] Y.Dong, A.K.Milne and B.C.Forster, “ Segmentation and Classification of Vegetated Areas Using Polarimetric SAR Image Data ”,*IEEE Trans. Geosci. Remote Sensing*, vol.39, no.2, pp 321-329, Feb. 2001.

[13] R.C. Dubes and A.K.Jain, “ Random field models in image analysis ”, *J.Appl.Statist.*, vol.16, 131-164,1989.

[14] R.C. Dubes, A.K.Jain, S.G. Nadabar and C.C. Chen, “ MRF model based algorithms for image segmentation ”, *In Proc. Int. Conf. on Pattern Recognition*, 808-814, 1990.

[15] S.Geman and D.Geman, “ Stochastic relaxation Gibbs distributions, and the Bayesian restoration of images ”,*IEEE Trans.Pattern Anal.Machine Intell.,* PAMI-6, 721-741, 1984.

[16] D.Hall and J.Llinas, “ An Introduction to Multisensor Data Fusion ”, *In Proc. of the IEEE*, Vol.85, no.1 pp. 6-23, Jan, 1997.

[17] S.Haykin, *“ Neural Networks - A Comprehensive Foundation ”*, Macmillan College Publishing Company, New York, 1994.

[18] M.V.Ibanez and A.Simo, “ Parameter estimation in Markov random field image modeling with imperfect observations. A comparative study.”, *Pattern Recog. Letters*, vol 24, pp. 2377-2389, 2003.

[19] 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.

[20] S. Le Hegarat-Mascle, I. Bloch and D. Vidal-Madjar,“ Introduction of neighborhood information in evidence theory and application to data fusion between radar and optical images with partial cloud cover ”, Pattern Recog. vol. 31, 14, pp. 1811-1823. Oct. 2003.

[21] 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.

[22] R A Johnson and D. H. Wichern," *Applied multivariate Statistical Analysis* ", 3rd ed. Prentice Hall of India Pvt. Ltd, New Delhi, 1996.

[23] B. Kartikeyan and A.Sarkar, " A unified approach for image segmentation using exact statistic ",*Computer Vision Graphics Image Process.*,vol.48, pp.217-229, 1989.

[24] I.Y. Kim and H.S.Yang," An Integration Scheme for Image Segmentation and Labelling based on Markov Random Field Model ", *IEEE Trans. Pattern Analysis Mach.Intell*, vol.18, pp.69-73,1996.

[25] S. Krishnamachari and R.Chellappa, " Multiresolution Gauss-Markov Random field models for texture segmentation ", *IEEE Trans.Image Process.* vol.6, no.2, 251-267, 1997.

[26] A.Lopes, R.Touzi and E. Nezry," Adaptive speckle filters and scene heterogeneity ",*IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp.992-1000, Nov. 1990.

[27] 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.

[28] W. Pieczynski, "Unsupervised Dempster-Shafer fusion of dependent sensors ", *Proceedings of the 4th IEEE Southwest Symposium Image Analysis and Interpretation, 2000.(SSIAI 2000)* , 2-4 April 2000,pp. 247-251.

[29] A. Sarkar, M.K.Biswas and K.M.S.Sharma, "A simple unsupervised MRF model based image segmentation approach" , *IEEE Trans.Image Process* vol. 9, pp. 801-812, May 2000.

[30] A. Sarkar, K.M.S.Sharma and R.V.Sonak, " A new approach for subset 2-D AR Model identification for describing textures " , *IEEE Trans.Image Process.* vol.6,pp. 407-413, Mar.1997.

[31] A. Sarkar, MK Biswas, B.Kartikeyan, V.Kumar, K.L.Majumdar 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.

[32] A.H.Schistad Solberg, A.K.Jain and T.Taxt, " Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images "*IEEE Trans. Geosci. Remote Sensing*, vol.32, pp 768-778, July 1994.)

[33] 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.

[34] G. Shafer, " *A Mathematical Theory of evidence.* " Princeton, NJ: Princeton University Press, 1976.

[35] K. M. S. Sharma and A. Sarkar, " A Band Selection technique for classification of multispectral remotely sensed data ", *In Pattern Recognition,Image Processing and Computer Vision-Recent Advances*, P.P.Das and B.N.Chatterji, Eds, New Delhi, India, Narosa, 1996.

[36] K. M. S. Sharma and A. Sarkar," A Modified Contextual Classification Technique for Remote Sensing Data "*Photogramm. Eng. Remote Sens.*, vol. 64 , no.4, pp.273-280, 1998.

[37] 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

[38] C.S.Won and H.Derin," Unsupervised segmentation of noisy and texture images using Markov random field ",*Comput Vision Graphics Image Process.*,vol.54, 308-328,1992.