Application of Dempster–Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing

Sylvie Le Hégarat-Mascle, Isabelle Bloch, and D. Vidal-Madjar

***Abstract*— The aim of this paper is to show that Dempster–Shafer evidence theory may be successfully applied to unsupervised classification in multisource remote sensing. Dempster–Shafer formulation allows to consider unions of classes, and to represent both imprecision and uncertainty, through the definition of belief and plausibility functions. These two functions, derived from mass function, are generally chosen in a supervised way. In this paper, we describe an unsupervised method, based on the comparison of monosource classification results, to select the classes necessary for Dempster–Shafer evidence combination and to define their mass functions. Data fusion is then performed, discarding invalid clusters (e.g., corresponding to conflicting information) thank to an iterative process.**

Unsupervised multisource classification algorithm is applied to MAC-Europe'91 multisensor airborne campaign data collected over the Orgeval French site. Classification results using different combinations of sensors (TMS and AirSAR) or wavelengths (L and C bands) are compared. Performance of data fusion is evaluated in terms of identification of land cover types. The best results are obtained when all three data sets are used. Furthermore, some other combinations of data are tried, and their ability to discriminate between the different land cover types is quantified.

I. Introduction

IMAGES acquired over the same site by different sensors are generally partially redundant, as they represent the same scene, and partially complementary, since the sensors have different characteristics and physical interaction mechanisms are different. For many applications of image classification problems, the information provided by a single sensor is incomplete resulting in misclassification. Fusion with redundant data can help reduce imprecision, and fusion with complementary data can provide a more complete description. In both cases, classification results should be better.

Data fusion may be performed at different stages [1]: pixel, feature, and decision level. At decision level, first monosource classification results are combined in the last step, generally in a supervised approach. At pixel level, many data fusion methods have been proposed. The simplest approach is to concatenate the data from the different sensors as if they were measurements from one single sensor [11]. In that case, establishing a good model for multisource data is the difficult part. More sophisticated methods of statistical multisource classification have been proposed [3], [10], [16], [17]. However, they require knowledge of a considerable amount of information on the measurement physics of sensors, and their applications are usually supervised.

Mathematical theory of evidence was first introduced by Dempster in the 1960's, and later extended by Shafer [18]. This theory, which allows to represent both imprecision and uncertainty, appears as a more flexible and general approach than the Bayesian one. Another of its advantages is its ability to consider not only single (or individual) classes, but also unions of classes. Applications were developed in medical imaging [5], object detection [7], [14], and remote sensing classification [12]. However, even if the authors generally underline the advantages of Dempster–Shafer approach, the studied applications are either theoretical or using small data sets, and we did not find any application of Dempster–Shafer theory to unsupervised classification problem.

In remote sensing applications, the expected number of classes may be large, in particular for agricultural areas. Therefore, an accurate estimation of the class characteristics from the training areas is tedious and time consuming. Moreover, even when supervised methods are able to show the existence of *a priori* unknown classes (for instance, by introduction of a rejection class), they are generally unable to separate these *a priori* unknown classes and to estimate their characteristics. Finally, for some applications, such as the MARS project [6], an unsupervised classification is first performed to identify image classes. Comparison with ground truth is then accomplished. In such a case, the advantage to operate in an unsupervised way is that it allows to identify the actual feature of each of the supervised classes. Therefore, even if unsupervised classifications are generally known to be more computationally expensive (in particular, they generally require a greater number of classes) than supervised classifications, there is an interest to develop unsupervised techniques. The aim of this paper is to propose a unsupervised multisource classification method based on Dempster–Shafer evidence theory.

The data set used for this study was acquired over the Orgeval agricultural site ([13]), located about 20 km east of Paris, during the MAC-Europe'91 campaign. Multifrequency polarimetric radar images and multi-spectral optical images

Manuscript received April 3, 1996; revised September 24, 1996. This work was supported in part by the French Space Agency (CNES) and Matra Cap Systémes company. The participation to the NASA MAC-Europe deployment was made possible by support from the French National Programme for Remote Sensing from Space (PNTS).

S. Le Hégarat-Mascale and D. Vidal-Madjar are with CETP, 78140 Vélizy, France (e-mail: sylvie.mascle@cetp.ipsl.fr).

I. Bloch is with ENST, 75013 Paris, France.

Publisher Item Identifier S 0196-2892(97)03677-2.

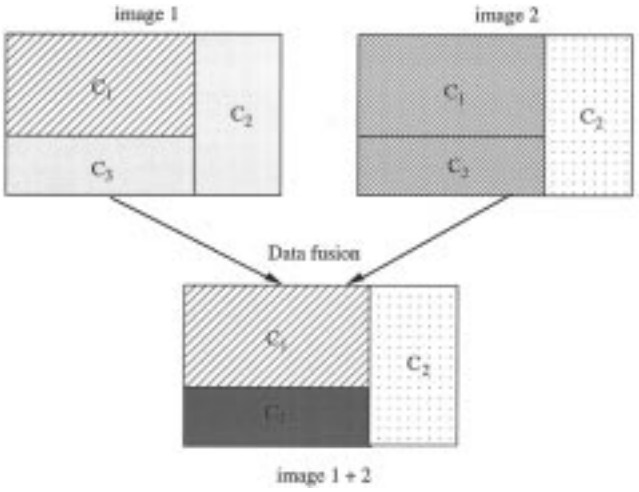

Fig. 1. Three classes example: distinctions between classes in images 1 and 2.

were respectively, recorded by AirSAR [21] and TMS (Thematic Mapper Simulator), both from NASA/JPL.

The remainder of the paper is organized as follows. In Section II, some advantages of Dempster–Shafer approach for data fusion are presented, and a simple example is introduced that will be discussed throughout the paper to illustrate how this theory can be applied to multisource classification. In Section III, main aspects of the Dempster–Shafer evidence theory are summarized. In Section IV, multisource unsupervised classification is discussed. In Section V, classification results, using different combinations of sensors or wavelengths, are analyzed in terms of land cover identification. Section VI is our conclusion.

II. Advantages of Dempster–Shafer Evidence Theory for Data Fusion

Let us first introduce a simple example to illustrate the interest of data fusion: consider the simple case of three classes C_1, C_2, and C_3, and two different sources Σ^1 and Σ^2 which are respectively able to distinguish C_1 from the two other classes but not C_2 from C_3, and C_2 from the two other classes but not C_1 from C_3. Fig. 1 shows a representation of the classes which can be identified in images 1 and 2 (respectively provided by Σ^1 and Σ^2). In this figure, we clearly see that complementary information may be used to detect new classes, such as C_3 (see Fig. 1) which was indistinguishable using image 1 (e.g., Fig. 4 in the application of Section V) alone or image 2 (e.g., Fig. 5) alone, or to obtain a nonambiguous description of the targets when data collected by a single sensor fail to discriminate between two particular targets of interest.

The three main numerical approaches to data fusion are the probabilistic methods, fuzzy set theory, and Dempster–Shafer evidence theory. The main advantage of fuzzy fusion approach is that the fuzzy set framework provides a lot of combination operators, which allows the user to adapt the fusion scheme to the specificity of the data at hand. However, to our knowledge, the operators are always selected in a supervised way.

The main limitation of Bayesian inference is that it cannot model imprecision about uncertainty measurement. An event is said to be uncertain if its probability is not equal to one (or zero, in which case it is certain to be false). However, there may be an imprecision on probability measurement (a measurement can be said imprecise if there is a non null error bar on its value). A first advantage of Dempster–Shafer evidence theory is its ability to deal with ignorance and missing information. In particular, it provides explicit estimations of imprecision and conflict between information from different sources. Indeed, probability theory may be seen as a limit of Dempster–Shafer evidence theory when it is assumed that there is no imprecision, and that only uncertainty has to be taken into account.

Another major advantage of Dempster–Shafer theory is it can deal with any union of classes. This is particularly useful to represent "mixed" pixels in classification problems. Besides the case of mixed pixels due to coarse spatial resolution, mixed pixels are present in monosource image modeling every time the source is unable to distinguish two classes, such as C_2 and C_3 in image 1 of the previous example.

Because of these advantages, we will focus on Dempster–Shafer method in the following sections.

III. Basic Principles of Dempster–Shafer Evidence Theory

We denote Θ the space of hypotheses. In image classification applications, Θ is the set of hypotheses about pixel class. Dempster–Shafer theory allows to consider any subset of Θ. In the following, we denote 2^Θ the set of the subsets of Θ. Applied to classification problems, it means that not only single classes (also called singletons) but also any union of classes can be represented. In the following, hypotheses about singletons and hypotheses about unions of classes are respectively called simple hypotheses and compound hypotheses.

By extension of the notations of the set theory, inclusion, intersection, and union between two hypotheses A and B are defined and denoted as follows:

$$\text{for a given event } x: \quad \forall A \in 2^\Theta, \forall B \in 2^\Theta$$
$$\begin{cases} A \subseteq B \Leftrightarrow \text{if } A \text{ is true, then } B \text{ is true} \\ (A \cap B) \text{ is true} \Leftrightarrow A \text{ is true and } B \text{ is true} \\ (A \cup B) \text{ is true} \Leftrightarrow A \text{ is true or } B \text{ is true.} \end{cases}$$

A. Representation of Evidence

The Dempster–Shafer evidence theory provides a representation of both imprecision and uncertainty through the definition of two functions: plausibility ($\mathbf{Pls}$) and belief ($\mathbf{Bel}$), which are both derived from a mass function ($\mathbf{m}$). $\mathbf{m}$ is defined for every element A of 2^Θ, such that the mass value $\mathbf{m}(A)$ belongs to the [0, 1] interval and

$$\mathbf{m}: \begin{cases} \mathbf{m}(\emptyset) = 0 \\ \displaystyle\sum_{A \in 2^\Theta} \mathbf{m}(A) = 1 \end{cases} \tag{1}$$

where $\emptyset$ is the empty set.

When the mass affected to a compound hypothesis $A \cup B$ is nonzero, it means that we have an option not to make the decision between A or B but rather leave the pixel in the $A \cup B$

class. In particular, assigning a non null mass to Θ allows to not classify some pixels, for which there is a global ignorance.

The belief and plausibility functions, derived from $\mathbf{m}$, are respectively defined from 2^{Θ} to [0, 1]:

$$\mathbf{Bel}(A) = \sum_{B \subseteq A} \mathbf{m}(B) \tag{2}$$

$$\mathbf{Pls}(A) = \sum_{B \cap A \neq \emptyset} \mathbf{m}(B). \tag{3}$$

These two functions, which have been sometimes referred to as lower and upper probability functions, have the following properties:

$$\begin{cases} \mathbf{Bel}(A) \leq \mathbf{Pls}(A) & (4) \\ \mathbf{Pls}(A) = 1 - \mathbf{Bel}(\overline{A}) & (5) \end{cases}$$

where $\overline{A}$ is the complementary hypothesis of A: $A \cup \overline{A} = \Theta$ and $A \cap \overline{A} = \emptyset$.

In the case of Bayes theory, uncertainty about an event is measured by a single value (probability) and imprecision about uncertainty measurement is assumed to be null. In the case of Dempster–Shafer theory, the belief value of hypothesis A may be interpreted as the minimum uncertainty value about A, and its plausibility value, which is also the "unbelief" value of the complementary hypothesis $\overline{A}$ [see (5)], may be interpreted as the maximum uncertainty value of A. Thus, uncertainty about A is represented by the values of the interval [$\mathbf{Bel}(A)$, $\mathbf{Pls}(A)$], which is called the "belief interval" and the length of this belief interval gives a measurement of the imprecision about the uncertainty value.

B. Evidence Combination

Dempster–Shafer theory provides a method to combine the previous measures of evidence of different sources. If $\mathbf{m_i}$ is the basic probability assignment provided by source $\Sigma^i (1 \leq i \leq p)$, the combination: $\mathbf{m} = \mathbf{m_1} \oplus \cdots \oplus \mathbf{m_p}$, also called orthogonal sum, is defined, according to the Dempster's combination rule [18], by

$$\begin{cases} \mathbf{m}(\emptyset) = 0 & \\ \text{if } K \neq 1, \mathbf{m}(A) = \dfrac{\displaystyle\sum_{B_1 \cap \cdots \cap B_p = A} \prod_{1 \leq i \leq p} \mathbf{m_i}(B_i)}{1 - K} & (6) \\ \text{where } K = \displaystyle\sum_{B_1 \cap \cdots \cap B_p = \emptyset} \prod_{1 \leq i \leq p} \mathbf{m_i}(B_i). & (7) \end{cases}$$

From (7), we see that $K (K \in [0, 1])$ represents the mass which would be assigned to the empty set, after combination, in the absence of normalization [division by $(1 - K)$ in (6)]. Thus, K is often interpreted as a measure of conflict between the different sources and it is introduced in (6) as a normalization factor. The larger K is (with $0 \leq K \leq 1$), the more the sources are conflicting and the less sense has their combination. Finally, the orthogonal sum does not exist when K is equal to 1. In this case, the sources are said to be totally or flatly contradictory, and it is no longer possible to combine them.

More details about Dempster–Shafer algebraic properties can be found in [9]. In particular, it is shown that the Dempster's rule of combination is commutative and associative.

C. Decision Making

Having computed the mass, plausibility and belief values for each simple and compound hypothesis of the multisource model, we need a criterion, which is called "decision rule," to decide which hypothesis is the more "realistic." Nowadays, the choice of this criterion remains application dependent. The three most popular decision rules are ([9], [18]): 1) maximum of plausibility, 2) maximum of belief, and 3) maximum of belief without overlapping of belief intervals. Rule 1) is judged as the best by some authors [2]; maximum belief over the simple hypotheses is the most used; rule 3), called absolute decision rule, is very strict. Other rules such as $\max\{\mathbf{Bel}(A) + \mathbf{Pls}(A)\}$ [which may also be written $\max\{\mathbf{Bel}(A) - \mathbf{Bel}(\overline{A})\}$] are compromises.

IV. Application to Unsupervised Multisource Classification

A. Mass Function Definition

There is no general answer to the problem of the mass function definition. In image processing, the most widely used mass functions are directly derived, at the pixel level, from the probabilities [12], [14], or from the distance to class centers [5]. In most cases, no other compound hypothesis but Θ is considered. This approach may be well adapted in cases where the information from the different sources is mainly redundant (same classes in the images), and data fusion is only used to confirm the decision taken from the data of a single source. Here we consider the case where the information provided by the different sources is mainly complementary.

First, we consider the example (presented in Section II and Fig. 1) of three classes C_1, C_2, and C_3. The hypotheses to be considered in Dempster–Shafer formulation are: $\emptyset$ (whose mass is null, according to (1), and therefore, it is no longer considered), simple hypotheses: C_1 (for notation convenience, we denote C_i the hypothesis about the membership of a pixel to class C_i), C_2, C_3, and compound hypotheses: $C_1 \cup C_2$, $C_1 \cup C_3$, $C_2 \cup C_3$, $\Theta = C_1 \cup C_2 \cup C_3$.

If there is no ambiguity between two classes (i.e., we assume there is no ignorance about these two classes), affecting a null mass to their union seems relevant. Conversely, when two classes C_i and C_j are not distinguishable by a sensor, it seems reasonable to give a non null mass to their union $C_i \cup C_j$. For the choice of the mass functions of C_i and C_j, two basic strategies may be chosen:

- affecting a null mass to these two classes: $\mathbf{m}(C_i) = \mathbf{m}(C_j) = 0$, and $\mathbf{m}(C_i \cup C_j) \neq 0$;
- affecting the same mass to these two classes and to their union: $\mathbf{m}(C_i) = \mathbf{m}(C_j) = \mathbf{m}(C_i \cup C_j) \neq 0$.

The former assumes that the ignorance about the membership of a pixel to C_i or C_j is total. In the latter, the mass is arbitrary distributed among C_i, C_j and a part of the ignorance represented by $\mathbf{m}(C_i \cup C_j)$. Therefore, the second strategy is intermediate between the model of total ignorance about C_i or C_j relative to $C_i \cup C_j$ and the Bayesian modeling, where ignorance is assumed to be null: $\mathbf{m}(C_i \cup C_j) = 0$, $\mathbf{m}(C_i) = \mathbf{m}(C_j) \neq 0$. In this paper, we consider the second mass assignment strategy, because experimental results (on our data) show that, in case of the first strategy, the final number of classes (see Section IV-C) is about 50; therefore, the interpretation of the results is tedious, or impossible for the classes which do not have pixels located in the area where ground truth is known.

Finally, the mass functions are normalized such that: $\sum_{A \subset \Theta} \mathbf{m}(A) = 1$ (1). So, in the case of the example shown on Fig. 1, we define the mass functions as in [5]:

$$\begin{cases} \mathbf{m_1}(C_1 \cup C_2) = \mathbf{m_1}(C_1 \cup C_3) = \mathbf{m_1}(\Theta) = 0 \\ \mathbf{m_1}(C_2) = \mathbf{m_1}(C_3) = \mathbf{m_1}(C_2 \cup C_3) = t \\ \mathbf{m_1}(C_1) = 1 - 3t \end{cases} \tag{8}$$

$$\begin{cases} \mathbf{m_2}(C_1 \cup C_2) = \mathbf{m_2}(C_2 \cup C_3) = \mathbf{m_2}(\Theta) = 0 \\ \mathbf{m_2}(C_1) = \mathbf{m_2}(C_3) = \mathbf{m_2}(C_1 \cup C_3) = u \\ \mathbf{m_2}(C_2) = 1 - 3u \end{cases} \tag{9}$$

where $t \in [0, \frac{1}{3}]$ and $u \in [0, \frac{1}{3})]$ have been determined for each pixel during learning process.

This example is very simple as we know *a priori* information about the class discrimination by the two sources. In case of unsupervised classification, we do not know the classes, nor the ability of the sources to recognize them.

In the case of unsupervised classification, the two classes C_1^1 and $C_{2\cup3}^1$ (resp. C_2^2 and $C_{1\cup3}^2$) would be detected on image 1 (resp. 2). As $C_1^1 \cap C_2^2$ would be empty, the simple hypotheses to be considered would be: $C_1^1 \cap C_{1\cup3}^2$, $C_{2\cup3}^1 \cap C_2^2$, and $C_{2\cup3}^1 \cap C_{1\cup3}^2$, which are easily identified as C_1, C_2, and C_3. This example shows a quite simple unsupervised way to define simple hypotheses for Dempster–Shafer data fusion: the set of singletons we will consider is the set of nonempty intersections between classes of different sources. This definition is based on the assumption that the cluster characteristics are reliable enough to avoid overlapping between classes, as most of the clustering algorithms do.

If we denote $\{A_1, \cdots, A_{c_1}\}$ and $\{B_1, \cdots, B_{c_2}\}$ the two respective sets of c_1 classes in the image from sensor Σ^a and c_2 classes in the image from sensor Σ^b, before data fusion, the simple hypotheses which have a non null mass function are:

$$\{A_i \cap B_j \text{ such that } A_i \cap B_j \neq \emptyset,\ i \in [1, c_1],\ j \in [1, c_2]\}.$$

The set of compound hypotheses is:

$$\{(A_{l_1} \cup \cdots \cup A_{l_m}) \cap (B_{k_1} \cup \cdots \cup B_{k_n}) \text{ such that } A_{l_i} \cap B_{k_j} \neq \emptyset,\ l_i \in [1, c_1],\ k_j \in [1, c_2]\}.$$

As we assume that there are no ambiguities between clusters discriminated by sensor Σ^i, $i \in \{a, b\}$, we affect a null mass function $\mathbf{m_i}$ to every compound hypothesis which represents union between clusters of the image from Σ^i. Then, for each pixel x_s, we define the non null mass functions as follows:

For simple hypotheses:

$$\mathrm{H} = A_i \cap B_j \text{ such that } A_i \cap B_j \neq \emptyset \quad \begin{cases} \mathbf{m_a}(\mathrm{H}) = \mathbf{m_a}(A_i) = \dfrac{p_a(x_s/A_i)}{Z_a} \\ \mathbf{m_b}(\mathrm{H}) = \mathbf{m_b}(B_j) = \dfrac{p_b(x_s/B_j)}{Z_b}. \end{cases} \tag{10a}$$

For compound hypotheses ($n \geq 2$ or $m \geq 2$):

$$H = A_i \cap (B_{k_1} \cup \cdots \cup B_{k_n}) \text{ such that } A_i \cap B_{k_j} \neq \emptyset,\ j \in [1, n] \quad \begin{cases} \mathbf{m_a}(\mathrm{H}) = \mathbf{m_a}(A_i) = \dfrac{p_a(x_s/A_i)}{Z_a} \\ \mathbf{m_b}(\mathrm{H}) = 0 \end{cases} \tag{10b}$$

$$H = (A_{l_1} \cup \cdots \cup A_{l_m}) \cap B_j \text{ such that } A_{l_i} \cap B_j \neq \emptyset,\ i \in [1, m] \quad \begin{cases} \mathbf{m_a}(\mathrm{H}) = 0 \\ \mathbf{m_b}(\mathrm{H}) = \mathbf{m_b}(B_j) = \dfrac{p_b(x_s/B_j)}{Z_b} \end{cases} \tag{10c}$$

$$\text{otherwise } \mathbf{m_a}(\mathrm{H}) = \mathbf{m_b}(\mathrm{H}) = 0$$

where $P_a(x_s/A_i)$ [resp. $p_b(x_s/B_j)$] is the conditional probability for pixel x_s to belong to cluster A_i in image 1 (resp. cluster B_j in image 2); Z_a and Z_b are normalization terms such that: $\Sigma \mathbf{m_a} = 1$ and $\Sigma \mathbf{m_b} = 1$ (1). If we denote n_i (resp. n_j) the number of non null intersections between A_i and B_j, $j \in [1, c_2]$, (resp. $i \in [1, c_1]$), there are $\sum_{k=1}^{n_i} \mathbf{C}_{n_i}^k$ hypotheses having a mass value equal to $p_a(X_s/A_i)/Z_a$ [where $\mathbf{C}_{n_i}^k$ is the number of combinations of k elements among n_i: $\mathbf{C}_{n_i}^k = n_1!/k!(n_i - k)!$]. Finally, since $\sum_{k=1}^{n_i} \mathbf{C}_{n_i}^k = (2^{n_i} - 1)$, we have

$$Z_a = \sum_{i=1}^{c_1} \{(2^{n_i} - 1) \times p_a(x_s/A_i)\} \tag{11a}$$

$$Z_b = \sum_{j=1}^{c_2} \{(2^{n_j} - 1) \times p_b(x_s/B_j)\}. \tag{11b}$$

Another way to define mass functions, in the initialization step, would be to define the masses as functions of both probabilities and sizes of the intersections between classes (provided by respective one-source classifications). In such a case, the sizes of the intersections would be similar to *a priori* probabilities in Bayes theory. However, on our data, we found that small clusters (and therefore the less represented land cover types) were penalized, and that classification results were not as good (in terms of land cover type identification, as defined in next Section V) as using mass functions only depending on probabilities.

B. Combination and Decision

The mass functions are combined according to the orthogonal sum defined by Dempster–Shafer (6). Table I summarizes the results obtained with the example of three classes C_1, C_2, and C_3 (Fig. 1). We note that, as all the compound hypotheses

TABLE I
MASS AND BELIEF COMBINATION RESULTS FOR EXAMPLE OF FIG. 1, WITH $1-K=2(t+u-4.t.u)$.

A	C_1	C_2	C_3	$C_1 \cup C_2$	$C_1 \cup C_3$	$C_2 \cup C_3$	Θ
$\mathbf{m_1}$(A)	1-3t	t	t	0	0	t	0
$\mathbf{m_2}$(A)	u	1-3u	u	0	u	0	0
$\mathbf{m}$(A)	$\frac{2u(1-3t)}{1-K}$	$\frac{2t(1-3u)}{1-K}$	$\frac{4.t.u}{1-K}$	0	0	0	0
$\mathbf{Bel}$(A)	$\frac{2u(1-3t)}{1-K}$	$\frac{2t(1-3u)}{1-K}$	$\frac{4.t.u}{1-K}$	$\frac{2(t+u-6.t.u)}{1-K}$	$\frac{2u(1-t)}{1-K}$	$\frac{2t(1-u)}{1-K}$	1
$\mathbf{Bel}(\overline{A})$	$\frac{2t(1-u)}{1-K}$	$\frac{2u(1-t)}{1-K}$	$\frac{2(t+u-6.t.u)}{1-K}$	$\frac{4.t.u}{1-K}$	$\frac{2t(1-3u)}{1-K}$	$\frac{2u(1-3t)}{1-K}$	0

have a null mass after combination, the $\mathbf{Bel}$ and $\mathbf{Pls}$ functions are equal, thus the belief interval reduces to zero: there is no longer imprecision after data combination.

In case of unsupervised data fusion, the combined mass functions of simple hypotheses $A_i \cap B_j$, such that $A_i \cap B_j \neq \emptyset$, are proportional to the sum of the products $\mathbf{m_a}(H_1).\mathbf{m_b}(H_2)$, where $\mathbf{m_i}$ is the mass function of the image acquired by sensor Σ^i, $i \in \{a, b\}$, and H_1 and H_2 are two hypotheses such that $H_1 \cap H_2 = A_i \cap B_j$. From (10a) and (10b), the only hypotheses H_1 which have a non null mass $\mathbf{m_a}(H_1)$ are either $A_i \cap B_j$, or $H_1 = A_i \cap (B_j \cup B_{k_1} \cup \cdots \cup B_{k_n})$ such that $\forall 1 \leq m \leq n$, $A_i \cap B_{k_m} \neq \emptyset$. Then, all hypotheses H_1 (having a non null mass) have the same mass value $\mathbf{m_a}(A_i \cap B_j) = p_a(x_s/A_i)/Z_a$. A similar reasoning may be used for H_2 and $\mathbf{m_b}(H_2)$.

The normalization constant K is the sum of the products $\mathbf{m_a}(H_1).\mathbf{m_b}(H_2)$ with H_1 and H_2 such that $H_1 \cap H_2 = \emptyset$. If we only consider hypotheses which have non null mass $\mathbf{m_a}(H_1)$ and $\mathbf{m_b}(H_2)$, H_1 and H_2 are such that $H_1 = A_i \cap (B_{k_1} \cup \cdots \cup B_{k_n})$ and $H_2 = (A_{1_1} \cup \cdots \cup A_{1_m}) \cap B_j$. Then, the condition for $H_1 \cap H_2 = \emptyset$ is either $i \neq l_p, \forall 1 \leq p \leq m$, or $j \neq k_p, \forall 1 \leq p \leq n$.

Finally, n_i and n_j having the same meaning as in the previous section, the combined mass of $A_i \cap B_j$ is

$$\begin{aligned}\mathbf{m}(A_i \cap B_j) = &\frac{\mathbf{m_a}(A_i \cap B_j) \times \mathbf{m_b}(A_i \cap B_j)}{1-K} \\ &\times \{1 + (2^{n_i-1} - 1) + (2^{n_j-1} - 1) \\ &+ (2^{n_i-1} - 1)(2^{n_j-1} - 1)\} \qquad (12a)\end{aligned}$$

where

$$\begin{aligned}K = &\sum_{A_i \cap B_j \neq \emptyset} \mathbf{m_a}(A_i \cap B_j) \times \mathbf{m_b}(A_i \cap B_j) \\ &\times \{(2^{n_i-1} - 1) \times (2^{n_j-1} - 1) + (2^{n_i-1} - 1) \\ &\times 2^{n_j-1} + 2^{n_i-1} \times (2^{n_j-1} - 1)\} \qquad (12b)\end{aligned}$$

and all compound hypotheses have null resulting mass.

In the case where the mass functions of simple hypotheses are never null (derived from Gaussian probabilities for example), a decision rule such as the maximum of belief over all hypotheses will always favor compound hypotheses. In this case, we prefer a decision rule taken over simple hypotheses. In a first step we choose the following decision rule:

$$\max[\mathbf{Bel}(A)], \qquad \text{if } \mathbf{Bel}(A) \geq \mathbf{Bel}(\overline{A}) \qquad (13)$$

over singleton classes. If there is no class satisfying this condition, the pixel is said "unclassified."

This rule, which may seem severe, has been chosen to consider only pixels which can be classified with a large amount of confidence in the unsupervised multisource classification described in next section. To illustrate the meaning of this decision rule, let us assume that the image has to be classified (according to the maximum of belief decision criterion) in two classes A and $\overline{A}$, where $\overline{A}$ represents any other class or union of classes than A. One pixel satisfying (13) would be labeled "A" rather than "not A." In such a case, we may assume that, if a new class B, intermediate between two classes included in $\overline{A}$, is now taken into account, the pixel would not change its label from "A" to "B," therefore, the class A is said "reliable."

C. Unsupervised Multisource Classification Algorithm

The initial set of singleton classes, considered for data fusion, is the set of non empty intersections between monosource classes. However, some of these singleton classes may be empty (or contain only very few pixels) after Dempster–Shafer combination. There are two main reasons for this.

- There may be some non empty class intersections due only to classification errors in monosource classification. In this case of "false" intersection, if the ill-classified pixels are sufficiently well recognized by the other sensor, classification errors will disappear after data fusion and the singleton class due to this false intersection may be empty.
- Even if there is no monosource classification error, there may some pixels which are classified in class C_1 in image 1, and to class C_2 in the image 2, and which will be classified neither in class C_1 nor in class C_2 (nor in the union) after data fusion. This is the case in the simple example of Fig. 1, if $t = u = \frac{9}{40}$; if the decision criterion is the maximum of belief over simple hypotheses, the pixel will be labeled C_3 since $\mathbf{Bel}(C_1) = \mathbf{Bel}(C_2) = \frac{13}{44}$, and $\mathbf{Bel}(C_3) = \frac{18}{44}$ (cf. Table I); and, if the decision criterion is the maximum of belief over all hypotheses except Θ, the pixel will be classified in $C_1 \cup C_3$ or $C_2 \cup C_3$

rather than in $C_1 \cup C_2$ [$\mathbf{Bel}(C_1 \cup C_3) = \mathbf{Bel}(C_2 \cup C_3) = \frac{31}{44}$, $\mathbf{Bel}(C_1 \cup C_2) = \frac{26}{44}$].

When a singleton class (defined as intersection between two monosource classes) is empty after data fusion, it is discarded as being invalid for the remainder of the classification. In the following, we describe an unsupervised iterative method to estimate the clusters:

0) classify images 1 and 2, find $\{A_1, \cdots, A_{c_1}\}$ and $\{B_1, \cdots, B_{c_2}\}$ the two respective sets of classes in images 1 and 2, and compute the conditional probabilities p_a and p_b;
1) $k = 0$;
   the data fusion set of $c^{(k)}$ classes is: $\Theta^{(k)} = \{A_i \cap B_j, \forall (i, j)$ such that $A_i \cap B_j \neq \emptyset\}$;
2) $k = k + 1$;
   for each pixel:
   a) compute the mass functions $\mathbf{m_a}^{(k)}$ and $\mathbf{m_b}^{(k)}$ (10) and (11).
   b) compute $\mathbf{m}^{(k)} = \mathbf{m_a}^{(k)} \oplus \mathbf{m_b}^{(k)}$ (12) and deduce $\mathbf{Bel}^{(k)}$;
   c) label the pixel according to decision criterion (13);
3) for each class $C_i \in \Theta^{(k)}$:

   if there "too few" pixels in C_i, suppress it:

$$\begin{cases} \Theta^{(k)} = \Theta^{(k)} - \{C_i\} \\ c^{(k)} = c^{(k)} - 1 \end{cases}$$

   if $c^{(k)} \neq c^{(k-1)}$, return to 2).

As a summary, the only clusters that we propose to consider are those which have a non null intersection before data fusion combination, and have "enough" pixels after Dempster–Shafer combination. In our case, the threshold for "empty" classes was chosen equal to 1/1000 of the total number of pixels. On our data, we found that the convergence is achieved after about ten iterations.

The proposed data fusion algorithm allows to consider the mixed aspect of some mixed pixels during classification process. However, after data fusion, no mixed pixels will be observed in the results: since the multisource classification that we describe is unsupervised, the only classes to be considered are the classes which are distinguishable from the information provided by the sources. Therefore, for any pixel which is mixed in one monosource modeling, there is another source which will provide the information to lift the ambiguity.

The importance of considering compound hypotheses for evidence estimation is illustrated in Fig 2. Fig. 2(a)–(c) shows the decision areas, in the case of the example given in Fig. 1, for the following three data fusion models: Bayes formulation with equal probabilities for indistinguishable classes [Fig. 2(a)], Dempster–Shafer formulation with maximum of belief over all simple hypotheses [Fig. 2(b)], and with the decision criterion given by (13) [Fig. 2(c)]. The decision areas are represented versus $[1 - \mathbf{m_1}(C_1)]$ and $[1 - \mathbf{m_2}(C_2)]$. We clearly see that not considering compound hypotheses disadvantages class C_3 relative to classes C_1 and C_2. This is due to the impossibility to identify C_3 using either image 1 alone or image 2 alone. On Fig. 2(c), the decision area called "unclassified" represents the mass values for which the criterion of (13) is not satisfied for any class.

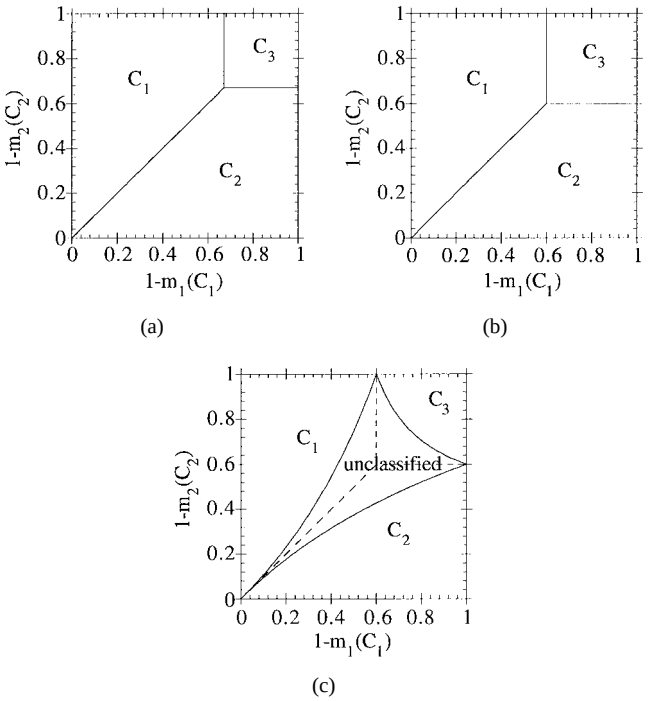

Fig. 2. Decision areas, in case of example of Fig. 1, given by: (a) Bayes formulation with equal probabilities for indistinguishable classes, (b) Dempster–Shafer formulation with maximum of belief over all simple hypotheses, and (c) Dempster–Shafer formulation with decision criterion given by (13).

D. Regularization Step and Global Classification Algorithm

The last step of classification is a regularization step, where a neighborhood term is introduced to reduce classification errors, particularly errors due to speckle in radar image. For this last step, we consider the case of Bayesian mass functions: only simple hypotheses are considered, and the three functions $\mathbf{m}$, $\mathbf{Bel}$, and $\mathbf{Pls}$ are equal and represent the Bayesian probability.

Referring to Geman's modeling [8], the field of the labels is supposed to be Markovian. Here, we consider the case of eight-connexivity neighborhood with cliques of order one and order two. The potential function of one-order clique is deduced from Bayesian mass functions like in Bayes formulation, and the pair-cliques potential function is defined according to a Potts model. The optimization algorithm we used to obtain Maximum *a posteriori* (MAP) classification is the Iterative Conditional Modes (ICM), a simpler and faster version of Geman's algorithm [8].

As summary, the global classification algorithm we will use has the following three steps:

- initialization step: unsupervised Bayesian classification of each data set alone, and initialization of Dempster–Shafer data fusion;
- Dempster–Shafer data fusion with iterative determination of the set of valid clusters;
- regularization step and final classification.

Fig. 3 shows a synoptic of this algorithm.

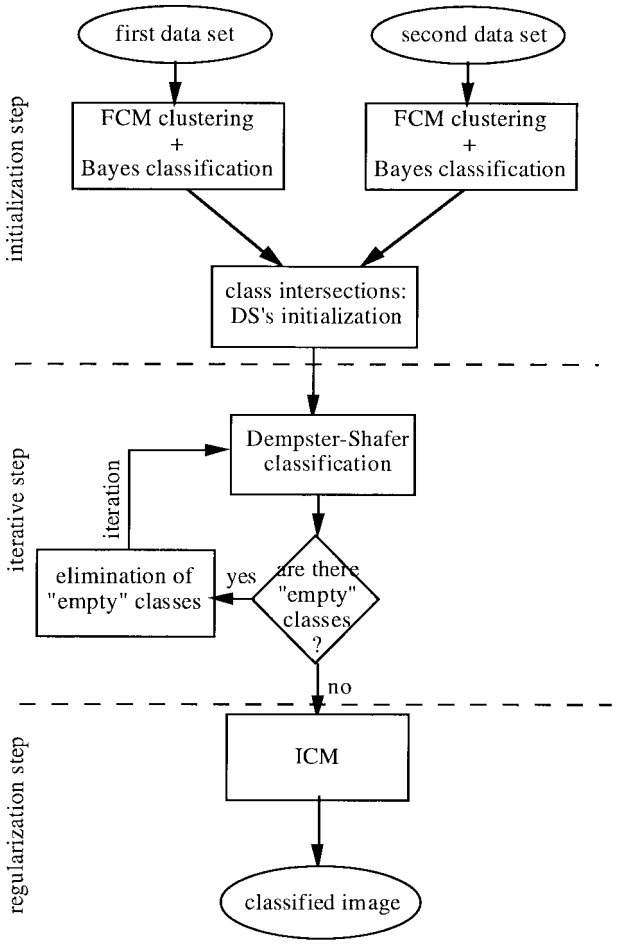

Fig. 3. Global data fusion algorithm.

V. Experimental Results

The data were collected during the MAC-Europe campaign (Summer 1991) over the French Orgeval site [13]. The area covered by the Orgeval site is about 8×8 km^2, and the test site is rather flat (the global variation in level is less than 40 m). The Orgeval basin is mainly an agricultural area surrounded by forests and containing villages.

Radar data were acquired with the NASA/JPL Aircraft SAR, three frequencies (P, L, and C bands) and full quad-polarization, on two days, June 15 and July 16. Optical data were collected with the NASA/JPL Thematic Mapper Simulator (12 bands), on July 5 and July 29. In this paper, we only report results corresponding to the AirSAR image of July 16, with incidence angle centered at 45°, and the July 29 Thematic Mapper Simulator (TMS) images in 2, 5, 6 (optical wavelengths, respectively, comprised between 0.45–0.52, 0.63–0.69, and 0.69–0.75 μm) and 10 (infra-red, $\lambda \in [2.08\text{–}2.35\ \mu\text{m}]$) bands. The last ones were selected because they present best contrast and greatest complementarity, measured in terms of conditional entropy as defined by Shannon in his mathematical theory of information [19].

A. AirSAR and TMS Image Models

The polarimetric scattering properties of targets and background clutters are described by the complex scattering matrix $S = \begin{pmatrix} \text{hh} & \text{vh} \\ \text{hv} & \text{vv} \end{pmatrix}$, where, in ij, j denotes the antenna polarization in transmission, and i the antenna polarization in reception [20]. In this paper, we consider the polarimetric feature vector proposed by E. Rignot [15]:

$$\mathbf{x_r} = \begin{bmatrix} 10.\log_{10}(\langle|\text{hh}|^2\rangle) \\ 10.\log_{10}(\langle|\text{hv}|^2\rangle) \\ 10.\log_{10}(\langle|\text{vv}|^2\rangle) \\ 10.\log_{10}(|\langle\text{hh}.\text{vv}^*\rangle|) \\ \dfrac{10}{\ln(10)}.\arg(\langle\text{hh}.\text{vv}^*\rangle) \end{bmatrix} \tag{14}$$

where $\langle\ \rangle$ means averaged value over a 7×7 window. The main advantage of operating in the log domain is to make the image speckle have the characteristics of an additive noise.

We assume that the distribution of the polarimetric feature vector $\mathbf{x_r}$ conditionally to a class i is a multi-variate Gaussian:

$$p(\mathbf{x_r}|\xi = i) = \frac{1}{\sqrt{2\pi^5.|W_i|}} \exp\left\{-\frac{1}{2}(\mathbf{x_r} - \mathbf{v_i})^t.W_i^{-1}.(\mathbf{x_r} - \mathbf{v_i})\right\} \tag{15}$$

where ξ is $\mathbf{x_r}$ label, $\mathbf{v_i}$ the ith cluster center, W_i is the $\mathbf{x_r}$ covariance matrix of cluster i, $|W_i|$ its determinant. This assumption is verified *a posteriori* from clustering results, with a Kolmogorov–Smirnov test at a level of significance equal to 0.95. Then, based on the knowledge of $\mathbf{x_r}$ statistical behavior, the concatenated vector approach is suitable for classification of one band AirSAR image, and then there is no need to apply Dempster–Shafer combination rule, which implies heavy calculations.

For the optical/infrared TMS data, the considered feature vector is

$$\mathbf{x_o} = \begin{bmatrix} x(10) \\ x(2) \\ x(5) \\ x(6) \end{bmatrix} \tag{16}$$

where $x(i)$ is the pixel value in band i ($i \in [10, 2, 5, 6]$). These four components have a common dynamic range of 255. We assume that $\mathbf{x_o}$ follows a multi-variate Gaussian distribution conditionally to the clusters, and the concatenated vector approach is convenient for optical data classification.

The radar spatial resolution, before and after averaging, and optical pixel size are, respectively, about 12, 36, and 25 m. The optical images have been projected in the radar geometry. Projection has been done by selecting the same reference points in the two images, and approximating the distortion between the images by a polynomial of degree equal to three (for 75 reference points). Thus, the mean error was equal to 1.962 (pixels) for the lines and 1.467 for the columns. Figs. 4 and 5 show, respectively, the L band AirSAR VV power image, and the TMS image (band 10), after projection on to the AirSAR image geometry. In the two cases, the image size is equal to 700 lines by 700 columns (representing about 9 km^2).

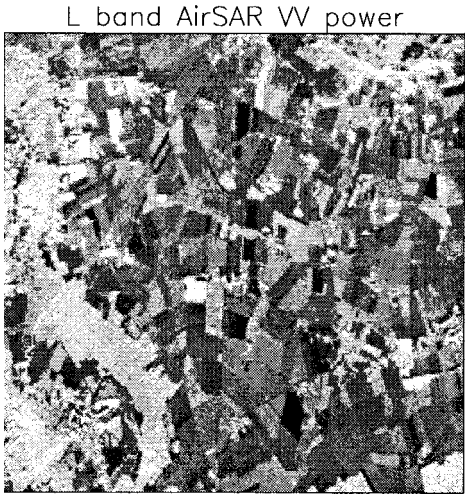

Fig. 4. L band AirSAR VV power image.

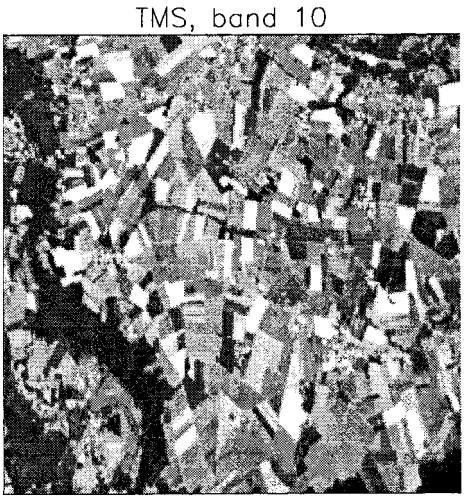

Fig. 5. TMS image (band 10), after projection in the AirSAR image geometry.

B. Land Cover Identification Rate Criterion

Here, we focus on the problem of land cover type identification. Table II shows the different land cover types present on the Orgeval site and an estimation of their percentage over the site, obtained from supervised classification result on AirSAR L and C bands polarimetric SAR images. The land cover types which are the most present on the site (they represent about 65% of the site) are: forest, wheat, and peas. Some other land cover types such as flax, and beans are very poorly represented. The averaged size of the fields is about 200 pixels (about 3 hectares), except for flax, broad beans, and string beans, for which it is about 100 (about 1.5 hectares).

In the case of unsupervised classification, the number of clusters may be different from the number of *a priori* known land cover types. Particularly, we found (from supervised analysis of the polarimetric characteristics of the different land cover types) that there may great differences, due to different states of maturity, in target characteristics between different fields of a same land cover type. Clearly, in such a case, the number of image clusters is greater than the number of culture types. However, we consider that the detection of a land cover type represented by more than one cluster is not hindered, since the clusters may be merged after the supervised interpretation process. According to these considerations, we propose a new classification performance criterion, called "identification rate":

TABLE II
PERCENTAGE OF THE DIFFERENT LAND COVER TYPES PRESENT OVER THE ORGEVAL SITE, NUMBER OF TEST AREAS

land cover type	percentage over the site	number of test areas	total number of test pixels
forest	≈15	9	5229
wheat	≈30	30	7524
peas	≈23	26	4167
corn	≈9.5	12	2007
barley	≈6.5	7	1170
flax	≈2.0	8	1224
broad beans	≈5.5	6	693
string beans	≈1.5	4	290
town	≈7.0	3	432

For each land cover type κ, the identification rate $\tau_{id}(\kappa)$ is defined by

$$\tau_{id}(\kappa) = 100.\sum_{i=1}^{c} p(\xi_i/k=\kappa).p(k=\kappa/\xi_i) \qquad (17)$$

where ξ_i is the pixel label taking its value in the set of c classes, and $p(\xi_i/k)$ is the conditional probability of label ξ_i knowing land cover type k. According to this definition, the detection and identification of a land cover type is only penalized by the fact that it is represented by one or several clusters also representing other land cover types.

For example, let us consider the case of four different land cover types. κ_1, κ_2, κ_3, κ_4, and four classes: ξ_1, ξ_2, ξ_3, ξ_4. Let us assume that classification results provide the confusion matrices shown in Table III: Table III(a) shows the distribution (in percentage of pixels) of a given land cover type κ_i, for $i \in [1, 4]$, in the different classes, and Table III(b) shows the distribution (in percentage of pixels) of a given class ξ_i, for $i \in [1, 4]$, in the different land cover types. Thus, according to (17), the identification rates are: $\tau_{id}(\kappa_1) = 100\%$, $\tau_{id}(\kappa_2) = 50\%$, $\tau_{id}(\kappa_3) = 41\%$, and $\tau_{id}(\kappa_4) = 9\%$.

These results are in agreement with the intuitive interpretation of Table III. The land cover type κ_1 will be, after classification, completely identified; κ_2 will only be half-detected; and κ_4 will be very poorly discriminated, since the class (ξ_3) which mainly represents it (at 90%) is in fact mainly (at 95%) representing other land cover types (κ_2 and κ_3). The advantage of classification result representation in terms of land cover type identification rate is the easiness of interpretation (in case where we do not want to penalize land cover represented by more than one class). However, in case of ill-discrimination of a land cover type, identification rates

TABLE III
EXAMPLE OF DISTRIBUTION (IN PERCENTAGE OF PIXELS) OF (a) A GIVEN LAND COVER TYPE κ_n IN FOUR DIFFERENT CLASSES AND (b) A GIVEN CLASS ξ_n IN FOUR DIFFERENT LAND COVER TYPES

	ξ_1	ξ_2	ξ_3	ξ_4
κ_1	25 %	75 %	0	0
κ_2	0	0	90 %	10 %
κ_3	0	0	90 %	10 %
κ_4	0	0	90 %	10 %

(a)

	κ_1	κ_2	κ_3	κ_4
ξ_1	100 %	0	0	0
ξ_2	100 %	0	0	0
ξ_3	0	50 %	45 %	5 %
ξ_4	0	50 %	5 %	45 %

(b)

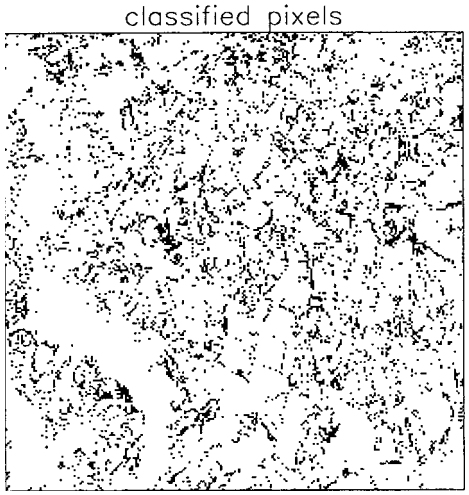

Fig. 6. Conflict image (with dynamic range between 0–255) of multisource classification between TMS and AirSAR L band images.

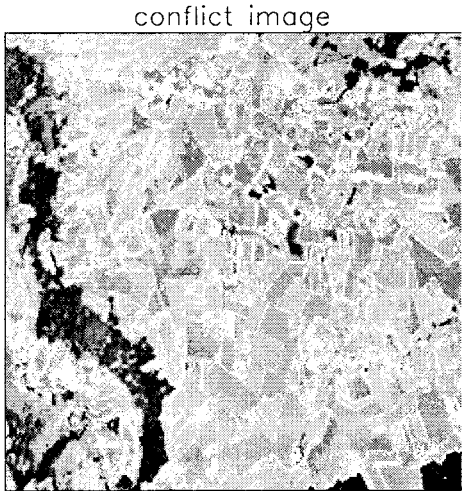

Fig. 7. Binary image of classified (coded in white) and unclassified (coded in black) pixels, according to criterion given by (13).

do not provide information about the other land cover type(s) with which it is confused. To access this information, we have to refer to the confusion matrices.

Identification rates are evaluated from test areas where ground truth is available. The number of test areas and the total number of test pixels for each of the different land cover types are shown in Table II. We note that, as there are less test areas in case of broad beans, string beans, and town, the identification rate values of these land cover types will be less reliable. The confidence intervals of the identification rate values have been evaluated in the following way: four identification rate values have been computed considering only one test-pixel (pixel belonging to a test area) out of four, the confidence interval is then defined as the difference between maximum and minimum values obtained for the four different sets of the test-pixels. In the next sections, Fig. 9 will show the identification rates of the Orgeval land cover types after data fusion between different data sets; the symbol indicates the value obtained by computing identification rates from all the test-pixels, and the error bar shows the minimum and maximum values (of identification rate) obtained by considering one test-pixel out of four.

C. Synergism Between TMS Images and L Band AirSAR Image

Firstly, we only consider TMS images (bands 2, 5, 6, and 10) and L band AirSAR image. The numbers of classes used for monosource classification are supervised parameters of the proposed data fusion algorithm. It was applied with monosource numbers of classes taking any values between six and 16. The input to data fusion is the output of monosource classifications; then, if one of the two monosource numbers of classes is too small (i.e., if monosource information is to much simplified), the data fusion will not perform optimally. On the other hand, when the monosource numbers of classes are both large enough, the final data fusion result becomes stable (equivalent number of classes, classification results, and identification rates). In summary, thanks to the iterative process, the described data fusion algorithm is not sensitive (except on computational requirements) to the monosource numbers of classes, provided that they are both large enough. In case of synergism between TMS and L band AirSAR data, the identification rates (of the different land cover types) and the final number of classes become stable for monosource classification numbers of classes greater than eleven. Thus, in this case of data fusion, the monosource numbers of classes have been chosen to be both equal to 11: smaller numbers of classes lead to noticeably worse classification results (in terms of land cover identification) and larger numbers of classes do not improve it.

Firstly, for each data set, radar and optical, the cluster characteristics have been estimated using the fuzzy c-means algorithm [4], [15], and the image has been classified accord-

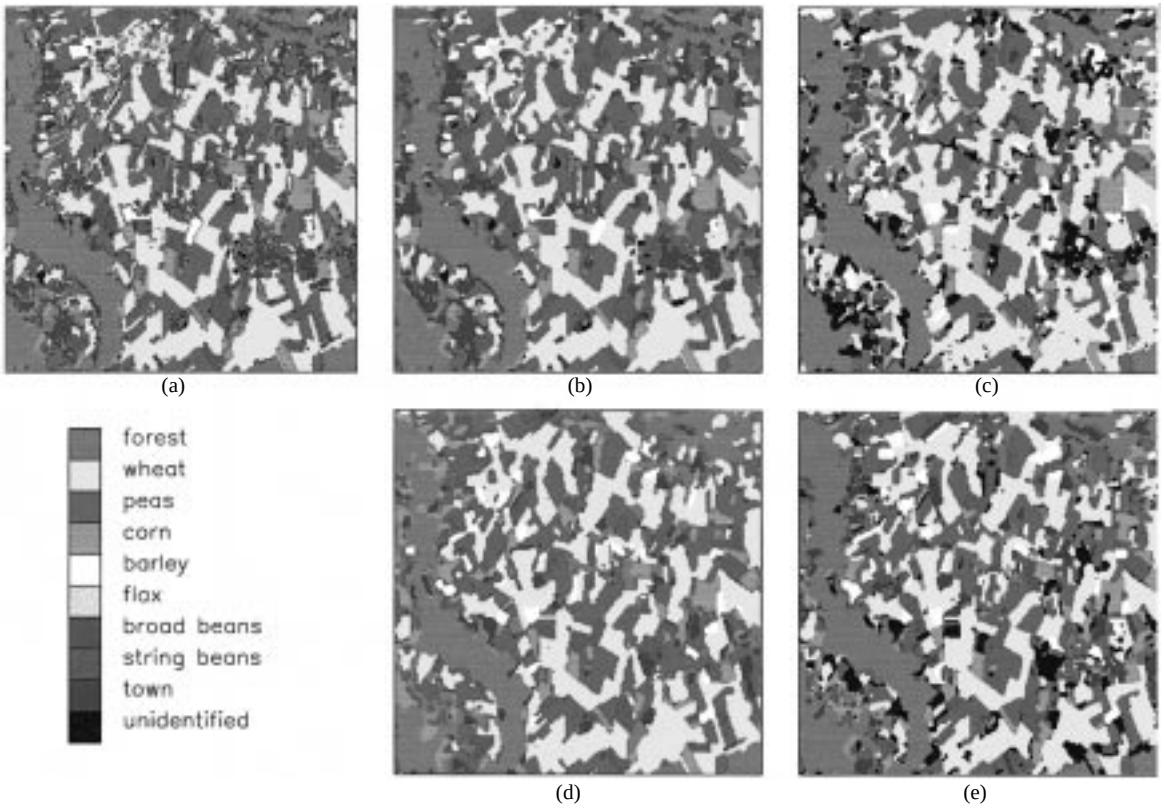

Fig. 8. Results of multisource classification (a) "O+L," before the regularization step, (b) "O+L," after the regularization step, (c) "L+C," (d) "O+C," and (e) "O+L+C" (clusters representing the same land cover are coded by the same label).

ing to Bayes criterion. Then, the two classified images are put in correspondence and all the non null intersections between their classes will be considered in the initial set of classes of Dempster–Shafer classification: the initial number of clusters is equal to 72. The mass functions are computed from multivariate Gaussian probabilities and combined according to the Dempster's rule. The final number of clusters is found to be equal to 34.

Fig. 6 shows the image of conflict values (12b). The minimum, mean, and maximum values of conflict are, respectively, equal to 0.00, 0.49, and 0.72. We note that the conflict is the lowest for forest areas. Most wheat fields present a lower value of conflict than the other fields. In fact, we will see that this land cover type was already well recognized by both radar and optical sensors. We also note that conflict is more important at the borders of fields. This is probably due to slight errors in the projection of the optical image in radar geometry.

The decision criterion was defined by (13). If there is no hypothesis satisfying this criterion, the pixel is said unclassified. Fig. 7 shows the binary image of classified (in white) and unclassified (in black) pixels. The percentage of unclassified pixels is equal to 12.2%. We see that, even if our decision criterion was quite strict, only few pixels do not satisfy it. Initially, considering the 72 singleton classes, the number of unclassified pixels was equal to 15.8%. As during the iterative data fusion, the number of unclassified pixels decreases, we may assume that classes become more reliable (according to the definition of the Section IV-B).

Fig. 8(a) and (b) show the classification results respectively before and after the regularization step. The clusters have been identified in terms of land cover types, and the classes representing the same land cover have been coded by the same label. We note that classification errors within some fields and at their borders have been reduced by the regularization step.

Fig. 9(a) shows the identification rates of the different land cover types present in the Orgeval site, provided by unsupervised classifications using respectively the three following data sets: polarimetric L band AirSAR image (called "L"); TMS bands 2, 5, 6, and 10 bands (called "O"); and both TMS bands and AirSAR L band (called "O+L"). We clearly see the improvement in land cover identification due to data fusion. We note that some land cover types such as forest, wheat and peas were already well identified using only one of the data sets (TMS or AirSAR, L band). In fact this land cover types are the most present on the Orgeval site, since they represent about 65% of the site. The other land cover types are less represented (see Table II) and thus more difficult to identify in a robust way as shown by the length of the error bars. Finally, we note that there are two land cover types: barley and flax, which are not well identified (their identification rates are respectively equal to 46 and 56%). In fact, these two land cover types are well discriminated relative to the other land cover types, but they are not distinguishable from each other using TMS and L band AirSAR information. However, we will see, in the next section, that C band AirSAR can discriminate them successfully.

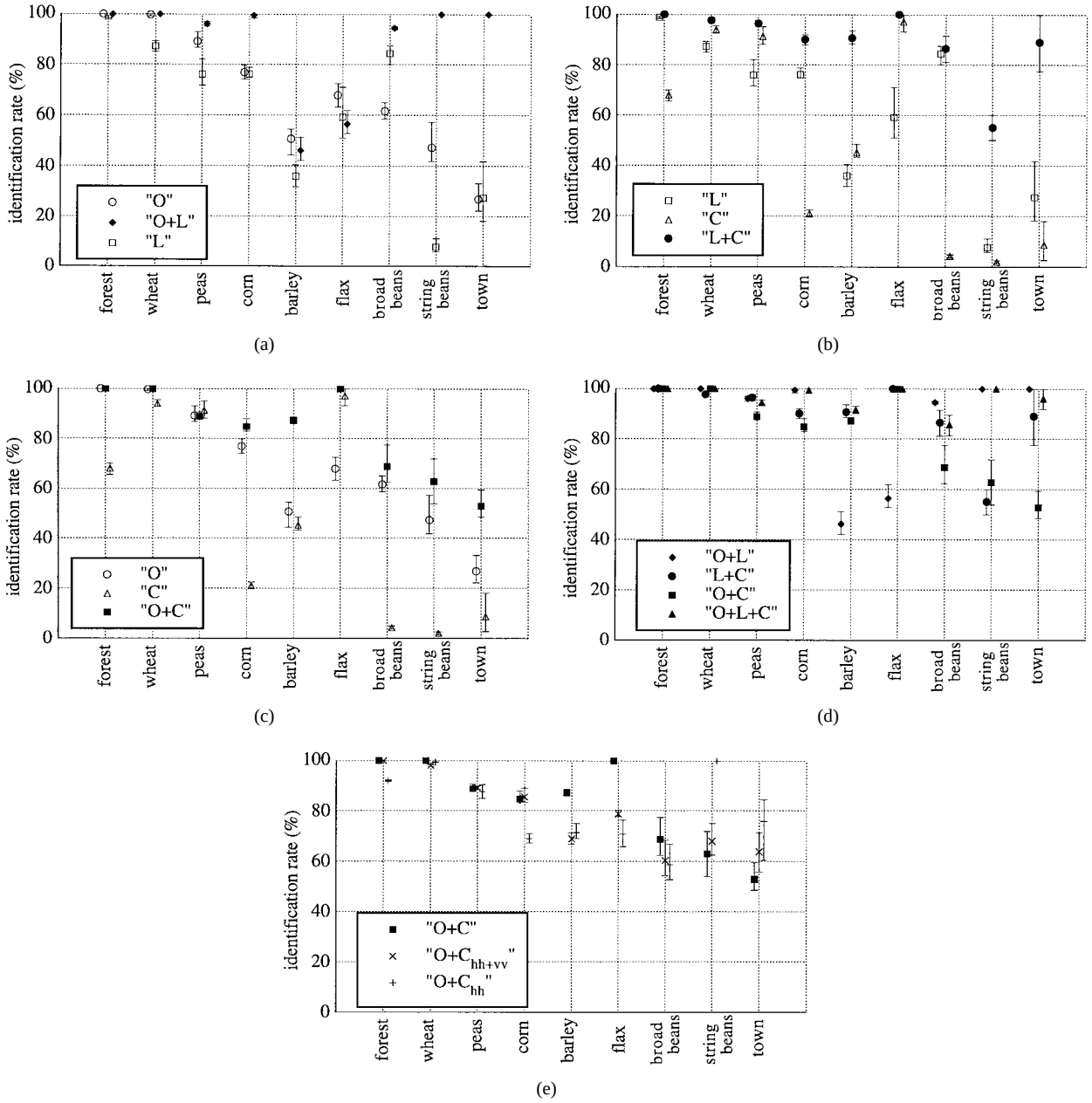

Fig. 9. Comparison of land cover identification rates (%) obtained from classifications using the following data sets: (a) "O," "L," and "O+L"; (b) "L," "C," and "L+C"; (c) "O," "C," and "O+C"; (d) "O+L," "L+C," "O+C," and "O+L+C"; and (e) "O+C," "O+C_{hh}," and "O+C_{hh+vv}."

D. Comparison between Multiband L–C Classification and Synergism between Optical and SAR Data

We now aim at comparing data fusion results versus the data sets used: TMS images, L band AirSAR data, and/or C band AirSAR data. The classification algorithm used is the same as described in the previous section. The monosource number of classes for C band AirSAR was chosen equal to eight. Fig. 8(c) and (d) show the classification results respectively using "L+C" (polarimetric C band AirSAR image is coded "C") and "O+C" data sets, with clusters representing the same land cover coded by the same label.

Fig. 9(b) shows the comparison of land cover identification rates computed from classification results respectively obtained using the three following data sets: "L," "C," and "L+C." Fig. 9(c) shows the comparison for the three data sets: "O," "C," and "O+C." As for Fig. 9(a), we clearly see the improvement due to data fusion. Error bars are lower than 5% for all the land cover types, except for broad beans, string beans, and town, for which error bars are about 10 or 20%. Multiband L–C classification results are very good: only identification rate of string beans is below 85%. From Fig. 9(b) or (c), we note flax is significantly better identified using C band SAR data than L band SAR or optical data, then it is well identified (and barley in the same way) by data fusion "L+C" or "O+C."

Comparing the Fig. 9(a)–(c), we can say that globally best data fusion results are achieved for "L+C," and "O+L": about 90% of the Orgeval image is identified with a rate greater than 90%, and error bars being inferior to 5%. "O+C" provides

an identification of 80% of the site with a rate greater than 80%, with about the same error bar values. Moreover, the land cover types which are "not well identified" are different with "L+C" and "O+L": string beans in the first case, and barley and flax in the second case. So, we suggest improving land cover identification by combination of the three data sets.

Thanks to the associative property of Dempster–Shafer combination rule, the data fusion algorithm we propose is easily applicable to more than two data sets. However, it is computationally heavy. Fig. 8(e) shows the multisource "O+L+C" classification results, and Fig. 9(d) shows a comparison between the results obtained by fusion of two among the three data sets "O," "L," and "C," and the three data sets (called "O+L+C"). In latter case, the identification rates are superior to 95% for all the land cover types (present on the Orgeval site) but barley and broad beans, whose identification rates are respectively equal to 92 and 86%. When more data sets (which are not completely redundant) are used in data fusion, the quantity of information increases and classification becomes more precise.

However, there is a need of some comments about the interpretation of unsupervised classes. In unsupervised classification, all the classes are *a priori* unknown classes. These classes are then *a posteriori* interpreted (in our case, identified in terms of land cover types). If one class cannot be interpreted, it is called an unidentified class, and so are its pixels. In our case, the class interpretation is done by comparison with test areas where ground truth is available (see Table II). However, as ground truth is only known on a subpart of the image, each time a class is not at all present on test areas, it cannot be identified. Table IV shows the final number of classes and the percentage of unidentified pixels (belonging to any unidentified class) on the different classified images: "C," "L," "O," "L+C," "O+L," "O+C," and "O+L+C." We note that in the classified image "O+L+C" the number of unidentified pixels is greater (≈10%) than in the other classification results (≈5%). This increase of the percentage of unidentified pixels in the case of data fusion "O+L+C" is due to both facts that the final number of classes is slightly greater than in the case of one-source or two-source classification (as shown in Table IV), and that ground truth is only known over a subpart of the image. Being aware of the statistical validity of every class, the practical conclusion of this comment is that the number of classes has to be a compromise between accurate detection of the different land cover types and the percentage of unidentified pixels. This is the reason for which the first strategy in mass function definition (see Section IV-A) was not chosen.

The last image combination we tried uses TMS and copolarized channels C band SAR data. The aim of this study was the evaluation of the performance (in terms of land cover type identification) of single band copolarized radar, such as ASAR (ESA/Envisat) or RadarSAT (the Canadian). The RadarSAT and ASAR sensors operate both at C band (5.3 GHz), with a wider range of incidence angles. Their polarization modes are respectively HH for RadarSAT (single polarization), and HH and VV for ASAR (dual polarization). Fig. 9(e) shows the land cover identification rates obtained from classifications using the three following data sets: "O+C," "O+C_{hh}" (TMS and C band HH polarization AirSAR data which simulate RadarSAT data) and "O+C_{hh+vv}" (TMS and C band HH and VV polarization AirSAR data which simulate ASAR data). We note that the results are not as good as previously, since the identification rates of corn, flax and barley are about 20% inferior to those obtained combining full polarimetric C band data and TMS images. However, the use of radar data improves land cover identification relative to classification with optical data alone. This is particularly true for barley, flax, string beans, and town.

TABLE IV
PERCENTAGE OF UNIDENTIFIED PIXELS ON CLASSIFIED IMAGES "C," "L," "O," "L+C," "O+L," "O+C," "O+L+C," "O + V_{hh+vv}," AND "O + C_{hh}."

classified image	number of classes	% of unidentified pixels
"C"	8	0.00
"L"	16	0.00
"O"	17	0.00
"L+C"	31	6.70
"O+L"	34	6.60
"O+C"	32	3.26
"O+L+C"	36	9.50
"O+C_{hh+vv}"	28	6.17
"O+C_{hh}"	29	8.30

E. Comparison with Other Unsupervised Data Fusion Methods

In this final section, we compare the data fusion algorithm presented in this paper with two other unsupervised data fusion methods:

a) the concatenated vector approach: cluster characteristics estimation by the fuzzy c-means algorithm, followed by MAP classification, with eight-connectivity Markov Random Field label image model, is applied as if the different data sets were collected by the same sensor;
b) the subdivision of the classes detected in the first image by the classes detected on the second image, which is equivalent to the initialization of our data fusion algorithm (Fig. 3).

These unsupervised data fusion methods have been chosen because they are quite simple, and we only aim at comparing the efficiency of the proposed unsupervised data fusion method with more simple data fusion algorithms.

Fig. 10 shows, for each land cover type κ, the difference $\Delta\tau_{id}(\kappa)$ between identification rates obtained a) by concatenated vector approach and b) by the class subdivision, and those obtained using data fusion method based on Dempster–Shafer evidence theory. The legends have the same meaning as in the previous sections. Negative values indicate that Dempster–Shafer data fusion performs better and positive values that it performs worse. Absolute differences less than five percents are not significant. From both Fig. 10(a) and (b), we note the global improvement due to Dempster–Shafer

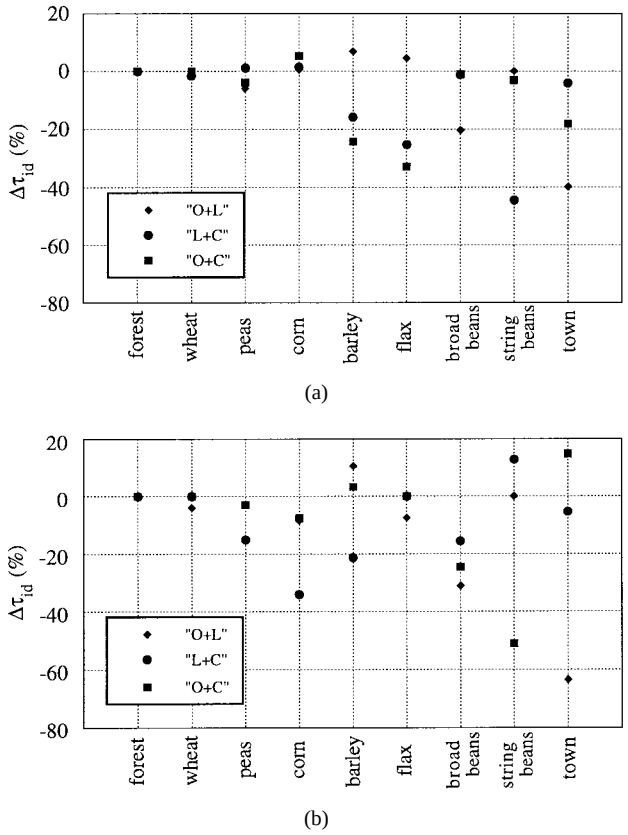

Fig. 10. Difference $\Delta\tau_{id}(\kappa)$ between the identification rates of the land cover types κ, obtained (a) by concatenated vector approach and (b) by the class subdivision, and those obtained using data fusion algorithm of Fig. 3.

data fusion. The improvement depends on the land cover type considered, the data sets used and the compared data fusion algorithm, since the concatenated vector and the class subdivision approaches do not provide the same classification results. For the land cover types which were already well discriminate using only one source information (i.e., forest, wheat, and peas), multisource classifications are equivalent. For "O+L," Dempster–Shafer method provides a significant improvement with an identification rate increase of more than 20% for broad beans and town. For "L+C," it performs better [$\Delta\tau_{id}(\kappa)$ about or greater than 20%] than the concatenated vector approach in identifying barley, flax, and string beans, and better than the class subdivision approach in case of corn, barley, and broad beans. For "O+C," greatest differences between identification rates [$\Delta\tau_{id}(\kappa) \geq 20\%$] are achieved for barley and flax [Fig. 10(a)], or broad beans and string beans [Fig. 10(b)].

VI. Conclusion

Data fusion at pixel level could be successfully used for a better identification of land cover types in remote sensing applications. Recent works [16], [17] have been proposed for statistical classification data. They incorporate much *a priori* information about sensors and image acquisition process. Here, we have described an unsupervised multisource classification method using Dempster–Shafer evidence theory. The advantage of the latter is to consider not only single classes, but also unions of classes, and thus to deal with mixed pixels. In particular, when two classes are indistinguishable by one sensor, Dempster–Shafer evidence theory provides the option not to make a decision between these two classes by affecting a non null mass value to their union. In that case, only the information provided by the other sensors will be used to discriminate between the two classes.

In the proposed data fusion algorithm, the final classes to be considered for data fusion are determined in an unsupervised way by combining the different monosource sets of classes and analyzing all the possible intersections. In particular, indistinguishable classes using only one image information are detected by comparison with the other image classification results, and the monosource mass functions are defined forcing indistinguishable classes and their union to have the same mass. Then, an iterative process allows to discard invalid clusters, which may be due either to monosource classification errors, or to conflicts between the information provided by the different sources.

In our case, lowest conflict values between images were observed on forest areas and wheat fields, both land cover types well identified on every image. Greatest conflict values were observed at the border of the fields. During the iterative step of the described data fusion algorithm, the decision rule was the maximum of belief over single classes under the constraint that singleton belief be superior to the belief of its complementary hypothesis. This condition, satisfied by about 80% of the pixels on the data we used, was imposed to insure of the cluster validity. It also provides an estimation of the reliability of data fusion classes. The last step consists in classifying each pixel even the ones that were not used in the cluster estimation and in smoothing the classified image through a regularization step.

The performance of classification is studied in terms of identification of the different land cover types. Comparing Dempster–Shafer data fusion to two other simple data fusion methods (the concatenated vector and the class subdivision approaches), we show that the former generally performs better (e.g., a 20% improvement in the identification rates for corn using L and C band data). On the MAC-Europe campaign data, the best two-data-set fusion results are obtained either using optical and L band SAR images, or multiband L and C SAR images. Use of the three data sets (optical, L and C band polarimetric SAR) provides identification rates greater than 85% for all the land cover types present on the Orgeval site. Because this method is unsupervised it is not site specific. However, the actual identification rates and their improvement due to the Dempster–Shafer approach, do depend on the site, the types of land cover and the state of the cultures.

Acknowledgment

The authors thank the reviewers for their very constructive remarks that helped us to improve the paper.

References

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

[2] J. A. Barnett, "Calculating Dempster–Shafer plausibility," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 599–602, June 1991.
[3] J. A. Benediktsson and P. H. Swain, "A method of statistical multisource classification with a mechanism to weight the influence of the data source," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Vancouver, B.C., Canada, July 10–14, 1989, vol. 2, pp. 517–520.
[4] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, pp. 191–203, 1984.
[5] I. Bloch, "Some aspect 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.
[6] H. de Boissezon, G. Gonzales, B. Pous, and M. Sharman, "Rapid estimates of crop acreage and production at European scale using high resolution imagery," in *Proc. Int. Symp. "From OPTICS to RADAR, SPOT, and ERS Applications,"* Paris, France, May 10–13, 1993, pp. 353–364.
[7] J. van Cleynenbreugel, S. A. Osinga, F. Fierens, P. Suetens, and A. Oosterlinck, "Road extraction from multi-temporal satellite images by an evidential reasoning approach," *Pattern Recognit. Lett.*, vol. 12, pp. 371–380, June 1991.
[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.
[9] J. W. Guan and D. A. Bell, *Evidence Theory and its Applications.* New York: North-Holland, 1991.
[10] I. Kanellopoulos, G. G. Wilkinson, and A. Chiuderi, "Land cover mapping using combined Landsat TM imagery and textural features from ERS-1 synthetic aperture radar imagery," in *Proc. SPIE: Image and Sig. Proc. Remote Sens.*, Rome, Italy, Sept. 26–30, 1994, vol. 2315, pp. 332–341.
[11] D. G. Leckie, "Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination," *Photogramm. Eng. Remote Sens.*, vol. 56, pp. 1237–1246, 1990.
[12] 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.
[13] M. Benallegue, O. Taconet, D. Vidal-Madjar, and M. Normand, "The use of radar backscattering signals for measuring soil moisture and surface roughness," *Remote Sens. Environ.*, vol. 53, no. 1, pp. 61–68, July 1995.
[14] H. Rasoulian, W. E. Thompson, L. F. Kazda, and R. Parra-Loera, "Application of the mathematical theory of evidence to the image cueing and image segmentation problem," in *Proc. SPIE: Sig. Image Proc. Sys. Performance Evaluation*, Orlando, FL, Apr. 19–20, 1990, vol. 1310, pp. 199–206.
[15] E. Rignot, R. Chellappa, and P. Dubois, "Unsupervised segmentation of polarimetric SAR data using the covariance matrix," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 697–704, July 1992.
[16] 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.
[17] 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.
[18] G. Shafer, *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press, 1976.
[19] C. E. Shannon and W. Weawer, *The Mathematical Theory of Communication.* Urbana, IL: Univ. of Illinois Press, 1963.
[20] F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications.* Norwell, MA: Artech House, 1990.
[21] J. J. van Zyl, R. Carande, Y. Lou, T. Miller, and K. Wheeler, "The NASA/JPL three frequency polarimetric AirSAR system," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Houston, TX, May 26–29, 1992, vol. 1, pp. 649–651.

Sylvie Le Hégarat-Mascle graduated from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1993, and received the Ph.D. in signal and image from ENST in 1996.

She is with the Centre d'Etudes des Environnements Terrestre et Planétaires, Vélizy, France. Her research interests include statistical pattern recognition, image analysis, multisensor classification, and remote sensing.

Isabelle Bloch graduated from Ecole des Mines de Paris, Paris, France, in 1986, and received the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, in 1990, and the "Habilitation à Diriger des Recherches" from University Paris 5 in 1995.

She is an Associate Professor in the Department Images, ENST. Her research interests include 3-D image and object processing, 3-D and fuzzy mathematical morphology, decision theory, data fusion in image processing, fuzzy set theory, evidence theory, medical imaging as well as aerial and satellite imaging.

D. Vidal-Madjar, photograph and biography not available at the time of publication.