Pergamon

Pattern Recognition, Vol. 31, No. 11, pp. 1811–1823, 1998

0031-3203/98 $19.00 + 0.00

PII: S0031-3203(98)00051-X

INTRODUCTION OF NEIGHBORHOOD INFORMATION IN EVIDENCE THEORY AND APPLICATION TO DATA FUSION OF RADAR AND OPTICAL IMAGES WITH PARTIAL CLOUD COVER

S. LE HÉGARAT-MASCLE,*,† I. BLOCH‡ and D. VIDAL-MADJAR*

*CETP (CNRS), 10-12 avenue de l'Europe, 78140 Vélizy, France
‡ENST, 46 rue Barrault, 75013 Paris, France

(*Received* 27 *October* 1997; *accepted* 5 *March* 1998)

Abstract—Two ways of introducing spatial information in Dempster–Shafer evidence theory are examined: in the definition of the monosource mass functions, and, during data fusion. In the latter case, a "neighborhood" mass function is derived from the label image and combined with the "radiometric" masses, according to the Dempster orthogonal sum. The main advantage of such a combination law is to adapt the importance of neighborhood information to the level of radiometric missing information. The importance of introducing neighborhood information has been illustrated through the following application: forest area detection using radar and optical images showing a partial cloud cover. © 1998 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Data fusion Multisource classification Evidence theory Missing information
Spatial neighborhood Remote sensing

1. INTRODUCTION

Due to the increasing number of sensors in data acquisition, data fusion has become an important field of research. In image processing, data fusion may be performed at different stages:[1] pixel, feature and decision level. At the decision level, information from each individual image is derived from some preliminary process and these first classification results are combined in the last step. At the pixel level, many data fusion methods have been proposed, such as probabilistic fusion and Bayesian inference. The simplest approach is to concatenate the data from the different sensors as if they were measurements from one single sensor.[2] In this case, establishing a good model for multisource data is the difficulty (due to the heterogeneity of the vector components derived respectively from the optical image and from the radar image). More sophisticated methods of statistical multisource classification have been proposed.[3–6] However, they are neither adapted to missing information (e.g. corresponding to the presence of clouds in remote sensing optical images), nor to the case of a class defined as the complement of another one (since such a class may be highly heterogeneous and cannot be defined by statistical features).

The Dempster–Shafer mathematical theory of evidence was first introduced by Dempster in the 1960s, and was later extended by Shafer.[7] By allowing the representation of both imprecision and uncertainty, evidence theory appears as a more flexible and general approach compared to the Bayesian approach. In particular, it can deal with missing information as well as with complementary classes. Applications have been developed in image processing (e.g., object detection,[8,9] medical imaging,[10] remote sensing classification,[11,12] as well as other domains such as pattern recognition and classification.[13]) However, to our knowledge, only very few works [14] attempt to incorporate spatial information in evidence theory, whereas a Markovian assumption is widely used in classical Bayesian modeling. In this paper, we introduce spatial information at two levels: in the definition of monosource mass functions, and in a new regularization step, applied after the combination of the different sources.

The application we consider is data fusion of radar and optical images with partial cloud cover. Identification of different land cover types is one of the most important topics in remote sensing applications. Optical images, such as SPOT images, allow good classification results, for example, using classical Bayesian methods. However, optical sensors may be affected by clouds, and there may be some countries (northern or equatorial) where cloud free optical images are difficult to obtain. In this case, radar images, such as ERS-1 images, may be used to supplement the missing

†Author to whom correspondence should be addressed. Tel.: 33 1 39254934; fax: 33 1 39254922; e-mail: sylvie.mascle@cetp.ipsl.fr.

optical image information, since they are not influenced by acquisition time and weather conditions. However, SAR images suffer from speckle, which makes application of classification procedure difficult. Thus, there is a great interest in developing data fusion methods for optical and radar images that reduce the imprecision in the data of the individual sensors, thereby improve classification results. In such approaches to data fusion, the ability of the Dempster–Shafer theory to model the incompleteness of each image, either due to cloudy pixels in the optical image or to speckle in the SAR image, is very useful. Moreover, this example will illustrate the importance of spatial neighborhood information, since thin cloud boundaries cannot be detected in a reliable way using only radiometric information. In particular, we will compare our results with those provided by the data fusion rule $\mathscr{R}$: "if there is a cloud in the optical image, then take the decision given by the radar image classification, otherwise take the decision given by the optical image classification" (note that this rule assumes that there is no imprecision in cloud detection in the optical image).

The remainder of the paper is organized as follows. In Section 2, the way Dempster–Shafer evidence theory may be used to represent different cases of ignorance in optical and SAR images is presented. It is illustrated in the simple case where perfect cloud detection is assumed. In Section 3, some ways for introducing spatial information in evidence theory are proposed, and a data fusion algorithm adapted to the presence of partial cloud cover in the images is described. In Section 4, this algorithm is applied to the problem of forested area discrimination using one SPOT image with partial cloud cover and one ERS image of the same scene. Different cases of simulated cloud cover are studied, and the performance of our data fusion method relative to the above defined classification rule $\mathscr{R}$ is quantified. Finally, the data fusion algorithm is applied to images with actual (not simulated) cloud cover, and the results are presented. Section V is our conclusion.

2. IGNORANCE REPRESENTATION USING DEMPSTER–SHAFER EVIDENCE THEORY

Let Θ denote the set of hypotheses about the membership of a pixel to a class. Dempster–Shafer theory allows one to consider hypotheses that correspond to arbitrary subsets of Θ, as well as hypotheses that correspond to the unions of classes (called compound hypotheses). This is in contrast to the probabilistic approach, in which only singletons (also called simple hypotheses) may be considered. This makes the modeling very flexible and many different situations can be taken into account: ambiguities between classes, mixture, missing information, etc.

In the following, the set of operations which is classically applied to the set of classes is extended to Θ. In particular, inclusion, intersection, and union between two hypotheses A and B are defined and denoted as follows:

$$\begin{cases} A \subseteq B \Leftrightarrow \text{if } A \text{ is true, then } B \text{ is true} \\ (A \cap B) \text{ is true} \Leftrightarrow A \text{ is true and } B \text{ is true} \\ (A \cup B) \text{ is true} \Leftrightarrow A \text{ is true or } B \text{ is true} \end{cases}$$

Since, as mentioned in the introduction, we will take (in Section 4) the problem of forest area discrimination as example, throughout this section, we refer to this problem to provide simple illustrations of Dempster–Shafer theory implementation. In these illustrations, the only considered classes are: forested areas and unforested areas (i.e. anything else than forest). The number of hypotheses in Dempster-Shafer formulation is therefore $2^2 = 4$. These hypotheses are the two simple hypotheses F (forest) and $\bar{F}$ (unforested areas), one compound hypothesis $\Theta = F \cap \bar{F}$, and the empty set $\emptyset$. $\emptyset$ represents the case where no class or union of classes among the considered ones is valid, such as in the case of unknown classes. Obviously, in our application, the hypothesis $\emptyset$ will never be chosen since the definition of $\bar{F}$ prevents us from "forgetting" some classes. Therefore, we will focus on the three remaining hypotheses: F, $\bar{F}$, and Θ.

2.1. *Ignorance representation in optical and radar image models*

The Dempster–Shafer evidence theory provides a representation of ignorance by assigning a non-zero mass function ($\mathbf{m}$) to compound hypotheses. $\mathbf{m}$, which is also called the basic probability assignment, is defined for every (simple or compound) hypothesis A, such that the mass value $\mathbf{m}(A)$ belongs to the interval $[0, 1]$ and satisfies the two following normalization conditions:

$$\mathbf{m}: \begin{cases} \mathbf{m}(\emptyset) = 0, \\ \sum \mathbf{m}(A) = 1, \\ A \subset \Theta. \end{cases} \tag{1}$$

Assigning a non-zero mass to a compound hypothesis $A \cup B$ 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-zero mass to Θ allows one not to classify some pixels, for which there is a global ignorance. For example, since the presence of a cloud in the optical image represents a case of total ignorance about the membership to F or $\bar{F}$ of the area covered by the cloud, it seems reasonable to define cloudy pixel mass function such that: $\mathbf{m}_O(F) = \mathbf{m}_O(\bar{F}) = 0$ and $\mathbf{m}_O(\Theta) = 1$ (where the subscript "O" means "optical" mass function). Conversely, in the case of pixels belonging with certainty to forested areas, we choose: $\mathbf{m}_O(F) = 1$ and $\mathbf{m}_O(\bar{F}) = \mathbf{m}_O(\Theta) = 0$.

In the Bayes theory, the uncertainty about an event is measured by a single value (probability), and the imprecision about uncertainty measurement is assumed to be null. The Dempster–Shafer evidence

theory provides a representation of both imprecision and uncertainty through the definition of two functions: plausibility (**Pls**) and belief (**Bel**). Both map the set of hypotheses to the interval [0, 1], and they are derived from mass function as follows:

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

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

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 $\bar{A}$ (since $\mathbf{Pls}(A) = 1 - \mathbf{Bel}(\bar{A})$), may be interpreted as the maximum uncertainty value about A ($\mathbf{Bel}(A) \leq \mathbf{Pls}(A)$). Then, the length of the interval $[\mathbf{Bel}(A), \mathbf{Pls}(A)]$, called the "belief interval", gives a measurement of the imprecision about the uncertainty value.

A special case which concerns our application is where only three hypotheses are considered: A, $\bar{A}$, and $\Theta = A \cup \bar{A}$. The length of the belief interval is equal to the Θ mass value, both for the A and the $\bar{A}$ hypotheses. For the hypothesis A, $\mathbf{Pls}(A) - \mathbf{Bel}(A) = \mathbf{m}(\Theta)$, since $\mathbf{Pls}(A) = \mathbf{m}(\Theta) + \mathbf{m}(A)$ and $\mathbf{Bel}(A) = \mathbf{m}(A)$. For the $\bar{A}$ hypothesis, $\mathbf{Pls}(\bar{A}) - \mathbf{Bel}(\bar{A}) = \mathbf{m}(\Theta)$.Therefore, a straightforward way to represent radar classification global imprecision is to attribute a constant non-zero mass $\mathbf{m}_{\mathbf{R}}(\Theta)$ (the subscript "R" means "radar") to every pixel in the radar image. This is a simple representation of the fact that the radar image information is less accurate than the optical image information.

In summary, the Dempster–Shafer formulation provides a way to model the two cases of ignorance respectively met in optical and radar images: total ignorance of land cover type in the case of cloud pixels (in the optical image), and a global ignorance factor for radar monosource classification.

2.2. Data combination and decision making

Having defined the evidence functions of each of the different sources, they are combined using Dempster's rule of combination,[7] also called orthogonal sum. If $\mathbf{m}_i$ denotes the basic probability assignment provided by source $\Sigma_i (1 \leq i \leq p)$, the orthogonal sum, denoted $\mathbf{m} = \mathbf{m}_1 + \cdots + \mathbf{m}_p$, is defined by:

$$\begin{cases} \mathbf{m}(\emptyset) = 0 \\ \text{if } K \neq 1, \ \mathbf{m}(A) = \dfrac{\displaystyle\sum_{a_1 \cap \cdots \cap a_p = A \ 1 \leq i \leq p} \ \prod \mathbf{m}_i(a_i)}{1 - K} \qquad (4) \\ \text{where } K = \displaystyle\sum_{a_1 \cap \cdots \cap a_p = \emptyset \ 1 \leq i \leq p} \ \prod \mathbf{m}_i(a_i) \qquad (5) \end{cases}$$

From equation (5), we see that K ($K \in [0, 1]$) represents the mass which would be assigned to the empty set $\emptyset$, after combination, in the absence of normalization [division by $(1 - K)$ in equation (4)]. Thus, K is often interpreted as a measure of conflict between the different sources: the larger K is (with $0 \leq K \leq 1$), the more the sources are conflicting and the less it makes sense to combine them. When K is equal to 1, the sources are said to be totally or flatly contradictory, and it is no longer possible to combine them. K has been introduced in equation (4) as a normalization factor; however, non-normalized forms may be preferred, since they keep track of conflicts through the value of $\mathbf{m}(\emptyset)$.[15]

More details about the algebraic properties of Dempster–Shafer theory can be found in reference (16). In particular, it is shown that the Dempster's rule of combination is commutative and associative, which allows one to combine the available sources in any order.

Having computed the combined evidence functions (**m**, **Bel**, and **Pls**), the classification must be done according to a "decision criterion". Several decision rules, such as maximum of plausibility, maximum of belief or compromises [e.g. $\max(\mathbf{Bel}(A) + \mathbf{Pls}(A))$], have been proposed.[7,16] At the present time, the choice of criterion remains application dependent. However, in our case, where the only compound hypothesis considered is Θ, all the previously mentioned decision rules are equivalent criteria.

Table 1 summarizes Dempster–Shafer computation in the case of two singleton classes (F and $\bar{F}$). Each mass function is defined, for each hypothesis, on all image pixels. Mass values on F and Θ (x and y for "O", z and u for "R") are estimated during a learning phase obtained from monosource classification. The unforested area mass functions are given by the normalization condition: $\sum_{A \subseteq \Theta} \mathbf{m}(A) = 1$ [equation (1)]. In the general case, x and y (resp. z and u) can take any value between 0 and 1, provided that they satisfy the normalization condition imposed by equation (1): $0 \leq x + y \leq 1$ (resp. $0 \leq u + z \leq 1$).

2.3. Simple example: Data fusion of optical and radar images in the case of no imprecision in cloud detection (binary optical mass values)

In this section, we aim at providing a first illustration of Dempster–Shafer evidence theory through the application to data fusion of radar and optical images in the presence of a partial cloud cover.

We consider the simple ideal case where we assume there is no imprecision in cloud detection (including their shadows, where we consider the optical information not to be exploitable), and there is no classification error in the optical image for the unclouded pixels. Conversely, we assume a non-zero classification error in the radar image. Clearly, in such a case, the best data fusion result will be obtained by only considering the radar image on the parts of the image where optical information is missing (cloud and shadow). We call this classification rule $\mathscr{R}$: "if there is a cloud in the optical image, then take the decision given by the radar image classification, otherwise take the decision given by the optical image classification".

Table 1. Mass and belief functions for forested area detection using multisource (resp. subscript "O" and "R") images; $K = x(1 - z - u) + z(1 - x - y)$

	Forested areas (F)	Unforested areas ($\bar{F}$)	$\Theta = F \cup \bar{F}$
$\mathbf{m}_O$	x	$1 - x - y$	y
$\mathbf{m}_R$	z	$1 - z - u$	u
$\mathbf{m} = \mathbf{m}_O \oplus \mathbf{m}_R$	$\frac{x(z+u)+yz}{1-K}$	$\frac{(1-x)(1-z)-uy}{1-K}$	$\frac{uy}{1-K}$
Bel	$\frac{x(z+u)+yz}{1-K}$	$\frac{(1-x)(1-z)-uy}{1-K}$	1

Table 2. Optical-radar data fusion results in the simple case where there is no imprecision in cloud detection

Optical image class	$\mathbf{m}_O(F)$	$\mathbf{m}_O(\bar{F})$	$\mathbf{m}_O(\Theta)$	Radar image class	$\mathbf{m}_R(\Theta)$	Sign of $\mathbf{m}_R(F) - m_R(\bar{F})$	$\mathbf{Bel}(F)$	$\mathbf{Bel}(\bar{F})$	Data fusion class
$\bar{F}$	0	1	0	$\bar{F}$	u	−	0	1	$\bar{F}$
$\bar{F}$	0	1	0	F	u	+	0	1	$\bar{F}$
Θ	0	0	1	$\bar{F}$	u	−	$\mathbf{m}_R(F)$	$\mathbf{m}_R(\bar{F})$	$\bar{F}$
Θ	0	0	1	F	u	+	$\mathbf{m}_R(F)$	$\mathbf{m}_R(\bar{F})$	F
F	1	0	0	$\bar{F}$	u	−	1	0	F
F	1	0	0	F	u	+	1	0	F

Assuming that there are no errors in the optical image classification is equivalent to only considering binary mass values for the optical image model: $\forall A \in \{F, \bar{F}, \Theta\}$, $\mathbf{m}_O(A)$ is equal to 0 or 1 (with $\mathbf{m}_O(\Theta) = 1$ representing the case of pixels labeled "cloud" or "shadow" in optical classification). Conversely, the imprecision in the radar image classification result (due to non-null classification error) is modeled by taking $\mathbf{m}_R(\Theta) = u > 0$ for every pixel. Table 2 shows the results of Dempster–Shafer data fusion for these assumptions for each pixel label according to optical (or radar) monosource classification.

$\mathbf{Bel}(F)$ and $\mathbf{Bel}(\bar{F})$ values are deduced from Table 1 by replacing x and y by their values [respectively given by $\mathbf{m}_O(F)$ and $\mathbf{m}_O(\Theta)$]. In every case, we have $0 \leq K < 1$, which means that there is no totally conflicting case, even when the optical and radar classes are contradictory. This is due to the fact that a global ignorance factor on the radar image has been introduced by imposing $\mathbf{m}_R(\Theta) > 0$. The only case where the conflict is null ($K = 0$) is when $\mathbf{m}_O(\Theta) = 1$, which represents the case where no optical information is available, i.e. data fusion reduces to radar image classification. Finally, we note that the only pair of $(\mathbf{m}_O(F), \mathbf{m}_O(\Theta))$ values which is not present in Table 2 is (1, 1). Indeed, this represents an impossible case because of the mass normalization condition: $\sum_{A \subseteq \Theta} \mathbf{m}(A) = 1$ [equation (1)].

The last column of Table 2 shows the Dempster–Shafer data fusion results, the decision rule being equal to the maximum of belief over simple hypotheses. Comparing it to the first column (optical class) and fourth column (radar class), we see that, in the case of a binary optical image mass function, Dempster–Shafer data fusion is equivalent to the classification rule $\mathscr{R}$. However, as we will see in the next section, Dempster–Shafer modeling also allows the representation of more general cases, compared to those presented in this simple example. In particular, the assumption about no imprecision in cloud detection is quite unrealistic, as shown by the results presented in Section 4.

In fact, the problem is that cloud boundaries may have the same radiometric values as some unforested areas. Thus, in the following, we define clouds as a cluster of pixels that has "high" radiometric pixel values near its center, and "medium" radiometric pixel values near its borders. In particular, such a definition means that the cloud center may be detected by thresholding method, and that, in the case of pixels having "medium" radiometric values, we need spatial information to decide between cloud border or unforested area. Then, this cloud pixel modeling clearly illustrates the need of introducing spatial information in the evidence theory.

3. INTRODUCTION OF SPATIAL INFORMATION

In this paper, we propose to introduce neighborhood information in Dempster–Shafer evidence theory at two different levels: in the definition of monosource mass functions, and, in a final regularization step which is performed after the combination of the monosource mass functions. These two levels are both based on the classical Markovian field assumption, respectively, for the label images associated to the monosource images, and for the label image associated to the multisource data.

3.1. *Definition of monosource image mass functions*

We propose to consider two terms in our definition of the monosource image mass functions: a radiometric term $\mathbf{m}^r$, and a neighborhood term $\mathbf{m}^s$. The first one may be classically defined from pixel radiometric value distance to cluster center, or more simply, as we will do, from the monosource classification result.

For the neighborhood term, we find it reasonable to assume that the contribution of each neighbor r is a decreasing function of the distance d (in pixels) between r and s, the considered pixel. In our case, for a neighborhood window such that $d < d_{\max}$, a neighbor labeled i in monosource image classification increases $\mathbf{m}^s(i)$ by the value: $(1/Z_i)\,(1 - d/d_{\max})$, where Z_i is a parameter which defines the importance of the pixel influence over its neighbors versus its class i. For example, in our application, taking $Z_\Theta \leq Z_F$ and $Z_F = Z_{\bar{F}}$ allows one to give more importance to neighborhood information in the case of cloud neighbors than in the two other cases. In summary, we define $\mathbf{m}^s(i)$ by

$$\mathbf{m}^s(i) = \left\{ \sum_{r \in N_s,\, \xi_r = i} \frac{1}{Z_i}\left(1 - \frac{d}{d_{\max}}\right) \right\} \Big/ Z \qquad (6)$$

where ξ_r is the label of pixel r; N_s, the neighborhood of the considered pixel s, is such that $d < d_{\max}$; and Z is a normalization constant such that $\mathbf{m}^s$ satisfies equation (1).

At this stage, different combination rules of $\mathbf{m}^r$ and $\mathbf{m}^s$ may be envisaged: sum, product, Dempster orthogonal sum, etc. Here, we propose to simply sum them, before (re)normalizing the global mass function to satisfy equation (1). This is a compromise type of fusion[17] that takes into account neighborhood information, in the sense that when $\mathbf{m}^r(i) = 1$ and $\mathbf{m}^s(i) = 0$, their combination $\mathbf{m}(i)$ is equal 0.5. We choose it because we consider neighborhood information extracted from monosource image to be less reliable than information which will be derived after multisource data combination.

3.2. *Final regularization step*

Final regularization steps are based on the rather classical assumption that homogeneous label configurations are more likely, in the considered images, than inhomogeneous ones. Such methods generally use neighborhood information to derive a homogeneity index, which is introduced in the decision rule. For example, a very simple regularization method assigns to a pixel the label which occurs the most often in its neighborhood. In the same way, the Ising model, applied in the case of Markov Random Field (MRF) label images,[18] defines a neighborhood energy term to be proportional to the number of neighbors having the same label. Such regularization methods are generally well adapted to the case of isolated mis-classified pixels (surrounded by well-classified pixels); however, they may not be adapted to the cases of classification errors presenting structures or blocks. In such cases one solution would be to increase the neighborhood size, but in this case small image structures could be lost.

In this paper, we propose a new regularization method based on an iterative data fusion scheme, according to the Dempster combination rule, between the current label image mass function $\mathbf{m}_r$ and $\mathbf{m}$ (called the "data mass function"). In our application, $\mathbf{m}$ is the "SPOT-ERS" combined mass function. We choose $\mathbf{m}_r$ to be proportional to the number of neighbors having the same label:

$$\mathbf{m}_r(i) = \frac{1}{|N_s|} \cdot \sum_{r \in N_s} \delta(i, \xi_r) \qquad (7)$$

where $\delta(.,.)$ is the Kronecker symbol:

$$\begin{cases} \delta(i,i) = 1, \\ \delta(i,j) = 0, \quad \text{if } i \neq j. \end{cases}$$

We note that the use of the orthogonal sum supposes that we are more confident in the spatial mass function than we were in the previous section. This seems reasonable since the considered label image is now derived from multisource information, i.e. more reliable and complete information than monosource information.

The advantage of the Dempster–Shafer combination rule is that ignorance may be taken into account in the case where $\mathbf{m}(\Theta) \neq 0$. The importance of the neighborhood information is directly adapted to the level of available information from data images. This is illustrated on Fig. 1, which shows the decision areas between two classes A and $\bar{A}$ in the case where $\mathbf{m}_r(\Theta) = 0$, and only the classes A and $\bar{A}$ are considered in the decision rule [i.e., $\mathbf{Bel}(A) = \mathbf{m}(A)$ and $\mathbf{Bel}(\bar{A}) = \mathbf{m}(\bar{A})$]. If we denote by α the ratio between A and $\bar{A}$ data mass values: $\alpha = \mathbf{m}(A)/\mathbf{m}(\bar{A})$, then $\mathbf{m}(A) = (1/(1+\alpha))[1 - \mathbf{m}(\Theta)]$, $\mathbf{m}(\bar{A}) = (\alpha/(1+\alpha))$

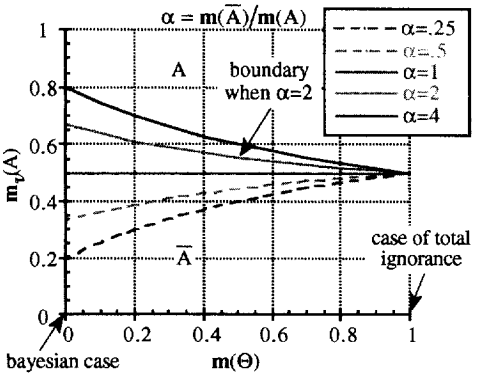


Fig. 1. Decision areas between A and $\bar{A}$, assuming $m_r(\Theta) = 0$, for different values of α ratio.

Table 3. Radiometric term of optical mass function

	Shadow	Forest	Certain unforest	Unforest or cloud	Certain cloud
$\mathbf{m}^r_O(F)$	0	1	0	0	0
$\mathbf{m}^r_O(\bar{F})$	0	0	1	0.5	0
$\mathbf{m}^r_O(\Theta)$	1	0	0	0.5	1

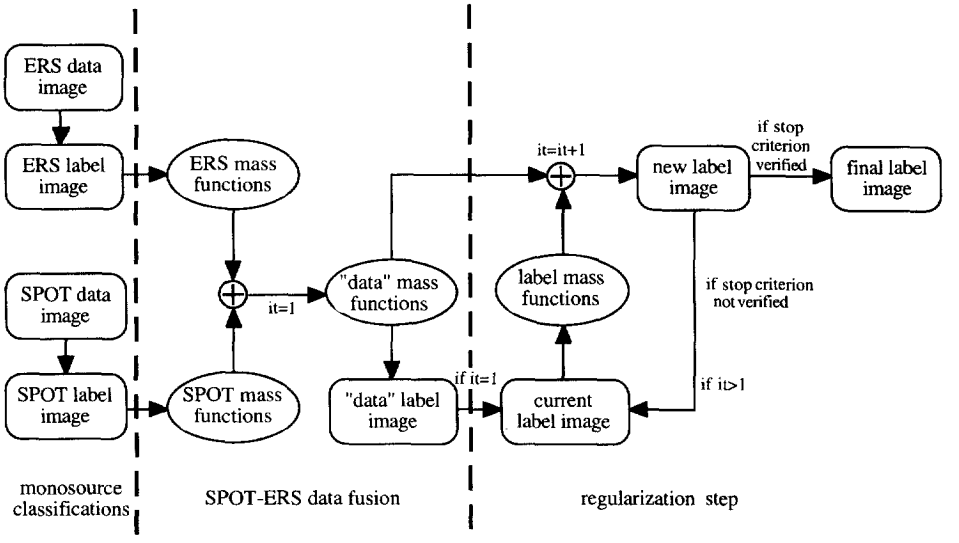


Fig. 2. Global data fusion algorithm divided in the three following steps: initialization (monosource classifications), data fusion, and regularization step.

$[1 - \mathbf{m}(\Theta)]$, $\mathbf{m}_\gamma(\bar{A}) = 1 - \mathbf{m}_\gamma(A)$, and the value of the border between the two decision areas is given by the ratio

$$\rho = \frac{1}{1+\alpha}\left[\frac{\alpha + \mathbf{m}(\Theta)}{1 + \mathbf{m}(\Theta)}\right].$$

In the Bayesian case, $\mathbf{m}(\Theta) = 0$, which is represented along the Y-axis, $\rho = \alpha/(\alpha + 1)$. This is the case where the necessary mass $\mathbf{m}_\gamma(A)$ to decide class A, whereas $\mathbf{m}(A)$ is inferior to $\mathbf{m}(\bar{A})$ (i.e. in the absence of neighborhood information the pixel would be labeled $\bar{A}$, $\alpha \leq 1$), is maximum. As $\mathbf{m}(\Theta)$ increases, the minimum value of $\mathbf{m}_\gamma(A)$ that is necessary to decide A decreases. For example, when $\alpha = 2$ and $\mathbf{m}_\gamma(A) = 0.6$, if $\mathbf{m}(\Theta) = 0.1$, the pixel will be labeled $\bar{A}$, but if $\mathbf{m}(\Theta) = 0.3$, it will be labeled A. Finally, when $\mathbf{m}(\Theta) = 1$, $\mathbf{m}(A) = \mathbf{m}(\bar{A}) = 0$ and $\rho = 0.5$, only spatial information is considered: decide A if $\mathbf{m}_\gamma(A) \geq \mathbf{m}_\gamma(\bar{A})$.

3.3. Application to SPOT-ERS data fusion in the presence of partial cloud cover

Monosource image radiometric mass functions $\mathbf{m}^r_O$ and $\mathbf{m}^r_R$ (where the subscripts "R" and "O" refer to optical and radar images) are derived respectively from monosource image classifications: in two classes {forest and unforested areas} for the radar image, and four classes {forest, unforested areas, cloud and shadow} for the optical one.

In the radar image case, for any pixel labeled F (resp. $\bar{F}$) in radar image classification $\mathbf{m}^r_R(F) = 1-u$, $\mathbf{m}^r_R(\bar{F}) = 0$ [resp. $\mathbf{m}^r_R(F) = 0$, $\mathbf{m}^r_R(\bar{F}) = 1 - u$)], and $\mathbf{m}^r_R(\Theta) = u$ (cf. Table 1, with $z \in \{0, 1-u\}$). As said in Section 2.1, choosing $\mathbf{m}^r_R(\Theta) \neq 0$ lets us represent the imprecision in radar image classification. Therefore, we choose $\mathbf{m}^r_R(\Theta)$ equal to the radar image error (about 0.3).

In the optical image case, there are no overlapping radiometric values between classes except between unforested areas and clouds. Therefore, assuming no classification error in the case of a cloud-free optical image, the radiometric term $\mathbf{m}^r_O$ of optical image mass function is chosen as shown in Table 3.

For the spatial term of the monosource mass function, the size of the neighborhood has been chosen equal to the mean size of cloud borders in our images, i.e. $\pm$ 4 pixels. Due to the special importance of neighborhood information in the case of clouds, we set $Z_\Theta = \frac{1}{2} Z_F$ and $Z_F = Z_{\bar{F}}$ (cf. Section 3.1).

Having defined optical image and radar image mass functions ($\mathbf{m}_R$ and $\mathbf{m}_O$) for every pixel, they are combined according to the Dempster combination rule: $\mathbf{m} = \mathbf{m}_R \oplus \mathbf{m}_O$. The first data fusion label image is then deduced according to maximum of belief (which is equivalent to maximum of plausibility in our case) over the singleton classes. Finally, the

regularization step is performed. The stop criterion used is the absence of change in the label image. In our case, the convergence of this last step has been empirically verified after about 10 iterations.

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

— initialization step: monosource classifications of each data set;
— mass function definition with the introduction of the neighborhood information deduced from monosource classification, and data fusion according to the Dempster combination rule;
— regularization step based on neighborhood information deduced from current data fusion result, and final classification.

Figure 2 illustrates this algorithm.

4. RESULTS AND DISCUSSION

The data fusion algorithm we described was applied to the problem of forested area discrimination using SPOT and ERS-1 images. The pixel resolutions of these two sensors are similar: respectively 20 m for the optical images and 12.6 for the radar images (case of 4-look data). In this section, we will consider two applications of the proposed data fusion method. In the first one, we apply it to monodimensional data: ERS image (C band, VV polarization) and SPOT channel 1 ($\lambda \in [0.50\text{–}0.59\ \mu m]$) image, with simulated cloud cover. In the second one, we consider multidimensional data: panchromatic SPOT image (two visible channels: "green" $\lambda \in [0.50\text{–}0.59\ \mu m]$ and "red" $\lambda \in [0.61\text{–}0.68\ \mu m]$, and a near infrared channel: $\lambda \in [0.79\text{–}0.89\ \mu m]$). Moreover, the images of this second set of data contain actual cloud cover.

4.1. Application to the case of simulated cloud cover

The first data set we consider corresponds to images acquired over a site near Canterbury (Kent, United Kingdom). They were respectively acquired in 1990 and December 1991. There was no cloud cover in the original SPOT image. Some simulated cloud cover has been added in the following way: actual clouds present in other SPOT images have been pasted over original pixels, and shadows have been simulated, given an illumination angle, by modifying the original pixel radiometry.[19] Figures 3(a) and (e) show the SPOT images with the different simulated cloud covers; respectively, equal to 12% and 33% of the whole image.

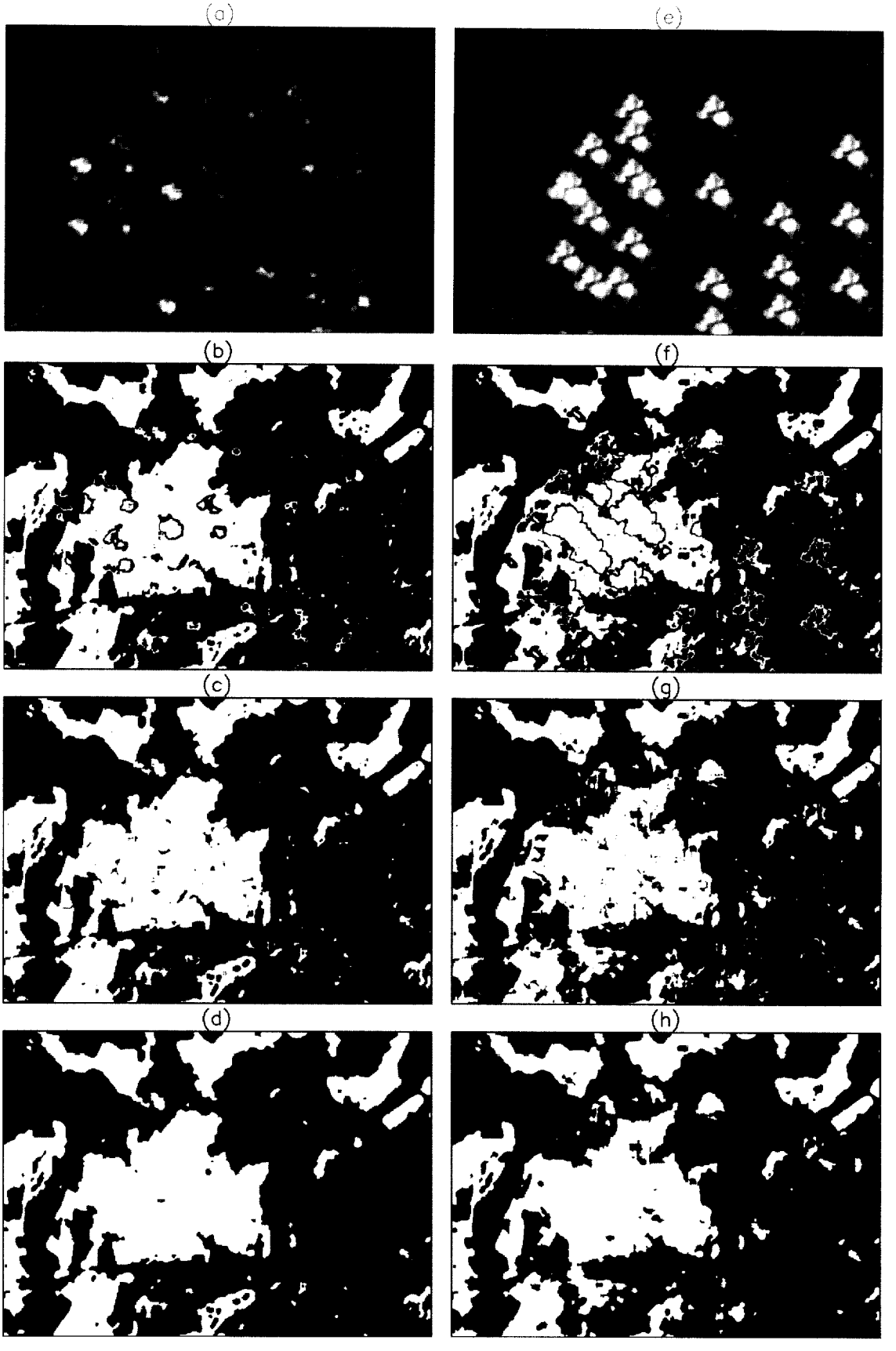
Figure 3 shows the classification results obtained by classification rule $\mathscr{R}$, by Dempster–Shafer data fusion (cf. Fig. 2) before regularization step (called "DS") and after by Dempster–Shafer data fusion regularization step (called "DSr"), for the two cases of cloud cover equal to 12% and 33%. We note that in both cases classification rule $\mathscr{R}$ fails at cloud borders because of their misdetection by the thresholding method. The improvement through data fusion based on the evidence theory is obvious comparing Fig. 3(b) [resp. 3(f)] to Fig. 3(c) [resp. 3(g)]. However, we note there are still some small structures of mis-classified pixels (classification noise). Applying the classical MRF regularization method (with Ising model), we found that these structures are too large to be cleaned. Conversely, after performing the regularization step we described in Section 3.2 with a neighborhood size equal to 5 pixels in line and in column, most of them disappear [cf. Fig. 3(d) and (h)]. Therefore, we conclude in favor of this new regularization method which is more flexible than the probabilistic model.

Similar results were obtained in the case of cloud cover equal to 66%, except that, the more significant is the cloud cover in SPOT image, the more the final classified image looks like ERS monosource classification. Figure 4 shows the percentage of well-classified pixels, versus the percentage of cloud cover in the image, again performing data fusion according to classification rule $\mathscr{R}$, and to Dempster–Shafer modeling with and without the regularization step. The performance of ERS monosource classification and the SPOT monosource classification have been indicated as references: they are equal to 72% for ERS image, and decreases from 92% to 40%, as a function of cloud cover, for the SPOT images. In this latter case, detected clouds are not classified.

Comparing the three data fusion methods we note the improvement, previously shown in Fig. 3, due to the introduction of spatial information in the evidence theory. However, we note that the number of corrected pixels is only a small percentage of the total number of image pixels. Moreover, the results do not improve monotonically as a function of the percentage of cloud cover, since the greatest improvement is achieved for a cloud cover equal to 33% rather than 66%. Actually, since classification rule $\mathscr{R}$ only fails at cloud borders, improvement is rather an increasing function of the percentage of cloudy pixels in SPOT images which are: (i) undetected according to the threshold value, and (ii) well-classified in the ERS-1 image classification. This second condition is set to eliminate the problem of ERS classification errors in the comparison of the data fusion algorithm performance.

Figure 5 shows the difference in percentage of well-classified pixels between evidential data fusion ("DS" or "DSr") and classification rule $\mathscr{R}$, versus cloudy pixels possibly corrected, i.e. satisfying (i) and (ii). We note that the percentage of actually corrected pixels relative to the percentage of possibly corrected pixels is about 40% before regularization step and about 50% after regularization step.

In summary, the efficiency of the data fusion method we described in Fig. 2 has been verified, and its superiority to classification rule $\mathscr{R}$ has been established on our data, both qualitatively (cf. Fig. 3) and quantitatively (cf. Figs 4 and 5). Particularly, it was shown that taking into account imprecision in cloud


(a)
(e)
(b)
(f)
(c)
(g)
(d)
(h)

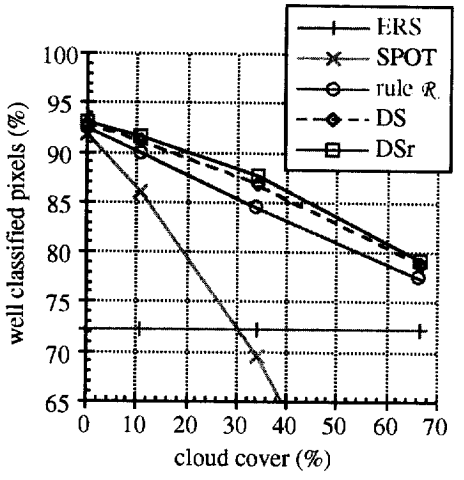


Fig. 4. Comparison between the three multisource classification algorithms: classification rule $\mathcal{R}$, Dempster–Shafer data fusion before and after regularization step (resp. denoted "DS" and "DSr").

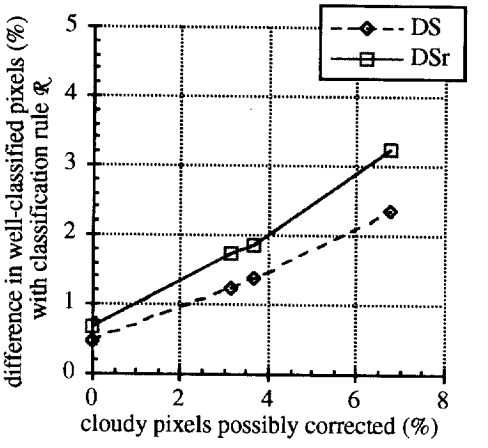


Fig. 5. Difference in percentage of well-classified pixels between data fusion algorithm of Fig. 2 with or without regularization step and classification rule $\mathcal{R}$, versus the percentage of cloud "borders" (difference between actual cloud cover and cloud class obtained as result of SPOT image monosource classification), possibly corrected using ERS image.

border detection allows partial correction of misclassifications due to cloud non-detection by the thresholding method. However, the cloud cover considered represents only a specific case of clouds: cumulus clouds. Therefore, we now aim at applying the data fusion method to some real data exhibiting more various kinds of clouds, such as low cirrus.

4.2. Application in the case of actual cloud cover

The second data set we used was acquired at the east of Paris (France) in summer 1996. We have at our disposal a panchromatic SPOT image, an ERS-1 image and a LandSAT image. There is cloud cover only in the SPOT image. The LandSAT image was used to create the ground truth map of forested and unforested areas. The three images have been projected in the same geometry (we chose the SPOT image as reference). Projection has been done by selecting the same reference points in each pair of images, and approximating the distortion between the images by a polynomial of degree two. Thus, the mean errors were between 1 and 2 pixels both in raw and in the column directions.

Figure 6(a) shows the channel 1 of the SPOT image. In Fig. 6(a), we can see a field of fair weather cumulus clouds with transition toward deeper convection, and probably some cirrus clouds, e.g. over the small rectangular forest area in the left of the image. As in the previous case, the study of histograms and pixel values shows that there is overlap between the range of cloud pixel radiometric values and the range of unforested area radiometric values. For example, in Fig. 6(a), we can see that there are some clouds over the forested areas having lower values than some fields. Using the three SPOT channels, we achieve classification error equal to 47.63%. Figure 6(b) shows the ERS image. The ERS image monosource classification error is equal to 15.01%. We note the improvement in ERS classification performance, since in the previous example classification error was about 30%. The main reason is that the image we now consider was acquired in the presence of vegetation (summer), which induces different backscattering mechanisms, and helps to distinguish between the different land cover types. However, we note that the forest class is overestimated by ERS classification, since it cannot distinguish between forested areas and some dense vegetation areas, such as corn fields.

The extension of the previously presented data fusion method to the case of multidimensional data is simply done by modifying the radiometric mass value computation during the learning process. For a given pixel, the radiometric term in the SPOT mass function is now defined according to its three radiometric values in the three different SPOT channels. In particular, the infrared channel is used to remove ambiguities between forest areas and cloud shadows, which are now present if we only consider the visible SPOT channels (contrary to the previous case of simulated cloud cover), such as at the top of the SPOT image [Fig. 6(a)], and ambiguities between roads [in the left part of Fig. 6(a)] or airport [in the right upper part of Fig. 6(a)] and clouds. The two images acquired in the visible SPOT channels show great redundancy. However, in the case of contradictory information, we used

Fig. 3. SPOT images with simulated cloud covers respectively equal to (a) 12% and (e) 33%, and corresponding data fusion results respectively obtained by (b and f) classification rule $\mathcal{R}$, the presented method (c and g) before regularization step and (d and h) after regularization step.

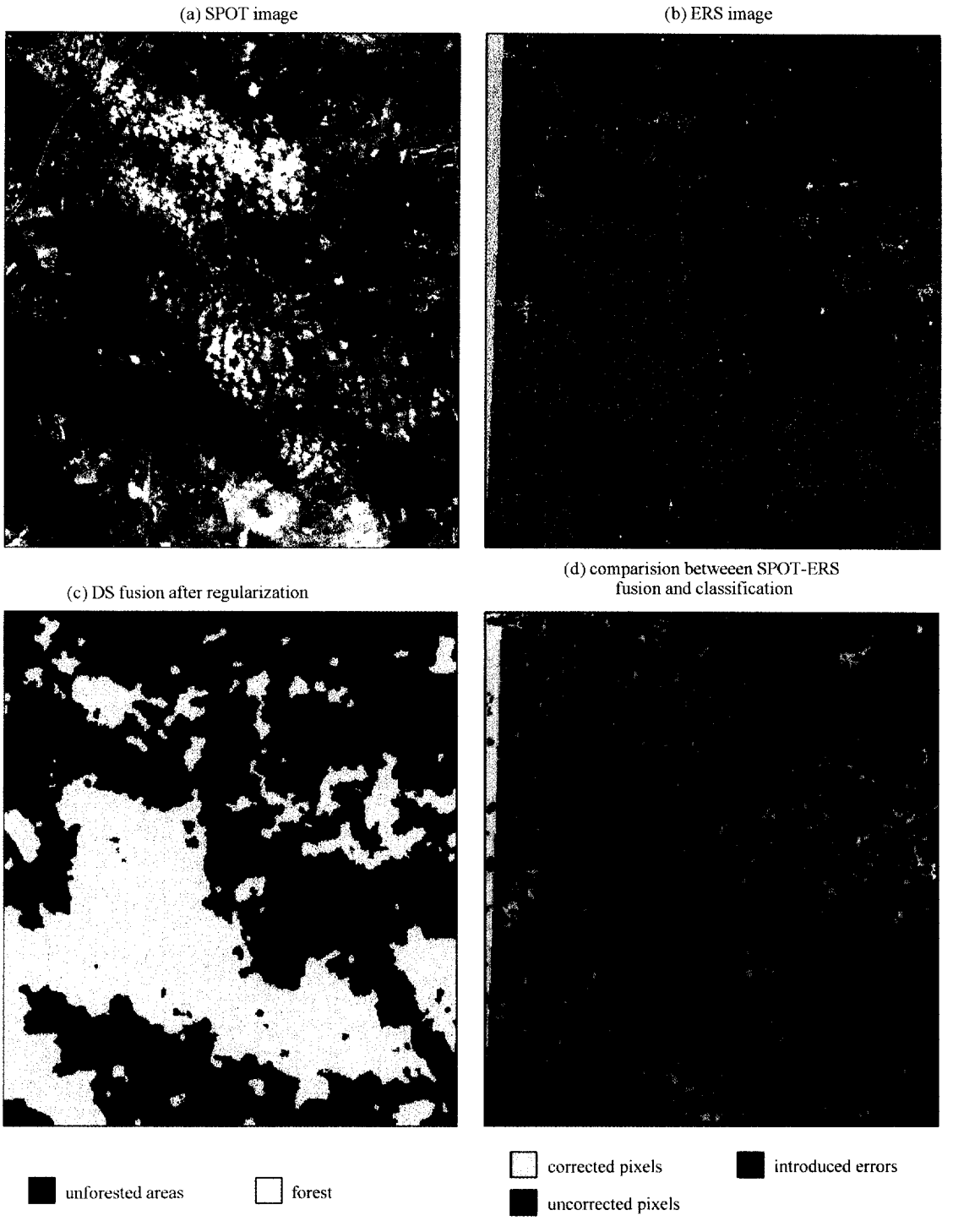


Fig. 6. (a) channel 1 of SPOT image with actual cloud cover, (b) ERS image, (c) data fusion result obtained by the described algorithm, and (d) comparison with ERS monosource classification.

a distance function to combine them, and to define the mass function. The neighborhood term of the SPOT mass is defined as previously, and combined with the radiometric one to provide the SPOT mass function $\mathbf{m}_O$.

The ERS mass function $\mathbf{m}_R$ is defined, as previously, from ERS image monosource classification. The imprecision in radar image classification results is taken into account by giving a non-zero mass to Θ, constant

and about equal to the radar image classification error, except on the left border of the image where no data are available [see Fig. 6(b)], where it is set equal to 1. Finally, we found that a slight improvement may be achieved in the results if we introduce a neighborhood term in the ERS mass function (like in the SPOT one). Indeed, radar images suffer from speckle, which induces some classification errors (which cannot be totally removed by performing the regularization step during radar image monosource classification).

Having defined the SPOT and ERS mass functions, data fusion is performed according to the algorithm of Fig. 2. Figure 6(c) shows the final classification result obtained after the regularization step (the size of the neighborhood being the same as in the previous case of application). The classification error is equal to 7.6%. It was equal to 11.1% before the regularization step. We note the improvement due to the regularization step, and the performance of the method in comparison with classification rule $\mathscr{R}$, which leads to a classification error of 18.0%. Comparing these results with ERS monosource classification, we note that the classification error has been reduced by a factor 2. Figure 6(d) shows a comparison between ERS classification and SPOT-ERS data fusion: the pixels which are well-classified in the two images are coded in black, the pixels mis-classified in ERS classification and well-classified by multisource classification are coded in yellow (called "corrected pixels" where "corrected" refers to data fusion), the pixels mis-classified by both ERS classification and multisource classification are coded in blue (called "uncorrected pixels"), and the pixels mis-classified by multisource classification when well-classified in ERS classification are coded in green (called "introduced errors"). First, we note that "corrected pixels" are much more numerous (they represent about 9.6% of the whole image) than the "introduced errors" ($\approx 2.2\%$ of the whole image), which results in a decrease of classification error. In particular, we recognize some yellow structures, e.g. near the left border of the image, which may be fields exhibiting high density vegetation cover, and were mis-classified using the ERS image alone. Then, we note that some mis-classified pixels by data fusion (represented either in blue or in green) are forest borders. These classification errors may be due to the imprecision in the image projection to a common geometry, as well as to the imprecision introduced in the outline by the regularization step. We also note that the other "introduced errors" concern a small forest area in the lower part of the image. Looking at the SPOT image, we see that it is under a quasi-transparent cirrus cloud, which modifies the pixel radiometry. In fact, this cloud remains undetected because it has no high radiometric pixel center. Therefore, this shows the limit of the described method, which fails in this case of a very thin cloud. Conversely, the cirrus cloud in the left of the image has been detected and the corresponding pixels well classified.

The "uncorrected pixels" represent about 5.4% of the whole image. However, we note that most of them are located under cloud cover in the SPOT image, and thus must not be considered in the performance analysis of the data fusion method (clearly, when no information is available from the optical image, data fusion is limited to the radar image classification performance). Among the other cases of "uncorrected pixels", we note the poor estimation of the boundary of the forested areas in the right part of the image, and the road which crosses the forest in the left low part of the image. This road was well classified (in unforested areas) before the regularization step, but then it was erased by the regularization.

In summary, the proposed data fusion method proved to be rather effective in the studied case of real data. However, its main limitation appears with the presence of transparent cirrus clouds.

5. CONCLUSION

For many applications of image classification problems, the information provided by a single sensor is incomplete or imprecise, resulting in misclassification. Data fusion can help reduce imprecision and provide a more complete description. The main advantage of Dempster–Shafer evidence theory for data fusion is that it provides a general framework to represent ignorance and imprecision. For example, in the described application, it was used to model the ignorance corresponding to the presence of the cloudy pixels in the optical image, and the imprecision introduced by the speckle in the radar image. In both cases, we assign a non-zero mass to Θ, the union of all classes, which provides us an option to not make a decision between the different classes.

In this paper, we propose to improve data fusion results by taking into account spatial neighborhood information. In our application, this was particularly needed because of the poor cloud border detection. Spatial information was introduced at two different levels:

- When defining a monosource mass function, for every pixel, a neighborhood term was defined. It was chosen to be proportional to the number of neighbors having the same label, and to be a decreasing function of pixel distance. Then, we use the addition law to combine this spatial term with a classically defined "radiometric" one, because it is derived from a monosource classification which may be erroneous.
- At the second level, we assume spatial information extracted from the data fusion label image to be much more reliable. We introduce this spatial information during an iterative regularization step. A spatial neighborhood mass function is defined from the current data fusion label image, and it is combined according to the Dempster combination law with the "blind" multisource mass

function (obtained before the regularization step). The advantage of using the Dempster combination law is that the importance of neighborhood information is automatically adapted to the level of ignorance at the considered pixel.

The presented method was applied to the problem of forested and unforested area detection, having at our disposal a radar image and an optical image showing a partial cloud cover. At each pixel, monosource mass functions have been defined from respective monosource classifications. Moreover, the non-null radar classification error has been modeled as a global ignorance factor, by assigning a constant non-zero mass to Θ. The proposed data fusion algorithm was applied to two different SPOT-ERS data sets. In both cases, the convergence of the regularization step was achieved after about ten iterations.

The first data set lets us compare the performance of the proposed algorithm to the simple classification rule $\mathscr{R}$ in the cases of four different simulated cloud covers (from 0% to 66%) in the SPOT image. It was shown that classification rule $\mathscr{R}$ leads to classification errors at cloud boundaries, due to their poor detection by thresholding methods, whereas using the evidential data fusion significantly improves the classification results. More precisely, we found that the observed improvement is an increasing function of the percentage of undetected cloudy pixels in the SPOT image, possibly corrected using ERS information, rather than a function of the percentage of clouds. Relative to these possibly corrected pixels, about a 50% correction was achieved. The second data set represents a case of actual cloud cover, with fair weather cumulus and cirrus clouds, which are much more difficult to detect than the simulated deeper convection cumulus. Performance was good, except in a special case of very thin cirrus clouds.

6. SUMMARY

In this paper, we propose to introduce spatial information in Dempster–Shafer evidence theory. This work concerns particularly pattern recognition and classification problems using incomplete data, such as in the presence of an hiding phenomena, or imprecise data, such as in the case of noisy data. Then, the main advantage of the evidence theory is that it provides a general frame work to represent the imprecision or incompleteness of each data set. Moreover, in such cases of missing radiometric information, the neighborhood information may be very useful. Therefore, we propose to introduce it. In the frame of the evidence theory, this was done at two different levels: before data fusion, in the definition of the monosource mass functions, and, during data fusion. In the latter case, a "neighborhood" mass function is derived from the current label image and combined according to the Dempster orthogonal sum with the "data" mass function representing radiometric information. We show that the main advantage of such a combination law is to adapt the importance of neighborhood information to the level of radiometric missing information.

All along the paper, we consider the following application to illustrate our purpose: data fusion of radar and optical images showing partial cloud cover. First, we show how the Dempster–Shafer evidence theory may be successfully used to model the partial cloud cover (in optical images) in terms of missing information, and the information imprecision, due to the speckle noise, in the SAR images. Then, because of cloud detection inaccuracy, we introduce spatial neighborhood information. This data fusion algorithm has been applied to the problem of forested and unforested area detection, and its performance is stated both in the cases of different simulated cloud cover and in the case of real data.

Acknowledgements—We thank Matra Cap System company for providing the first data set used, and Professor B. Denby for his proofreading. We also thank Professor S. Hutchinson for carefully reading the paper and improving the English.

REFERENCES

1. M. Abidi and R. Gonzalez, *Data Fusion in Robotics and Machine Intelligence*. Academic Press, New York (1992).
2. D. Leckie, Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination. *Photogrammetric Engineering and Remote Sensing* **56**, 1237–1246 (1990).
3. J. Benediktsson, and P. Swain, A method of statistical multisource classification with a mechanism to weight the influence of the data source, *Proc. Int. Geoscience Remote Sensing Symp. (IGARSS)*, Vol. 2 517–520. Vancouver, Canada (10–14 July 1989).
4. I. Kanellopoulos, G. Wilkinson, and A. Chiuderi, Land cover mapping using combined LandSAT TM imagery and textural features from ERS-1 synthetic aperture radar imagery, *Proc. SPIE: Image Signal Process. Remote Sensing*, Vol. 2315, 332–341. Rome, Italy (26–30 September 1994).
5. A. Schistad Solberg, A. Jain, and T. Taxt, Multisource classification of remotely sensed data: fusion of LandSAT TM and SAR images, *IEEE Transactions on Geoscience and Remote Sensing* **32**(4), 768–778 (1994).
6. A. Schistad Solberg, T. Taxt, and A. Jain, A Markov random field model for classification of multisource satellite imagery, *IEEE Trans. Geoscience Remote Sensing* **34**(1), 100–113 (1996).
7. G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ (1976).
8. H. Rasoulian, W. Thompson, L. Kazda, and R. Parra-Loera, Application of the mathematical theory of evidence to the image cueing and image segmentation problem, *Proc. SPIE: Signal and Image Processing Systems Performance Evaluation*, Vol.1310, pp. 199–206. Orlando, Florid, (19–20 April 1990).
9. J. van Cleynenbreugel, S. Osinga, F. Fierens, P. Suetens, and A. Oosterlinck, Road extraction from multitemporal satellite images by an evidential reasoning approach, *Pattern Recognition Letters* **12**, 371–380 (1991).
10. I. Bloch, Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account, *Pattern Recognition Letters* **17**(8), 905–916 (1996).

11. T. Lee, J. Richards, and P. Swain, Probabilistic and evidential approaches for multisource data analysis, *IEEE Trans. Geoscience Remote Sensing* **25**(3), 283–293 (1987).
12. S. Le Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar, Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing, *IEEE Trans. Geoscience Remote Sensing* **35**(4), 1018–1031 (1997).
13. T. Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, *IEEE Trans. Systems, Man, Cybernet.* **25**(5), 804–813 (1995).
14. N. Jiordana, and W. Pieczynski, Utilisation de la théorie de l'évidence pour la restauration de chaînes de Markov cachées multispectrales, *Proc. GRETSI*, Grenoble, France (15–19 September 1997).
15. P. Smets, The combination of evidence in the transferable belief model, *IEEE Trans. Pattern Analysis Machine Intelligence* **12**(5), 447–458 (1990).
16. J. Guan, and D. Bell, *Evidence Theory and its Applications*, North-Holland. North-Holland, New York (1991).
17. I. Bloch, Information combination operators for data fusion: a review with classification, *IEEE Trans. Systems, Man, Cybernet.* **26**(1), 52–67 (1997).
18. S. Geman, and D. Geman, Stochastic relaxation, Gibbs distribution and Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intelligence* **6**(6), 721–741 (1984).
19. L. Peytavin, F. Dansaert, and A. Sephton, Une méthode robuste de classification multi-source utilisant des processus aléatoires Markoviens, *Proc. Int. Conf.: from optics to radar, SPOT and ERS applications*, 165–173. Paris, France (May 1993).

About the Author—SYLVIE LE HÉGARAT-MASCLE is graduated from the *Ecole Nationale Supérieure des Télécommunications* (ENST), Paris, France, in 1993. She received her Ph.D. in signal and image from ENST in 1996. She is working at 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.

About the Author—ISABELLE BLOCH is Associate Professor at ENST Paris (Department Images). She graduated from *Ecole des Mines de Paris* in 1986, received Ph.D. from ENST Paris in 1990 and the "Habilitation à Diriger des Recherches" from University Paris 5 in 1995. Her research interests include 3D Image and Object Processing, 3D 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.