16

Three Strategies for Fusion of Land Cover Classification Results of Polarimetric SAR Data

Nada Milisavljević[1], Isabelle Bloch[2], Vito Alberga[1] and Giuseppe Satalino[3]
[1]*Signal and Image Centre, Royal Military Academy Brussels,*
[2]*TELECOM ParisTech (ENST) CNRS UMR 5141 LTCI Paris,*
[3]*Institute of Intelligent Systems for Automations (ISSIA)*
National Research Council (CNR) Bari,
[1]*Belgium*
[2]*France*
[3]*Italy*

1. Introduction

A fully polarimetric synthetic aperture radar (SAR) is a device that is able to transmit and receive both orthogonal (horizontal and vertical) components of an electromagnetic wave (Touzi et al., 2004). Its signals include the magnitude and phase information, rendered as complex variables. Many classification algorithms have been proposed for SAR images (Cloude & Pottier, 1997; Ferro-Famil et al., 2001; Hoekman & Vissers, 2003; Kersten et al., 2005; Lee et al., 1994; Lee et al., 1999, Lee et al., 2004), that can be grouped into three main types (Chen et al., 2003): 1) algorithms based on image processing techniques, 2) algorithms based on a statistical model, 3) algorithms based on the scattering mechanism of the electromagnetic waves. Our interest is in the first type, since such algorithms permit a general approach to the images, potentially after investigating the physical properties of natural media.

In a previous work (Alberga et al., 2006), several ways of representing polarimetric SAR data (Cloude & Pottier, 1996; Cameron et al., 1996; Cloude & Pottier, 1997; Freeman & Durden, 1998) have been analyzed and their usefulness for land cover classification compared. The classifiers used were the minimum distance classifier, the maximum likelihood classifier and a neural network - the Multi-Layer Perceptron (MLP), trained by the Back-Propagation (BP) learning rule. The MLP outperformed the other two classifiers. In addition, the MLP does not need any *a priori* knowledge on the statistics of the input data, thus it can be applied to any possible polarimetric observable, permitting an unbiased use of the classification results (Benediktsson, et al., 1990). For these reasons, the MLP is used also in this work.

The accuracy of the classification of a given polarimetric representation (*i.e.*, set of polarimetric parameters) was taken in (Alberga et al., 2006) as a measure of the usefulness of that set and compared with the accuracies obtained using other representations of the data. The substantial equivalence of the parameters in these terms was shown. However, no attempt was made to take advantage of the complementary information provided by the

Source: Sensor and Data Fusion, Book edited by: Dr. ir. Nada Milisavljević,
ISBN 978-3-902613-52-3, pp. 490, February 2009, I-Tech, Vienna, Austria

different parameters nor was their fusion performed in order to improve the classification performance. This is the scope of the present research.
Papers on polarimetric data fusion can be rarely found in the literature. In (Mascle et al., 1997; Le Hégarat-Mascle et al., 1998), an unsupervised classification is performed on each image separately and the results are fused using a strategy based on belief functions (Shafer, 1976; Smets, 1990).
In this paper, classification results coming from several sets of polarimetric parameters are fused following different strategies (SMART, 2004), two based on the belief function framework (Shafer, 1976; Smets, 1990) and one based on fuzzy theory (Dubois et al., 1999). After decision combination on pixel level, as a final step of the fusion module, a spatial regularization is performed.
In Section 2, the sets of polarimetric parameters under consideration are described, and the classifier used is presented in Section 3. The main aspects of the belief function theory are underlined in Section 4. The applied fusion approaches are detailed in Section 5. Section 6 presents the experimental data and the characteristics of the analyzed scene, while Section 7 reports on the results of data fusion and on their comparison with the classification results of each set of polarimetric parameters. Finally, Section 8 is devoted to conclusions, followed by acknowledgements and references.

2. Investigated polarimetric parameters

Fully polarimetric radars can transmit and receive both orthogonal components of an electromagnetic wave (Touzi et al., 2004). Thus, its vector nature is taken into account ensuring that complete scattering information carried by radar echo signals may be used for target detection and identification. Within this framework, different representations exist of the scattering interactions, such as, *e.g.*, the 2 x 2 scattering matrix [**S**] or the higher order ones, the coherency and covariance matrices.
In the field of terrain classification, the choice of a given representation has been related to considerations on both the statistics of the data and the physics of the scattering mechanisms. In particular, the use of incoherent parameters (*i.e.,* those derived from the second order matrices) has become predominant with respect to that of the coherent ones related to the [**S**] matrix. For multi-look data represented as covariance or coherency matrices, Lee et al. (Lee et al. 1994) defined a distance measure for the membership of a pixel to a class based on the complex Wishart distribution and this measure could be incorporated in several classification algorithms (Ferro-Famil et al., 2001; Lee et al. 1999). Accordingly, only second order representations were considered when operating these classifiers (Ferro-Famil et al., 2001; Lee et al. 2004); not much attention was paid, until now, to coherent parameters.
Different viewpoints are adopted when choosing a coherent or incoherent representation: in the first case, the hypothesis is made that the scattering interaction within a resolution cell involves only one or few point scatterers. Their phase can then be measured and analyzed or taken into account when deriving new parameters. Moreover, in the case of independent scattering mechanisms, these may be singularly recognized (coherent target decomposition theorems perform just this separation). Given their definition, the reliability of coherent methods is higher when dealing with man-made artifacts, which correspond better to such a scenario. Thus, in our specific experimental case, we expect good classification performance of coherent representations especially for urban areas or other targets with a "stable"

behavior. On the contrary, incoherent methods perform immediately an averaging of the returned signals that yields the loss of the direct reference to the phase of the elementary targets. However, they provide a statistically sounder description of the behavior of "dynamic" natural environments. For these environments, target decomposition theorems may be applied and permit to recognize different scatterers. Namely, they separate the responses of different mechanisms considered as "average" ones (*e.g.*, within forested areas they can distinguish between distributed volumes and surface scatterers).
Since coherent representations are better suited to targets approximating ideal point scatterers (as it happens in urban areas) and incoherent ones to randomly distributed targets (forests or fields), it can be expected that their use would provide complementary information that, by classification applications, could be better exploited by means of a fusion approach.
The representations reported in the following subsections have been taken into account in order to benefit from the different types of information they provide as well as because they are the most often used and known.

2.1 First and second order matrices

When a horizontally or vertically polarized wave is incident upon a target, the backscattered wave can have contributions in both horizontal and vertical polarizations. Thus, the backscattering of the target can be completely described by a scattering matrix:

$$[\mathbf{S}] = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}. \tag{1}$$

In other words, the four complex elements of the scattering matrix describe the transformation of the polarization of the incident wave to the polarization of the backscattered wave. For monostatic configurations (the ones in which the receive antenna is co-located with the transmit antenna), the scattering matrix becomes symmetric, *i.e.*, $S_{hv} = S_{vh}$.
The real and imaginary parts of the three complex terms of the scattering matrix provide six variables in total that can be reduced to five independent parameters by normalization with respect to a given phase term. For our tests, we normalize with respect to the phase of the *hh* term and the results of the classification, performed using this representation, are indicated as *cl1* in the remaining part of the chapter.
Real systems involve scatterers situated in dynamic environments and subject to space and/or time variations (*i.e.*, non-deterministic scatterers). This causes the electromagnetic waves to be partially polarized and thus prevents the scattering process from being described by a single matrix [**S**]. Hence, averaging processes are needed to cope with the statistical variation of the polarization. The covariance and coherency matrices, the definition of which includes such averaging, take into account these variations and permit their description.
In the monostatic case, the 3 x 3 covariance matrix has the form:

$$[\mathbf{C}_{(3)}] = \begin{bmatrix} \langle |S_{hh}|^2 \rangle & \sqrt{2}\langle S_{hh} S_{hv}^* \rangle & \langle S_{hh} S_{vv}^* \rangle \\ \sqrt{2}\langle S_{hv} S_{hh}^* \rangle & 2\langle |S_{hv}|^2 \rangle & \sqrt{2}\langle S_{hv} S_{vv}^* \rangle \\ \langle S_{vv} S_{hh}^* \rangle & \sqrt{2}\langle S_{vv} S_{hv}^* \rangle & \langle |S_{vv}|^2 \rangle \end{bmatrix}, \tag{2}$$

with its elements derived from the ones of the scattering matrix. By definition, the covariance matrix is Hermitian positive semidefinite, hence, its symmetric elements are complex conjugates and only nine independent parameters are necessary in order to completely characterize it. For classification, these nine parameters (the three real main diagonal elements and the real and imaginary parts of the three non-redundant off-diagonal elements) have been given as input to the classifier. In the following, the MLP output results based on the covariance matrix elements are denoted as *cl2*.
Alternatively, an incoherent representation is provided by the coherency matrix, which is defined as:

$$\left[\mathbf{T}_{(3)}\right]=\frac{1}{2}\begin{bmatrix} \left\langle\left|S_{hh}+S_{vv}\right|^2\right\rangle & \left\langle(S_{hh}+S_{vv})(S_{hh}-S_{vv})^*\right\rangle & 2\left\langle S_{hv}^*(S_{hh}+S_{vv})\right\rangle \\ \left\langle(S_{hh}+S_{vv})^*(S_{hh}-S_{vv})\right\rangle & \left\langle\left|S_{hh}-S_{vv}\right|^2\right\rangle & 2\left\langle S_{hv}^*(S_{hh}-S_{vv})\right\rangle \\ 2\left\langle S_{hv}(S_{hh}+S_{vv})^*\right\rangle & 2\left\langle S_{hv}(S_{hh}-S_{vv})^*\right\rangle & 4\left\langle\left|S_{hv}\right|^2\right\rangle \end{bmatrix}. \quad (3)$$

A direct interpretation of the measured values is possible, *e.g.*, in terms of the scattering model to be adopted (and this depending, in turn, on the surface roughness). The predominance of one [**S**] matrix term over the other is connected to the type of scattering on the illuminated surface (Ulaby et al. 1982; Born & Wolf, 1985; Curlander & McDonough, 1991). More precisely, $|S_{hh}|>|S_{vv}|$ when the incident beam is scattered according to the Fresnel model, valid for almost flat surfaces, whereas the case of $|S_{hh}|<|S_{vv}|$ is verified for scattering from rough surfaces described by the Bragg model.

2.2 Target decomposition theorems

Target decomposition (TD) theorems permit to identify different scattering mechanisms corresponding to sets of theoretical models (Cloude & Pottier, 1996; Corr & Rodrigues, 2002; Moriyama et al., 2004). This further means that these methods intrinsically perform a classification, since they recognize and weight the contributions of different model targets in a scene.
We have applied here the principal decomposition theorems: the Pauli and the Cameron decomposition, as examples of the coherent methods (that operate on the scattering matrix), and the Freeman decomposition, as an example of the incoherent ones (based on the covariance and coherency matrices).
The classification results corresponding to the Pauli decomposition are referred to as *cl3*. By means of the set of the Pauli matrices, it is possible to write a generic matrix

$$\begin{aligned}[\mathbf{S}]&=\begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}=\begin{bmatrix} a+b & c-jd \\ c+jd & a-b \end{bmatrix} \\ &=a\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}+b\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}+c\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}+d\begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}'\end{aligned} \quad (4)$$

where a, b, c and d are complex numbers. The first decomposition term represents single scattering from a plane surface or a sphere, the second and third term correspond to double-bounce scattering from diplane reflectors with a relative orientation of 45°, and the fourth term - to a scatterer that rotates every incident polarization by 90°. As it causes [**S**] to be non-symmetric, the fourth term disappears in reciprocal backscattering cases.

The Cameron decomposition (Cameron et al., 1996) (*cl4*) is a more generalized example of the model fitting seen with the previous decomposition. A generic matrix [**S**] (not only in the monostatic case) can be characterized by its tendency of being more or less symmetric according to the reciprocity rule and it can be split into two terms representing reciprocal and non-reciprocal scattering mechanisms. The reciprocal term represents a target which is more or less symmetric with respect to an axis in the plane orthogonal to the radar line-of-sight and, again, a distinction can be made between the most and the least dominant symmetric target components.[1] Thus, the decomposition follows the scheme (Cameron et al., 1996):

$$[\mathbf{S}] \begin{cases} \rightarrow [\mathbf{S}]_{rec} \begin{cases} \rightarrow [\mathbf{S}]_{sym}^{max} \\ \rightarrow [\mathbf{S}]_{sym}^{min} \end{cases} \\ \rightarrow [\mathbf{S}]_{non-rec} \end{cases}$$

The degrees of reciprocity and symmetry are evaluated in terms of projection angles of the scattering vectors onto the corresponding subspace and subsets via proper projection operators. As SAR data are calibrated in order to fulfill reciprocity constraints, the basic distinction among scatterers made by this method is based on their geometrical symmetry. For this reason, only the decomposition of the [**S**] matrix into its most dominant and least dominant symmetric terms, $[\mathbf{S}]_{sym}^{max}$ and $[\mathbf{S}]_{sym}^{min}$, has to be performed.

The principle behind incoherent decomposition theorems consists in modeling the scattering interaction so that the received power may be expressed as sum of contributions due to different basic mechanisms. As seen above, similar modeling based on the [**S**] matrix yields normally combinations of terms where typical scatterers, like spheres and diplanes, may be recognized. In (Freeman & Durden, 1998), another method is presented, less bound to pure mathematical models and more to real scatterers. The Freeman decomposition (*cl5*) describes the scattering as due to three physical mechanisms: first-order (surface) scattering, *s*, a double-bounce scattering mechanism (corner reflector), *d*, and canopy (or volume) scattering from randomly oriented dipoles, *v*. According to this model, the measured power *P* may be finally expressed as (Freeman & Durden, 1998):

$$P = \left\langle \left|S_{hh}\right|^2 \right\rangle + \left\langle \left|S_{vv}\right|^2 \right\rangle + 2\left\langle \left|S_{hv}\right|^2 \right\rangle = P_s + P_d + P_v \,, \tag{5}$$

with the three decomposition terms being related via a system of linear equations to the covariance matrix elements.

[1] Note the difference in the use of the word ``symmetry'' when referred to scattering matrices and to targets. According to the Cameron decomposition, scattering matrices which are symmetric due to the reciprocity constraint may describe targets which are geometrically more or less symmetric in the plane orthogonal to the radar line-of-sight (in the case of a helix, a symmetric scattering matrix represents a target which is not geometrically symmetric).

2.3 Entropy/α analysis

A generic coherency matrix may be diagonalized and decomposed by determining its eigenvalues and eigenvectors (Cloude, 1986; Cloude & Pottier, 1996; Cloude & Pottier, 1997). Using this decomposition, the different relevance of each scattering mechanism (within a given resolution cell) is expressed by means of its eigenvalues. Indeed, while the eigenvectors discriminate the presence of different scattering mechanisms, the eigenvalues underline their intensity. A quantity that measures the randomness of these scattering processes is the polarimetric scattering entropy, H:

$$H = \sum_{i=1}^{3} - P_i \log_3 P_i \,, \tag{6}$$

where:

$$P_i = \frac{\lambda_i}{\sum_{j=1}^{3} \lambda_j} \,. \tag{7}$$

The three λ_i are the calculated eigenvalues, conventionally ordered such that $0 \le \lambda_3 \le \lambda_2 \le \lambda_1$, and P_i represents the appearance probability of each contribute. H ranges from 0 to 1: $H = 0$ stands for a deterministic scattering process (the coherency matrix has only one non-zero eigenvalue), while $H = 1$ indicates a degenerated eigenvalues spectrum, typical of random noise processes (the coherency matrix has three identical eigenvalues).
To estimate the relative importance of the different scattering mechanisms, the polarimetric anisotropy (A) has been introduced:

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \,. \tag{8}$$

A medium entropy means that more than one single scattering mechanism contributes to the backscattered signal, but it is not clear how many additional mechanisms are present (one or two). In this case, a high A states that only the second scattering mechanism is important, whereas a low A indicates a remarkable contribute also of the third one.
Finally, a further parameter, the α angle, may be derived from the coherency matrix eigenvectors, which is associated to the type of scattering mechanism and can vary in the range $[0, \pi/2]$. $\alpha = 0$ stands for isotropic surfaces, $\alpha = \pi/2$ for isotropic diplanes or helices. Low values of α represent all-anisotropic scattering mechanisms with S_{hh} different from S_{vv}. The boundary between anisotropic surfaces and diplanes is represented by the case $\alpha = \pi/4$, which describes a horizontal dipole. An average $\overline{\alpha}$ angle is normally used in polarimetric SAR data analysis. Henceforth, the classification results of the $H/\overline{\alpha}/A$ parameters are referred to as *cl6*.

3. Classification algorithm

The neural network classifier chosen in this work is the Multi-Layer Perceptron architecture, with one hidden layer, trained by the Back-Propagation learning rule. The MLP is a fully connected feed-forward neural network, composed of nodes arranged in layers. It can be used to perform every non-linear input-output mapping, such as classification functions

(Hertz et al., 1991), or more complex tasks such as the approximation of continuous functions (Funahashi, 1989). For this purpose, it is necessary to submit the MLP to a training phase that searches the optimum set of weights minimizing a cost measure, usually given by the mean square error between estimated and expected outputs. This training phase, performed by the well-known BP learning rule (Hertz et al., 1991), requires a set of input-output examples. In this context, the input examples are obtained by a supervised procedure that identifies on the mono- or multi-band image, and for each class of interest, a corresponding region of points. Consequently, the output examples are given by the class labels j, j=1, 2, …, n (n being the number of classes), represented by n-dimensional Boolean vectors. The MLP dimension to be used is related to the current classification problem. In particular, the number of MLP input nodes corresponds to the number of polarimetric features under investigation, whereas the number of output nodes is set equal to the number n of classes identified on the images. The performance of each trained MLP is estimated on an independent test data set.

The same MLP classification procedure is applied to each of the six sets of polarimetric parameters described in Section 2, and the classification results are referred to as *cl1* - *cl6* (Table 1). The results are good for some sets and exhibit lower performances for other sets, in terms of global accuracy (see Section 7). However, the overall accuracy does not completely describe the behavior of the classifier. Looking more precisely at the performances for each class, it appears that even a globally good set of parameters can lead to classification results for a particular class which are worse than the ones obtained using another set of parameters. The main reason is that classes may be well separated in some polarimetric representations and not in other ones, and there is no single set of parameters that separates correctly all classes. These observations advocate for a fusion of all classification results obtained on the different sets of polarimetric parameters. An important contribution of this paper is to show how this fusion can be performed using the belief function theory.

Classification	Corresponding set of parameters
cl1	scattering matrix
cl2	covariance matrix
cl3	Pauli decomposition
cl4	Cameron decomposition
cl5	Freeman decomposition
cl6	$H/\overline{\alpha}/A$ parameters

Table 1. Set of parameters used to obtain each classification

4. On belief functions

Belief function theory or Dempster-Shafer (DS) evidence theory has been already widely used in satellite image processing (Mascle et al., 1997; Le Hégarat-Mascle et al., 1998; Cleynenbreugel et al., 1991; Tupin et al. 1999; Milisavljević & Bloch, 2003). DS theory allows representing both imprecision and uncertainty, using plausibility and belief functions derived from a mass function. The mass of a proposition A is a part of the initial unitary amount of belief that supports that the solution is exactly in A. It is defined as a function m from 2^{Θ} into [0, 1]. Θ is the decision space (frame of discernment) and it is a set of N possible

solutions, *e.g.*, classes C_i, $i = 1, 2, ..., N$ ($\Theta = \{C_1, C_2, \ldots, C_N\}$), while the power set, denoted 2^Θ, consists of 2^N subsets of Θ:

$$2^\Theta = \{\varnothing, C_1, C_2, ..., C_N, \{C_1 \cup C_2\}, ..., \Theta\}, \tag{9}$$

meaning that it contains not only single hypotheses (singletons) of Θ, but also all possible unions of the singletons, called compound hypotheses or disjunctions. Thus, in this formalism, any combination of possible decisions from the decision space can be quantified rather than considering only the singletons of Θ (Shafer, 1976; Smets, 1990). This is one of the main advantages of the DS approach, as it leads to a very flexible and rich modeling, able to fit a large class of situations, occurring in image fusion in particular. Another advantage of this method over the probabilistic ones is in allowing an easy way of representing the state of a total ignorance by means of the so-called vacuous basic belief assignment: $m(\Theta) = 1$, $m(A) = 0$, for all $A \neq \Theta$. The basic difficulty that some other theory, such as probabilistic, faces in these cases is the inability of distinguishing between lack of belief and disbelief (Shafer, 1976).

In image processing, mass functions may be derived at three different levels: the most abstract level, an intermediate level, and the pixel level. At the pixel level, which is the most interesting here, mass assignment is inspired from statistical pattern recognition. The most widely used approach is as follows: masses on simple hypotheses are computed from probabilities or from the distance to a class center (Appriou, 1993; Denœux, 1995; Appriou, 1998). Then a global ignorance $m(\Theta)$ is introduced as a discounting factor, often as a constant on all pixels (Lee & Leahy, 1990). In most cases, no other compound hypothesis is considered, and this drastically under-exploits the power of DS. The mass assignment in (Bloch, 1996) is based on a reasoning approach where knowledge about the information provided by each image is used to choose the focal elements (*i.e.*, subsets with non-zero mass values). A similar reasoning is used in (Milisavljević & Bloch, 2001). However, in the case of large numbers of classes, this process would become too tedious, and unsupervised methods are needed, such as (Mascle et al., 1997) for SAR imaging or (Ménard et al., 1996) for fusion of several classifiers.

If we have evidence issued from M sources, modeled in terms of previously defined mass functions, these masses are combined applying Dempster's rule of combination (Shafer, 1976; Smets, 1993). This rule has two main forms, normalized (Shafer, 1976) (by imposing $m(\varnothing) = 0$), and unnormalized (Smets, 1993), corresponding to the closed-world and the open-world assumptions, respectively. For m_j being the mass function associated with source j ($j = 1, 2, ..., M$), the unnormalized Dempster's rule of combination is:

$$\forall A \in 2^\Theta, \quad (m_1 \oplus ... \oplus m_M)(A) = \sum_{B_1 \cap ... \cap B_M = A} m_1(B_1) \cdot m_2(B_2) \cdot ... \cdot m_M(B_M) \,, \tag{10}$$

hence preserving the mass assignment to the empty or zero set, that indicates the strength of the possibility that "something else" happens (either the full set is not an exhaustive set of hypotheses or there is some contradiction between the sources, *e.g.*, some of them are not reliable).

Dempster's rule of combination is commutative and associative. Also, it behaves in a conjunctive way, meaning that when more sources are combined, mass is more focalized

(thus imprecision decreases) while conflict behaves in a disjunctive way (increases) (Bloch, 1996).
Sources under combination have to fulfill the condition of being independent in the cognitive sense (Shafer, 1976; Smets, 1993). This is related to the notion of distinctness, meaning "no double counting" of pieces of evidence (Dempster, 1967). In case of non distinct pieces of evidences, other combination rules should be employed, such as the cautious rule proposed in (Denœux, 2008). In our experiments, we assume independence and distinctness.
From a mass function, one can derive a belief function, being the degree of minimum or necessary specific (Smets, 1993) support for A:

$$Bel(A) = \sum_{B \subseteq A,\, B \neq \varnothing} m(B) \tag{11}$$

and a plausibility function, as the degree of maximum specific support for A:

$$Pl(A) = \sum_{A \cap B \neq \varnothing} m(B) \ . \tag{12}$$

After combination, the final decision is usually taken in favor of a simple hypothesis using one of several rules (Denœux, 1995; Denœux, 1997): *e.g.*, the maximum of plausibility (generally over simple hypotheses), the maximum of belief, the pignistic decision rule (Smets & Kennes, 1994), *etc*. For some applications, such as humanitarian demining, it may also be necessary to give more importance to some classes (*e.g.*, mines, since they must not be missed) at the decision level. Then maximum of plausibility can be used for the classes that should not be missed, and maximum of belief for the others (Milisavljević et al., 2003; Milisavljević & Bloch, 2005).
Not many papers can be found dealing with fusion of polarimetric data. In (Mascle et al., 1997), the main idea is to perform unsupervised classification on each image of a polarimetric data set separately. The intersections of the obtained classes define the set of discernment and the initial classes can be expressed as disjunctions of some of these intersections. This allows determining automatically both the singletons and the interesting disjunctions. This paper also shows the interest of using a fusion approach and of combining data from several polarimetric images. In (Le Hégarat-Mascle et al., 1998), the idea is to introduce neighborhood information as a mass function, so as to take a spatial regularity constraint into account and to consider it as a source of information.

5. Fusion

5.1 Fusion strategies

Fusion strategy no. 1 - including a global discounting factor (F1):
Some of the classification results in this work differentiate well two classes and some others do not. In addition, the overall reliability of each of the classification results is different. This fact should be taken into account in the fusion process so that the more reliable classification results influence the fusion result more than the less reliable ones. These are cases in which DS theory may be successful so it is our first choice of fusion approaches.
A first, simple method would consist in considering each of the classification results *cl1* - *cl6* as one information source. The focal elements would be simply the classes, using the outputs directly as mass functions. As no confidence values are provided but only decision

images, the mass would assume only values 0 or 1. This approach would inevitably result in a high conflict after the combination. Moreover, only the classes detected by all classifiers would be obtained as resulting focal elements, so no good result could be expected. This shows the interest of really using belief function theory or any other theory that takes into account the specificities of the classifiers, disjunctions of classes and ignorance (mass assigned to the full set, Θ).

In our first fusion strategy, we still consider each classifier output as one information source, but the focal elements are the singletons and Θ. The definition of $m(\Theta)$ takes into account both the fact that some classes are not detected (thus it should be equal to 1 at points where 0 is obtained for all detected classes) as well as global errors. We propose to use a discounting factor γ (Xu et al., 1992) equal to the overall accuracy of a classification result, *i.e.*, the sum of the diagonal elements of the confusion matrix, divided by the cardinality of the training areas. This discounting is applied to all masses defined as in the previous, straightforward approach. Then, if the output value of classification result cl_k (k = 1,2,...6) is C_i at a given pixel, the masses for that pixel and that classification output are assigned as follows:

$$m_k(C_i) = \gamma_k \,, \tag{13}$$

$$m_k(\Theta) = 1 - \gamma_k \,. \tag{14}$$

Note that this strategy explicitly uses the confidence matrix, which should be computed on the training areas for each classification output. Hence, at each step of the fusion, the focal elements are always singletons and Θ. After assigning masses by all classification outputs in the above way, the DS fusion is performed. Decision rule can be maximum of belief, of mass or of pignistic probability (all being equivalent in this case). This approach is very easy to implement and models in a simple way the fact that classification outputs may not give any information on some classes and may be imperfect. Results are explained by the conjunctive behavior of the Dempster's rule of combination.

Fusion strategy no. 2 - including class-dependent discounting factors (F2):

As a second alternative, we propose to use the confusion matrices for defining more specific discounting for each class. This approach is close to the one proposed in (Mercier et al., 2005; Mercier et al., 2008). Each output of the classifier is still one information source, and the focal elements are the singletons and Θ. From the confusion matrix computed from classifier output cl_k (k = 1,2,...6) and from the training areas, we use the diagonal coefficients $conf_k(i,i)$ for discounting. Thus, if the output value of cl_k is C_i at a given pixel, the masses for that pixel and that classifier output are assigned as follows:

$$m_k(C_i) = conf_k(i,i) \,, \tag{15}$$

$$m_k(\Theta) = 1 - conf_k(i,i) \,. \tag{16}$$

In comparison with the previous method, the richness of the information provided by a classifier output is better exploited as the class-dependent classification accuracy is used instead of a global accuracy of the classifier.

Fusion strategy no. 3 - fuzzy fusion (F3):

In order to compare the previous methods with a fuzzy approach, we have tested a simple method, where for each class we choose the best classification results, and combine them

with a maximum operator (possibly with some weights). Then a decision is made according to a maximum rule. The choice is made based on the confusion matrix for each classification result, by comparing the diagonal elements in all matrices for each class. This approach is interesting because it is very fast. It uses only a part of the information, which could also be a drawback if this part is not chosen appropriately. Some weights have to be tuned, which may need some user interaction in some cases. Although it may sound somewhat *ad hoc*, it is interesting to show what can be obtained using the best parts of all classifier outputs.
In a next step, in order to take more benefit from the information and to avoid the *ad hoc* tuning of the weights, we use for each class all the classifier outputs. If the value of classification result cl_k at a given pixel is C_i, it participates in the combination for that class discounted by the corresponding diagonal element of the confusion matrix, $conf_k(i,i)$. Such discounted classifier outputs are combined using a maximum operator. The decision is made again applying a maximum rule. Thus, in this approach, all classification results that have the same class as output participate in the combination performed for that class, discounted by their accuracy for that class.
Finally, we develop a third fuzzy fusion strategy, where for each class we use again all the classifier outputs. As in the previous strategy, if cl_k output at a given pixel is C_i, this classifier participates in the combination for that class discounted by the corresponding diagonal element of the confusion matrix, $conf_k(i,i)$. In addition, even if the output of classification result cl_k at a given pixel is C_j, $j \neq i$, it participates in the combination for class C_i if the confusion between the two classes, expressed by $conf_k(i,j)$ in the confusion matrix, is high enough (above a threshold). In that case, this element of the confusion matrix is used for discounting the classifier output prior to combination per class using a maximum operator. The decision is again made using a maximum rule. Since the results obtained in this way make the largest use of the information and are the best of the three strategies proposed in this subsection, this fuzzy strategy is used here. Note that this approach takes into account the fact that if the decision is C_j, the true class is possibly C_i, to some degree.
The actual list of classification results used for each class in our application is detailed in Subsection 7.3.

5.2 Spatial regularization

Spatial regularization is the final step in our fusion approach, applied to the output of each of the proposed strategies. Namely, it is very unlikely that isolated pixels of one class can appear in another class. Hence, several local filters have been tested, such as a majority filter, a median filter, or morphological filters, applied on the decision image. A Markovian regularization approach on local neighborhoods was tested too. All these filters give similar and good results, and improve results of the previous fusion steps (see Section 6). A recently proposed approach (Bloch, 2008) could be used as well, by integrating spatial information directly in the definition of mass functions.

6. Data and experimental approach

For the tests presented here, we have used single-look complex data of the area of Oberpfaffenhofen, Germany, acquired by the E-SAR airborne sensor of the German Aerospace Centre (DLR) during a measurement campaign in October 1999 (see Table 2). The data consist of L-band scattering matrices measured in the *hv*-basis. The size of the data set is 1000 x 4050 pixels (in range and azimuth, respectively).

Noise removal and speckle reduction are obtained by means of simple boxcar filtering with fixed window dimensions. The averaging window adopted for all sets of parameters is 5 x 11 pixels (in range and azimuth, respectively). In this way, data processing could be performed following a consistent approach, *i.e.*, choosing the same averaging window size to define the covariance and coherency matrices or to filter data derived from coherent methods.

The imaged area is situated approximately 25 km South-West of the city of Munich and includes several interesting features: the DLR centre, the former Fairchild Dornier airplane factory and the airfield shared by the two firms (see Figure 1). Not far from them, a small lake and the village of Gilching are located. Other important man-made structures are the motorway and the railway line stretching across the image. The vegetation patches consist of coniferous and mixed forests, meadows and crops.

A common classification procedure has been adopted for all the polarimetric parameters (Alberga et al., 2006). At first, a set of seven ground cover classes has been defined: "water", "houses", "roads", "trees", "grass", "field 1" and "field 2". For each of them, separated areas of training and test samples have been identified having a comparable number of pixels (at this scope, aerial photographs and a cartographic map have been used as complementary sources of information). Then, the training pixels from each class have been fed into the classifiers to perform the training stage. As a following step, all the data have been classified using the MLP. Finally, the test samples have been used to measure the classification performance. The fusion of the results has been performed at this stage, according to the three strategies described in Section 5 and providing a re-assignment of the pixels.

Fig. 1. Backscattered intensity image and regions of interest: [1] "water", [2] "houses, [3] "roads", [4] "trees", [5] "grass", [6] "field 1", [7] "field 2".

The class "roads" has been defined using only the runway of the airport, so it refers to a relatively wide asphalt surface. Regarding the class indicated as "houses", it is related to areas with family houses surrounded by gardens (often including trees). In other words, it represents an impure class, characterized by the presence of different scattering components: flat surfaces, dihedrals, volumes (surrounding vegetation), and rough surfaces (ground).

Flight date	6/10/1999
RF-band	L
Centre frequency	1.3 GHz
Wavelength	23 cm
Bandwidth	100 MHz
Range resolution	1.5 m
Azimuth resolution	0.89 m

Table 2. Main data and E-SAR system parameters

7. Results

In the following, different measures are used to assess the classification accuracy, based on the confusion matrices obtained on the test areas (note that the columns and the rows in the confusion matrices correspond to the ground-truth map and to the classification output map, respectively):

- κ coefficient, as a measure of the quality of the classified map compared to a randomly generated map:

$$\kappa = \frac{N \sum_{i=1}^{r} conf(i,i) - \sum_{i=1}^{r} conf(i,+)conf(+,i)}{N^2 - \sum_{i=1}^{r} conf(i,+)conf(+,i)}, \tag{17}$$

  where N is the total number of pixels in the confusion matrix, r is the dimension of the confusion matrix, *conf(i,i)* is the number of pixels on the main diagonal, *i.e.*, in row and column i of the confusion matrix, *conf(i,+)* is the total number of pixels in row i, and *conf(+,i)* is the total number of pixels in column i of the confusion matrix;
- overall accuracy γ, *i.e.*, the percentage of correctly classified pixels:

$$\gamma = \frac{\sum_{i=1}^{r} conf(i,i)}{N}, \tag{18}$$

- user's accuracy (*UA*), being the probability that a given pixel will appear on the ground as it is classified, so for class j, it can be defined as:

$$UA_j = \frac{conf(j,j)}{conf(j,+)}, \tag{19}$$

- producer's accuracy (*PA*), *i.e.*, the percentage of a given class that is identified correctly on the map, being for class j calculated as:

$$PA_j = \frac{conf(j,j)}{conf(+,j)}. \tag{20}$$

Figures 2-4 contain images of classification results *cl1* - *cl6*, which are inputs for the fusion module. The accuracy estimates obtained for each of the classification results are given in

Table 5. The use of the covariance matrix leads to the best overall accuracy. This confirms the capability of incoherent observables to describe complex scenarios better than coherent ones. Indeed, the Freeman decomposition has the second best overall performance. The limits of the $H/\overline{\alpha}/A$ parameters are bound to their definition: neither of the three parameters expresses an intensity or a power measurement but they directly provide a semantic interpretation of the type and statistical behavior of the imaged targets. At the class level, again, the Freeman decomposition provides some of the best *PA* and *UA* values but then also some coherent representations yield specific class maxima, indicating the capability of the given model to describe the typical target of that class.

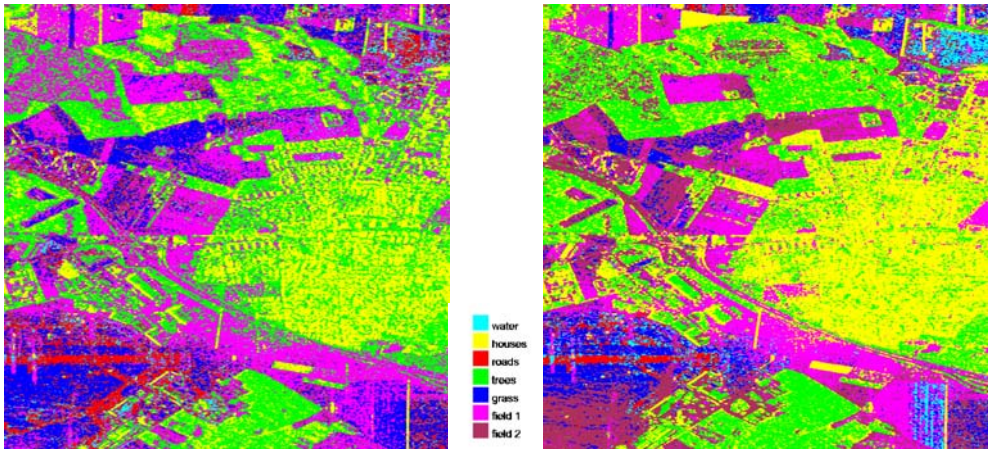

Fig. 2. MLP classification results: left - **[S]** matrix (*cl1*), right - covariance matrix (*cl2*).

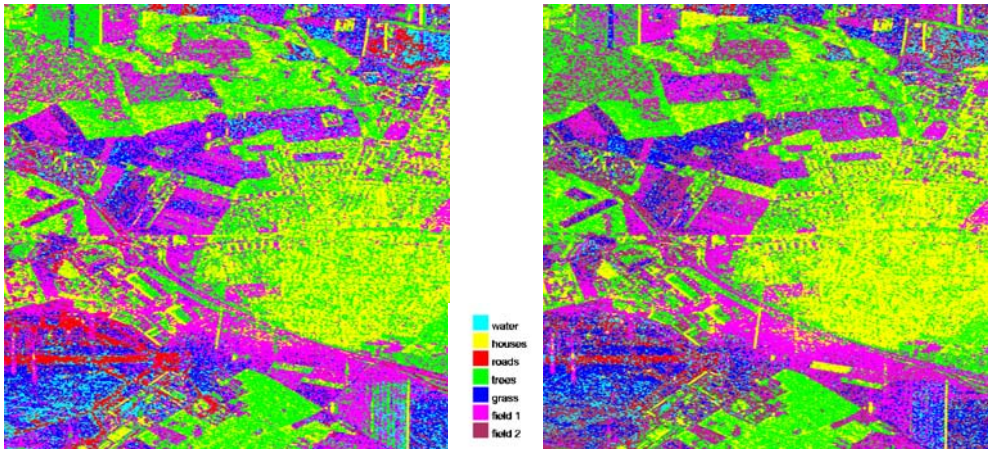

Fig. 3. MLP classification results: left - Pauli decomposition (*cl3*), right - Cameron decomposition (*cl4*)

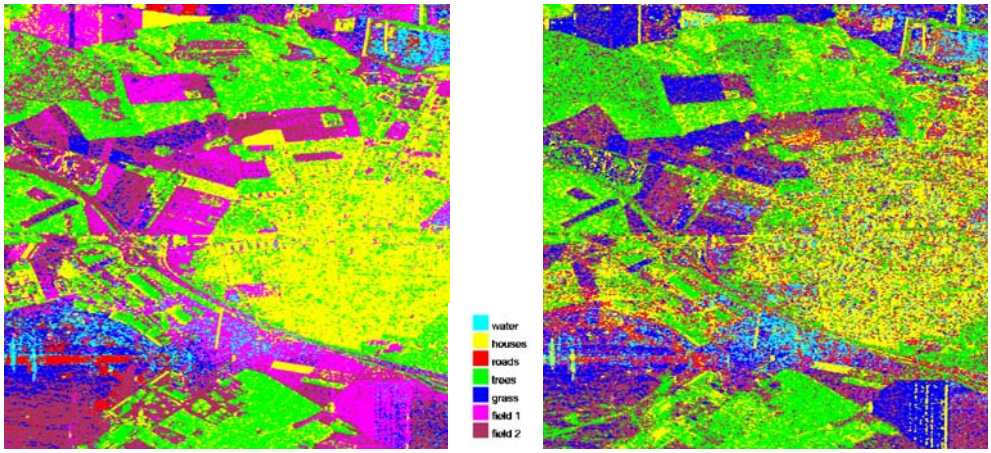

Fig. 4. MLP classification results: left - Freeman decomposition (*cl5*), right - $H/\overline{\alpha}/A$ (*cl6*).

Besides the fusion results, the fusion module provides confidence and stability images too. At each pixel, the confidence image contains the confidence degree of the decided class, while the stability image represents the difference between the confidence in the decided class and the confidence in the second most possible class. Examples of these two images are shown in Figure 8.

7.1 Results with F1

Here, we provide the results obtained using the method F1 (Subsection 5.1). The values of the discounting factors γ, corresponding to the six classification results *cl1* - *cl6* are given in Table 3. For each classification result, this factor is calculated as the normalized sum of the diagonal elements of the confusion matrix obtained on the training areas (*i.e.*, the overall accuracy).

Classification result	*cl1*	*cl2*	*cl3*	*cl4*	*cl5*	*cl6*
γ	0.66	0.84	0.66	0.71	0.77	0.55

Table 3. Discounting factors for F1

Figure 5 shows the result of fusion using the F1 approach and spatial regularization. Table 5 indicates that F1 outperforms the best classification results in most of the cases. The only exceptions are *UA* for "roads" and *PA* for "grass". The overall accuracy improvement with respect to the single representations ranges from 13% to 37% but the main advantage of the fusion approach lies in the fact that, at class level, all features are recognized with good *PA* and *UA*. For example, the low *PA* of "water" using the **[S]** matrix elements or of "field 1" using the $H/\overline{\alpha}/A$ parameters have been largely compensated by means of the information provided by the other representations.

7.2 Results with F2

The discounting factors for the method F2, described in Subsection 5.1, are given in Table 4. They correspond to the diagonal elements of the confusion matrices obtained on the training areas, for each of the six classification results.

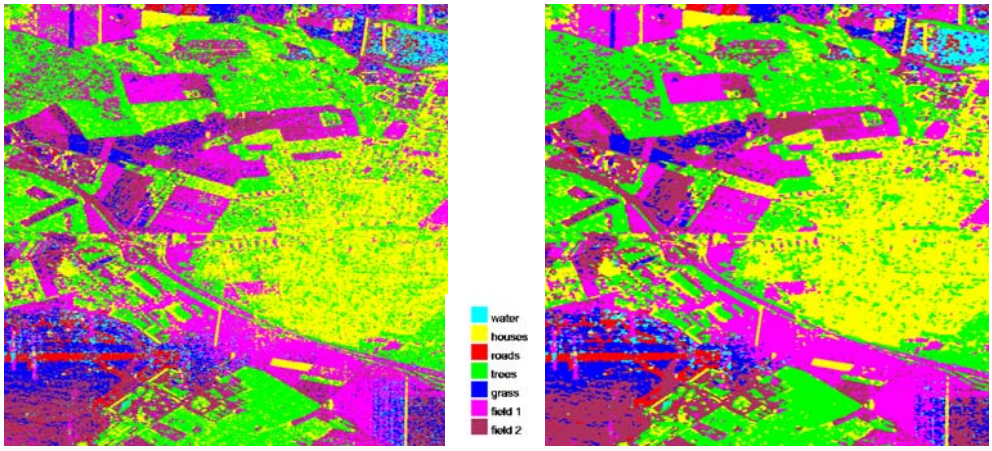

Fig. 5. Results of F1 before (left) and after (right) spatial regularization

Classification result → Class ↓	*cl1*	*cl2*	*cl3*	*cl4*	*cl5*	*cl6*
water	0.23	0.70	0.50	0.48	0.59	0.55
houses	0.70	0.89	0.68	0.83	0.86	0.57
roads	0.84	0.80	0.84	0.74	0.74	0.42
trees	0.89	0.98	0.89	0.88	0.92	0.87
grass	0.70	0.80	0.59	0.71	0.66	0.63
field 1	0.84	0.89	0.78	0.82	0.80	0.17
field 2	0.41	0.81	0.31	0.54	0.82	0.67

Table 4. Discounting factors for F2

According to Table 5, F2 provides better results than the best individual classification in most of the cases. However, there are only half of the classes for which this fusion method outperforms the previous one. Hence, regardless the specificity of the class discounting (varying depending on the class), no clear advantage is obtained by this type of fusion. An overall discounting factor for a given classification result (as for F1) seems to be sufficient. Figure 6 shows the result of the F2 fusion.

7.3 Results with F3

As explained in Subsection 5.1, with this approach we take into account the fact that for some classifier outputs, there exist pairs of classes whose confusion is quite strong. Thus, even if cl_k does not give class C_i as its output but class C_j, we include that classification result in the combination for class C_i if the corresponding coefficient of the confusion matrix obtained on the training data, $conf_k(i,j)$ has a value higher than some threshold.

The following classifier outputs have been used, all discounted by the corresponding coefficients in the confusion matrices:

- for "water": all six classification results in case of "water", plus "grass" and "roads" for *cl3*, "field 1" and "roads" for *cl6* and "roads" for *cl2*, *cl4* and *cl5*;
- for "houses": all six classification results in case of "houses", plus "trees" for *cl3* and *cl4*;

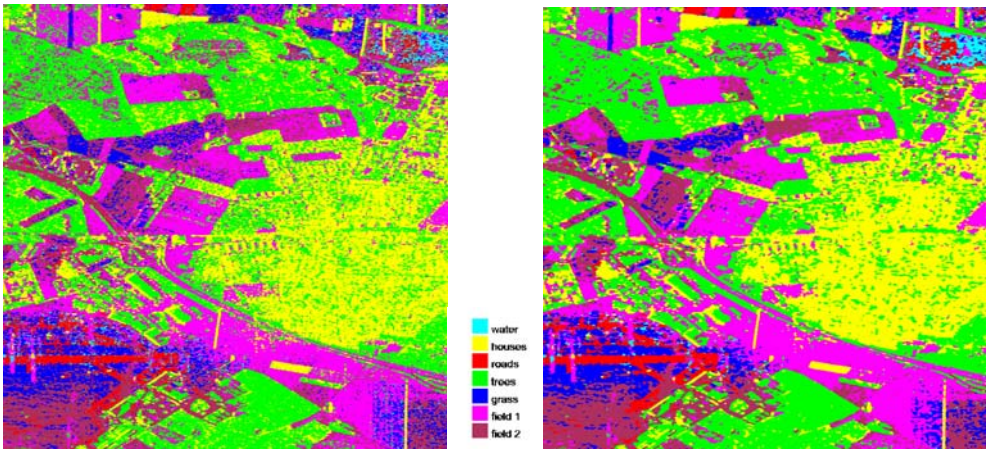

Fig. 6. Results of F2 before (left) and after (right) spatial regularization

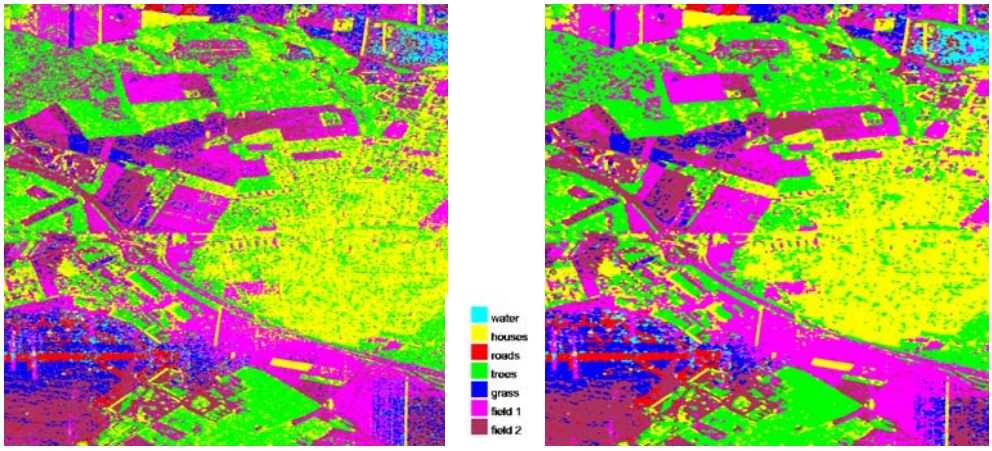

Fig. 7. Results of F3 before (left) and after (right) spatial regularization

- for "roads": all six classification results in case of "roads", plus "field 1" and "water" for *cl1* and *cl6*, and "water" for *cl3*, *cl4* and *cl5*;
- for "trees": all six classification results in case of "trees" and "houses";
- for "grass": all six classification results in case of "grass", plus "water" for *cl1* - *cl5*, "field 1" for *cl6*, and "field 2" for *cl1*, *cl2* and *cl4*;
- for "field 1": all six classification results in case of "field 1", plus "roads" and "grass" for *cl6*, "grass" for *cl3* and "field 2" for *cl1*;
- for "field 2": all six classification outputs in case of "field 2" and "grass", plus "roads" and "field1" for *cl6*.

Table 5 shows that F3 provides results similar to F1. Note that the typical sources of confusion (i.e., the classes characterized by reciprocal *PA* and *UA*) are also those whose intensity is roughly comparable (*e.g.*, "water" and "grass" or "houses" and "trees"). Thus, it seems that the radiometric information finally plays the most relevant role in the definition,

and hence classification, of a feature. Fig. 7 presents the result of F3 fusion. Confidence and stability images for this type of fusion are shown in Fig. 8.

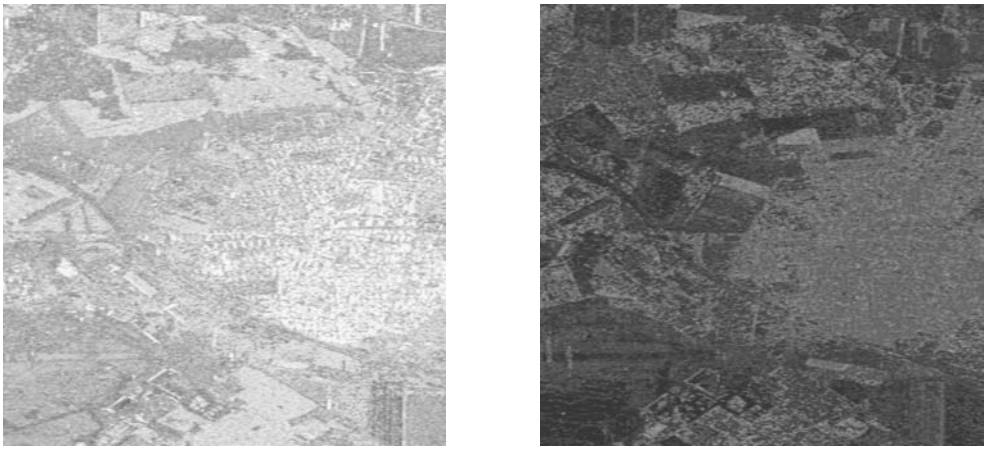

Fig. 8. Confidence (left) and stability (right) image for F3

		cl1	*cl2*	*cl3*	*cl4*	*cl5*	*cl6*	F1	F2	F3
γ		0.68	0.73	0.64	0.65	0.71	0.49	0.86	0.84	0.86
κ		0.62	0.68	0.58	0.59	0.66	0.41	0.84	0.81	0.84
PA	water	0.25	0.60	0.43	0.40	0.61	0.51	0.75	0.53	0.74
	houses	0.72	0.90	0.71	0.82	0.92	0.66	0.94	0.93	0.95
	roads	0.72	0.70	0.72	0.59	0.71	0.32	0.85	0.90	0.85
	trees	0.88	0.92	0.83	0.85	0.92	0.87	0.99	0.99	0.99
	grass	0.80	0.60	0.57	0.68	0.52	0.54	0.68	0.69	0.66
	field 1	0.84	0.62	0.79	0.69	0.54	0.05	0.89	0.92	0.91
	field 2	0.42	0.70	0.25	0.43	0.70	0.60	0.83	0.80	0.83
UA	water	0.73	0.84	0.62	0.59	0.79	0.62	0.96	0.99	0.95
	houses	0.86	0.81	0.80	0.83	0.81	0.70	0.96	0.97	0.96
	roads	0.52	0.92	0.67	0.56	0.85	0.35	0.89	0.75	0.89
	trees	0.70	0.85	0.69	0.75	0.89	0.69	0.93	0.90	0.94
	grass	0.54	0.49	0.39	0.49	0.54	0.50	0.69	0.66	0.68
	field 1	0.84	0.78	0.76	0.84	0.77	0.17	0.90	0.86	0.89
	field 2	0.45	0.44	0.37	0.34	0.37	0.25	0.62	0.66	0.63

Table 5. Results on the test data

8. Conclusion

In this chapter, three strategies for fusion of land cover classification results of polarimetric SAR data are proposed, two of them based on belief function theory and one based on fuzzy sets theory. As a final step of each of the fusion strategies, spatial regularization is performed. The proposed fusion strategies are applied to the outputs of a neural network classifier, corresponding to six different sets of polarimetric parameters as input. The values

of the different polarimetric parameters are extracted from single-look complex data of the area of Oberpfaffenhofen, Germany, acquired by the E-SAR airborne sensor of the German Aerospace Centre.
The proposed fusion strategies do not need statistical independence of their input information. Each of the strategies uses the confusion matrices obtained on the training data set as means for estimating the accuracy of each of the classification results and for weighing them prior to their fusion.
The test data set, different from the training data set, is used to compare the classification accuracy of the fusion results with the single representation classification results. The results obtained by each of the three fusion strategies show a significant improvement of the classification accuracy of the separate classification results.
The general improvement of the classification accuracy indicates the complementary nature of the information provided by the analyzed polarimetric representations. Although their classification performance is comparable, different characteristics of the scene are enhanced by each set of observables, so their fusion effectively takes advantage of their whole information content.

9. Acknowledgements

E-SAR data have been provided by the German Aerospace Centre (DLR). DLR takes no responsibility for the published results.
Part of the SAR data processing has been performed using the free RAT software developed at the Department of Computer Vision and Remote Sensing of the Technical University of Berlin, Germany.

10. References

Touzi, R.; Boerner, W-M.; Lee, J-S. & Lueneburg, E. (2004). A Review of Polarimetry in the Context of Synthetic Aperture Radar: Concepts and Information Extraction. *Canadian Journal of Remote Sensing,* Vol. 30, No. 3, pp. 380-407.

Cloude, S. R. & Pottier, E. (1997). An Entropy Based Classification Scheme for Land Application of Polarimetric SAR. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 35, No. 1, pp. 68-78.

Ferro-Famil, L.; Pottier, E. & Lee, J-S. (2001). Unsupervised Classification of Multifrequency and Fully Polarimetric SAR Images Based on H/A/alpha-Wishart Classifier. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 39, No. 11, pp. 2332-2342.

Hoekman, D. H. & Vissers, M. A. M. (2003). A New Polarimetric Classification Approach Evaluated for Agricultural Crops. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 41, No. 12, pp. 2881-2889.

Kersten, P. R.; Lee, J-S. & Ainsworth T. L. (2005). Unsupervised Classification of Polarimetric Synthetic Aperture Radar Images Using Fuzzy Clustering and EM Clustering. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 43, No. 3, pp. 519-527.

Lee, J-S.; Grunes, M. R. & Kwok, R. (1994). Classification of Multi-Look Polarimetric SAR imagery Based on Complex Wishart Distribution. *International Journal of Remote Sensing*, Vol. 15, No. 11, pp. 2299-2311.

Lee, J-S.; Grunes, M. R.; Ainsworth, T. L.; Du, L. J.; Schuler, D. L. & Cloude, S. R. (1999). Unsupervised Classification Using Polarimetric Decomposition and the Complex

Wishart Classifier. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 37, No. 5, pp. 2249-2258.

Lee, J-S.; Grunes, M. R.; Pottier, E. & Ferro-Famil, L. (2004). Unsupervised Terrain Classification Preserving Polarimetric Scattering Characteristics. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 42, No. 4, pp. 722-731.

Chen, C-T.; Chen, K-S. & Lee, J-S. (2003). The Use of Fully Polarimetric Information for the Fuzzy Neural Classification of SAR Images. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 41, No. 9, pp. 2089-2099.

Alberga, V.; Satalino, G. & Staykova, D. K. (2006). Polarimetric SAR Observables for Land Cover Classification: Analyses and Comparisons, *Proceedings of SPIE - SAR Image Analysis, Modeling, and Techniques VIII*, vol. 6363, Stockholm, Sweden, DOI: 636305.

Cameron, W. L.; Youssef, N. N. & Leung, L. K. (1996). Simulated Polarimetric Signatures of Primitive Geometrical Shapes. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 34, No. 3, pp. 793-803.

Cloude, S. R. & Pottier, E. (1996). A Review of Target Decomposition Theorems in Radar Polarimetry. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 34, No. 2, pp. 498-518.

Freeman, A. & Durden, S. L. (1998). A Three-Component Scattering Model for Polarimetric SAR Data. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 36, No. 3, pp. 963-973.

Benediktsson, J. A.; Swain, P. H. & Ersoy, O. K. (1990). Neural Network Approaches versus Statistical Methods in Classification of Multisource Remote Sensing Data. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 28, No. 4, pp. 540-551.

Mascle, S.; Bloch, I. & Vidal-Madjar, D. (1997). Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 35, No. 4, pp. 1018-1031.

Le Hégarat-Mascle, S.; Bloch, I. & Vidal-Madjar, D. (1998). Introduction of Neighborhood Information in Evidence Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover. *Pattern Recognition*, Vol. 31, No. 11, pp. 1811-1823.

Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton, NJ: Princeton Univ. Press.

Smets, P. (1990b). The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, pp. 447-458.

SMART consortium (2004). SMART - final report.

Dubois, D.; Prade, H. & Yager, R. (1999). Merging Fuzzy Information, In: *Handbook of Fuzzy Sets Series, Approximate Reasoning and Information Systems*, Bezdek, J.C.; Dubois, D. & Prade, H. (Eds.), ch. 6, Kluwer, Boston, USA.

Ulaby, F. T.; Moore, R. K. & Fung, A. K. (1982). *Microwave remote sensing - Volume II: Radar remote sensing and surface scattering and emission theory*, Addison-Wesley Publishing Company, Reading, USA.

Born, M. & Wolf, E. (1985). *Principles of optics*, 6th edition, Pergamon Press, New York, USA.

Curlander, J. C. & McDonough, R. N. (1991). *Synthetic Aperture Radar - Systems and signal processing*, John Wiley & Sons, New York, USA.

Corr, D. G. & Rodrigues, A. (2002). Alternative Basis Matrices for Polarimetric Decomposition, *Proceedings of EUSAR*, pp. 597-600, Cologne, Germany.

Moriyama, T.; Uratsuka, S.; Umehara, T.; Maeno, H.; Satake, M.; Nadai, A. & Nakamura, K. (2004). Feature Extraction of Urban Area Based on Polarimetric Decomposition, *Proceedings of EUSAR*, pp. 435-438, Ulm, Germany.

Cloude, S. R. (1986). Group Theory and Polarization Algebra. *Optik*, Vol. 75, No. 1, pp. 26-36.

Hertz, A.; Krogh, A. & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Addison Wesley, Redwood City, USA.

Funahashi, K. (1989). On the Approximate Realizations of Continuous Mappings by Neural Networks. *Neural networks*, Vol. 2, No. 3, pp. 183-192.

van Cleynenbreugel, I.; Osinga, S.A.; Fierens, F.; Suetens, P. & Oosterlinck, A. (1991). Road Extraction from Multi-temporal Satellite Images by an Evidential Reasoning Approach. *Pattern Recognition Letters*, Vol. 12, pp. 371-380.

Tupin, F.; Bloch, I. & Maître, H. (1999). A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 3, pp. 1327-1343.

Milisavljević, N. & Bloch, I. (2003). Sensor fusion in anti-personnel mine detection using a two-level belief function model. *IEEE Trans. Systems, Man and Cybernetics*, Part C, Vol. 33, No. 2, pp. 269-283.

Appriou, A. (1993). Formulation et traitement de l'incertain en analyse multi-senseurs, *Proceedings of GRETSI*, pp. 951-954, Juan les Pins, France.

Lee, R. H. & Leahy, R. (1990). Multi-spectral Classification of MR Images Using Sensor Fusion Approaches, *Proc. of SPIE Medical Imaging IV: Image Processing*, Vol. 1233, pp. 149-157.

Denœux, T. (1995). A k-nearest Neighbor Classification Rule based on Dempster-Shafer Theory. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 25, No. 5, pp. 804-813.

Bloch, I. (1996). Some Aspects of Dempster-Shafer Evidence Theory for Classification of Multi-Modality Medical Images Taking Partial Volume Effect into Account, *Pattern Recognition Letters*, Vol. 17, No. 8, pp. 905-919.

Milisavljević, N. & Bloch, I. (2001). A Two-Level Approach for Modeling and Fusion of Humanitarian Mine Detection Sensors within the Belief Function Framework, *Proc. of Applied Stochastic Models and Data Analysis*, Vol. 2, pp. 743-748, Compiègne, France.

Ménard, M.; Zahzah, E.H. & Shahin, A. (1996). Mass Function Assessment: Case of Multiple Hypotheses for the Evidential Approach, *Proceedings of Europto Conf. on Image and Signal Processing for Remote Sensing*, Vol. 2955, pp. 214-218, Taormina, Italy.

Smets, P. (1993). Belief Functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *International Journal of Approximate Reasoning*, Vol. 9, pp. 1-35.

Smets, P. (1995). The Transferable Belief Model for Uncertainty Representation, *Technical Report TR/IRIDIA/95-23*, IRIDIA, Université Libre de Bruxelles, Brussels, Belgium.

Milisavljević, N.; Bloch, I.; van den Broek, S.P. & Acheroy, M. (2003). Improving mine recognition through processing and Dempster-Shafer fusion of ground-penetrating radar data. *Pattern Recognition*, Vol. 36, No. 5, pp. 1233-1250.

Milisavljević, N. & Bloch, I. (2005). Improving Mine Recognition through Processing and Dempster-Shafer Fusion of Multisensor Data. In: *Computer-Aided Intelligent Recognition, Techniques and Applications*, Sarfraz, M. (Ed.), ch. 17, pp. 319-343, J. Wiley, New York, ISBN: 0-470-09414-1.

Xu, L.; Krzyzak, A. & Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, No. 3, pp. 418--435.

Mercier, D.; Cron, G.; Denœux, T. & Masson, M.-H. (2005). Fusion of multi-level decision systems using the transferable belief model, *Proceedings of FUSION'2005*, Philadelphia, PA, USA.

Mercier, D.; Quost, B. & Denœux, T. (2008). Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, Vol. 9, No. 2, pp. 246--258.

Denœux, T. (1997). Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, Vol. 30, No. 7, pp. 1095--1107.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, Vol. 38, pp. 325--339.

Denœux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. *Artificial Intelligence*, Vol. 172, No. 2-3, pp. 234-264.

Appriou, A. (1998). Uncertain data aggregation in classification and tracking processes. *Aggregation and Fusion of imperfect information*, Bouchon-Meunier, B. (Ed.), pp. 231-260, Physica-Verlag, Heidelberg.

Smets, P. & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, Vol. 66, pp. 191-234.

Bloch, I. (2008). Defining Belief Functions using Mathematical Morphology - Application to Image Fusion under Imprecision. *International Journal of Approximate Reasoning*, Vol. 48, pp. 437-465.