Kohonen-Based Credal Fusion of Optical and Radar Images for Land Cover Classification

Imen Hammami, Jean Dezert, Grégoire Mercier

Kohonen-based Credal Fusion of Optical and Radar Images for Land Cover Classification

Imen Hammami
University of Tunis el Manar
Tunis Sciences Faculty
Tunis, Tunisia.
imen.hammammi@gmail.com

Jean Dezert
The French Aerospace Lab
ONERA/DTIM/MSDA
Palaiseau, France.
jean.dezert@onera.fr

Grégoire Mercier
CTO, eXo maKina
Digital Technologies
Paris, France.
gregoire.mercier@exomakina.fr

***Abstract*—This paper presents a Credal algorithm to perform land cover classification from a pair of optical and radar remote sensing images. SAR (Synthetic Aperture Radar) /optical multispectral information fusion is investigated in this study for making the joint classification. The approach consists of two main steps: 1) relevant features extraction applied to each sensor in order to model the sources of information and 2) a Kohonen map-based estimation of Basic Belief Assignments (BBA) dedicated to heterogeneous data. This framework deals with co-registered images and is able to handle complete optical data as well as optical data affected by missing value due to the presence of clouds and shadows during observation. A pair of SPOT-5 and RADARSAT-2 real images is used in the evaluation, and the proposed experiment in a farming area shows very promising results in terms of classification accuracy and missing optical data reconstruction when some data are hidden by clouds.**

Keywords: Land cover classification, remote sensing, image fusion, Kohonen map, belief functions.

I. Introduction

In this paper we consider the problem of the joint classification of farming landscape images from a multispectral optical image denoted I_{MS} and a Synthetic Aperture Radar (SAR) image denoted I_{SAR}. We assume the image registration problems resolved in a pre-processing step and we focus only on the classification problem. Classically, the classification is obtained from an information fusion approach which can be addressed at distinct levels [1]:

- **A pixel level** (low level fusion) [2]: Ideally, the images should be fused at this pixel information level, in which raw data extracted from each pixel, such as spectral or temporal information of the considered sources, are used. However, the design of efficient techniques for achieving this fusion is very difficult due to the complexity of joint processing of heterogeneous data, such as optical and radar data. This pixel-based heterogeneous fusion requires the use of accurate co-registered images that are often derived from a resampling process in a pre-processing step. The fusion techniques at this level require high computational resources and time.
- **At feature level** (mid level fusion) [3]: The feature (objet) level fusion is a mid (intermediary) level fusion approach. It is mainly based on the extraction of one, or more, feature maps by computing the relevant descriptors of objects from each input image under analysis. Then, a fusion strategy must be applied to combine attributes from different images related to a same spatial area. This approach induces a loss of information in the processing due to the replacement of original raw data by the attributes extracted which usually are not 100% reliable.
- **At decision level** (high level fusion) [4]: The decision level operates directly on individual decisions found by applying a proper processing for each image analyzed separately. Although this level is considered the most robust among the three, the fusion solution solution is not globally optimal since it seeks to optimize each source individually at first. There is no joint image processing for making the classification.

Due to the complexity of development of efficient low level fusion methods and their high computational resource requirement, most of users working in remote sensing prefer to work at mid level fusion or high level fusion. In this paper, we propose a new method for joint classification from optical and radar images developed at the mid fusion level exploiting features information being a good compromise between low level and high level fusion techniques. Our main contribution lies in the joint use of images to extract the features in each image before making their fusion for the classification. For this, Kohonen's unsupervised map classification framework [5], [6] is used to provide an effective way for dealing with the heterogeneity of the data sources. Our technique is based on the belief functions because they offer a mathematical framework for modeling epistemic or subjective uncertainty (i.e., uncertainty resulting from imperfect knowledge), as well as stochastic uncertainty (i.e., uncertainty resulting from data diversity). In the context of this work, the multispectral optical image I_{MS} is considered as reliable if no clouds or shadows affect the data. The SAR image I_{SAR} is known to be noisy [7] due to unavoidable speckle noise effects. Moreover, with no polarimetric capability (which is the case in this study), the information that can be extracted from the SAR image is much less reliable than the one from the optical data. Different features are extracted from I_{SAR} and I_{MS}. For I_{SAR}, we use six local texture descriptors [8], [9] to take into account the information computed from the neighborhoods surrounding

the SAR pixel under concern. At the feature level, each pixel (i,j) of the SAR image is characterized by a 6-tuple[1] denoted $\kappa(i,j)$. The *featured-SAR image* is denoted $I_{\kappa_{\text{SAR}}}$. The surface reflectance of the different bands of the multispectral optical image I_{MS} are used directly as features. To guarantee that the classes are defined in a homogeneous fashion both from the optical and from the SAR observations, a first coarse joint classification is performed to link the spectral signatures of I_{MS} and the SAR texture descriptors of $I_{\kappa_{\text{SAR}}}$. For this, a simple K-means classifier is used with an appropriate distance that accounts for the heterogeneity of the joint observation and cross-calibration factor, which in turn accounts for the relative dynamics between the two observations. Once this coarse joint classification is done, at each pixel (i,j) of both registered images the fusion is achieved thanks to belief functions. This of course requires the construction of BBAs $m_{\text{MS}}(\cdot)$ and $m_{\kappa_{\text{MS}}}(\cdot)$ defined with respect to a common classification frame of discernment, and a decision-making strategy for making the classification. This paper is organized as follows. Sections II and III recall respectively the basics of belief functions, and Kohonen's map technique. Section IV proposes a technique of estimation of belief functions from heterogeneous data in the case of complete and missing optical data. Section IV. Some results obtained from real remote sensing images are presented in Section V, followed by a conclusion with perspectives in the last section.

II. Basics of belief functions

Shafer introduced Belief Functions (BF) [10] to model the epistemic uncertainty, to reason under uncertainty and to combine distinct sources of evidence. BF are based on preliminary Dempster's works [11], [12]. We consider a finite discrete frame of discernement (FoD) $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, with $n > 1$, and where all exhaustive and exclusive elements of Θ represent the set of the potential solutions of the problem under concern. The set of all subsets of Θ is the power-set of Θ denoted by 2^Θ. The number of elements (i.e. the cardinality) of 2^Θ is $2^{|\Theta|}$. A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping $m(\cdot) : 2^\Theta \rightarrow [0,1]$ satisfying the conditions $m(\emptyset) = 0$ and $\sum_{A \in 2^\Theta} m(A) = 1$. The quantity $m(A)$ is the mass of belief of subset A committed by the source of evidence (SoE). A focal element X of a BBA $m(\cdot)$ is an element of 2^Θ such that $m(X) > 0$. The set of all focal elements of $m(\cdot)$ is denoted $\mathcal{F}_\Theta(m) \triangleq \{X \subseteq \Theta | m(X) > 0\} = \{X \in 2^\Theta | m(X) > 0\}$.

The belief and plausibility functions are defined by

$$Bel(A) = \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X) \quad \text{and} \quad Pl(A) = \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X). \quad (1)$$

The width $Pl(A) - Bel(A)$ of the belief interval $[Bel(A), Pl(A)]$ is called the *uncertainty on* A committed by the SoE. It represents the imprecision on the (subjective) probability of A granted by the SoE which provides the BBA $m(\cdot)$. When all elements of $\mathcal{F}_\Theta(m)$ are only singletons, $m(\cdot)$ is called a *Bayesian BBA* [10] and its corresponding $Bel(\cdot)$ and $Pl(\cdot)$ functions are homogeneous to a same (subjective) probability measure $P(\cdot)$. The functions $m(\cdot)$, $Bel(\cdot)$ and $Pl(\cdot)$ are one-to-one [10]. If a source of evidence is known reliable at some degree with a reliability factor[2] $\alpha \in [0,1]$, then its BBA $m(\cdot)$ must be discounted (modified) to work with a new BBA defined by $m'(X) = \alpha \cdot m(X)$ for $X \neq \Theta$, and by $m'(\Theta) = \alpha \cdot m(\Theta) + (1 - \alpha)$. This is the classical Shafer's discounting method [10]. A more refined reliability discounting technique called contextual discounting can be found in [13]. In the combination of BBAs, one can also take into account (if one wants) the importance factor of each source. The importance level of the sources enters in many multi-criteria decision-making problems involving BF. For this, one commits to each source a priority factor[3] $\beta \in [0,1]$ and we apply the importance discounting method proposed in [14]. More precisely, we have to work with a new (non normalized) BBA defined by $m''(X) = \beta \cdot m(X)$ for $X \neq \emptyset$, and by $m''(\emptyset) = \beta \cdot m(\emptyset) + (1 - \beta) = 1 - \beta$. Of course, after the combination of BBAs a normalization step is necessary before making the final decision.

In DST framework, Shafer [10] did propose to combine $s \geq 2$ distinct sources of evidence represented by BBAs $m_1(.), \ldots, m_s(.)$ over the same FoD with Dempster's rule (i.e. the normalized conjunctive rule) which is commutative and associative. The justification and behavior of Dempster's rule have been strongly disputed from both theoretical and practical standpoints as reported in [15]–[17]. In particular, the high (or even very low) conflict level between the sources can be totally ignored by Dempster's rule which is a very serious problem [18]. Also, Shafer's conditioning (based on Dempster's rule) is inconsistent with the probabilistic conditioning and Generalized Bayes' Theorem [19]. We don't assume, nor claim, that there must be a single optimal belief fusion method that fits all situations, but it is reasonable and it makes sense to test and evaluate other combination techniques in applications. In this work, we focus on the Proportional Conflict Redistribution rule #6 (PCR6) developed in DSmT (Dezert-Smarandache Theory [20]) because it preserves the specificity of information through its conflict redistribution principle and it usually provides better performances than Dempster's rule as reported by many researchers working in different fields of applications. Of course, the price to pay for using PCR6 is its higher computational complexity with respect to Dempster's rule. The decision-making from a BBA can be done in many ways (max of belief, max of plausibility, max of some approximate probability measure built from the BBA, etc) [21].

[1] More precisely, $\kappa(i,j) \triangleq (\mu, \sigma, \beta_1, \beta_2, f_6, f_5)$, i.e the mean, standard deviation, skewness, kurtosis, sum-average and inverse different moment - see [9] for details and their definitions in appendix.

[2] $\alpha = 1$ means that the source is 100% reliable, whereas $\alpha = 0$ means that the source is totally unreliable.

[3] $\beta = 1$ means that the source is considered as 100% important for the analyst, whereas $\beta = 0$ means that the source is considered as not important at all.

III. KOHONEN'S MAP TO MODEL BF

Kohonen's map, performs an unsupervised neural network-based competitive learning [5]. It defines a non-linear mapping from an input space (say $\mathbb{R}^p$) onto a regular array of $N \times M$ nodes as illustrated in Fig. 1.

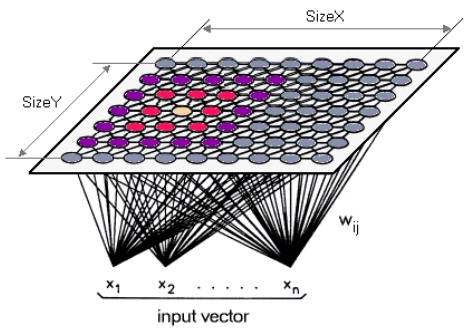


Figure 1. Schematic view of Kohonen's Self-Organizing Map [22].

A reference vector, also called a weighting vector, $\boldsymbol{w}(i,j) \in \mathbb{R}^p$ is associated with the node at each position (i,j) with $1 \leq i \leq N$ and $1 \leq j \leq M$. During the projection of $\boldsymbol{x} \in \mathbb{R}^p$ to the map, the input vector is compared to each $\boldsymbol{w}(i,j)$ by using an appropriate metric. The best match $\boldsymbol{w}_{\boldsymbol{x}}$ (the node that minimizes this metric) is defined as the output of the Self-Organizing Map (SOM): thus, the input data $\boldsymbol{x}$ is mapped onto the SOM at location $(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$, where $\boldsymbol{w}(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$ is the most similar neuron to $\boldsymbol{x}$ (also called the winning neuron). Thanks to its topology-preserving property, samples in $\mathbb{R}^p$ which are not too far from each other are projected onto the map in the same area. Thus, the neurons $\boldsymbol{w}$ are located on the map according to their similarity. Moreover, it can be considered that the SOM achieves a non-uniform quantization that transforms $\boldsymbol{x}$ to $\boldsymbol{w}_{\boldsymbol{x}}$ by minimizing the given metric.

We present briefly the method for the construction of BBA from Kohonen's SOM. This method has been presented in details in [23], [24], and it has already been applied successfully for SPOT images classification. In this paper we apply it for joint classification of optical and SAR images.

We consider a given FoD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\}$ which correspond to the set of classes for the pixels in the images under analysis. The aim is to classify each pixel into one of these classes. Typically, θ_1 may represent the cereal class, θ_2 the bare soil class, etc. The BBA construction requires a SOM and an initial classification that defines the classes θ_i, $i = 1, \ldots, K$. The class centers in $\mathbb{R}^p$ are denoted as $\mathcal{C}_1, \ldots, \mathcal{C}_K$ while their projection onto the map are associated with the weighting vectors $\{\boldsymbol{w}_{\mathcal{C}_1}, \ldots, \boldsymbol{w}_{\mathcal{C}_K}\}$. The mass of the class θ_i for $i = 1, \ldots, K$ is defined directly on the map by

$$\begin{cases} m(\boldsymbol{x} \in \theta_i) \sim 1, & \text{if } \boldsymbol{w}_{\boldsymbol{x}} = \boldsymbol{w}_{\mathcal{C}_i} \\ m(\boldsymbol{x} \in \theta_i) \sim \dfrac{d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_i})^{-1}}{\sum_{\ell=1}^{K} d_{\text{map}}(\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_\ell})^{-1}}, & \text{otherwise} \end{cases} \tag{2}$$

where $d_{\text{map}}(\cdot,\cdot)$ stands for the distance along Kohonen's map. It is mainly based on Euclidean norm, and it uses the index that locates the two vectors on the map. That is

$$d_{\text{map}}(\boldsymbol{w}_1, \boldsymbol{w}_2) = \sqrt{(i_{\boldsymbol{w}_1} - i_{\boldsymbol{w}_2})^2 + (j_{\boldsymbol{w}_1} - j_{\boldsymbol{w}_2})^2} \tag{3}$$

if $\boldsymbol{w}_1$ (resp. $\boldsymbol{w}_2$) is located at position $(i_{\boldsymbol{w}_1}, j_{\boldsymbol{w}_1})$ (resp. $(i_{\boldsymbol{w}_2}, j_{\boldsymbol{w}_2})$) on the map.

The mass of compound hypotheses, say $m(\boldsymbol{x} \in \theta_k \cup \theta_\ell)$, is defined directly in the feature space because our definition of disjunctions of hypotheses expresses the lack of discrimination. We relate this mass value to a scaling effect between the sample $\boldsymbol{x}$ under concern and the disjunction of classes θ_k and θ_ℓ as follows

$$m(\boldsymbol{x} \in \theta_k \cup \theta_\ell) \sim 1 - \tanh(\beta \cdot z) \tag{4}$$

with

$$z = \frac{d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)}{d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)} \qquad 0 < k, \ell \leqslant K, k \neq \ell.$$

and where the β parameter stands for the level of ambiguity. $d_{\mathbb{R}^p}(\cdot,\cdot)$ denotes a distance metric in $\mathbb{R}^p$. It can be defined through Euclidean norm $\mathcal{L}^2(\mathbb{R}^p)$, but also through a spectral perspective, such as the spectral angle mapper or the spectral information divergence. It may also be based on Kullback-Leibler divergence or mutual information, when dealing with SAR [25].

The justification for modeling $m(\boldsymbol{x} \in \theta_k \cup \theta_\ell)$ by (4) is as follows. If a sample $\boldsymbol{x}$ is very close to the class center $\mathcal{C}_k$ in comparison with any other class center $\mathcal{C}_\ell$, then there is no ambiguity in believing that $\boldsymbol{x}$ belongs to class θ_k, whereas if the distances between $\boldsymbol{x}$ and the class centers $\mathcal{C}_k$ and $\mathcal{C}_\ell$ are of the same scale then it becomes uncertain to commit $\boldsymbol{x}$ with the class θ_k or with the class θ_ℓ, and it is more prudent in this case to commit $\boldsymbol{x}$ with the disjunction $\theta_k \cup \theta_\ell$.

The mass committed to the total ignorance Θ is based on the distance of a sample $\boldsymbol{x}$ to the map. It therefore requires a comparison of the distances in the feature space, as well as the distances in Kohonen's space. It can be expressed as follows:

$$m(\boldsymbol{x} \in \Theta) \sim 1 - \min\left(\frac{d_{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{w}_{\boldsymbol{x}})}{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})}, \frac{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}}, \boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})}{d_{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{w}_{\boldsymbol{x}})}\right) \tag{5}$$

where $\mathcal{C}_{\boldsymbol{x}}$ is the class center of $\boldsymbol{x}$, $\boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}}$ is its projection on the map. The formula (5) can be used as an outlier detector. In order to work with normalized BBAs in the fusion process, we apply a classical normalization step.

Ideally, if we consider that the both sources are totally reliable, this BBA construction could be applied separately with I_{MS} (using a SOM_{MS} and Euclidean distance metric) to build $m_{\text{MS}}(\cdot)$ for each pixel of I_{MS}. Similarly, one could apply this method with $I_{\kappa_{\text{MS}}}$ (using a SOM_{SAR} and Euclidean distance metric) to build $m_{\text{SAR}}(\cdot)$ for each pixel of $I_{\kappa_{\text{MS}}}$. In practice however, the things are no so perfect and SAR images appear as non-reliable source of information because of speckle noise which makes their extracted features less reliable. The optical images can also appear as partially

unreliable in the areas where there are clouds or cloud shadows because the optical information can be missing at these pixel locations. To overcome this serious problem a *joint SOM* is defined through a hybrid neuron definition that takes into account the p spectral signatures of the optical data, and the six texture descriptors of the SAR data.

A. Joint and enslaved Kohonen's SOM for BBA construction

In a general context, let $\boldsymbol{x} = \{x_1, x_2, \ldots, x_p\} \in \mathbb{R}^p$ and $\boldsymbol{y} = \{y_1, y_2, \ldots, y_q\} \in \mathbb{R}^q$ be the two heterogeneous observations provided by two heterogeneous sensors. The input samples of the proposed hybrid SOM are done through the co-located observations $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$ with which a distance must be associated. This (hybrid-joint) distance is a linear combination of two metrics defined respectively in $\mathbb{R}^p$ and in $\mathbb{R}^q$, that is

$$d(\boldsymbol{z}, \boldsymbol{z}') \triangleq d_{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{x}') + \alpha \cdot d_{\mathbb{R}^q}(\boldsymbol{y}, \boldsymbol{y}'), \tag{6}$$

with $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{z}' = (\boldsymbol{x}', \boldsymbol{y}')$ belong to $\mathbb{R}^{p+q}$. The parameter α is a cross-calibration factor, that accounts for the relative dynamics between $\boldsymbol{x}$ and $\boldsymbol{y}$. Since $d(\boldsymbol{z}, \boldsymbol{z}')$ is defined though a cartesian product metric space and two distances $d_{\mathbb{R}^p}(\cdot, \cdot)$ and $d_{\mathbb{R}^q}(\cdot, \cdot)$, it remains a distance.

According to this definition of a joint feature space and its related distance metrics, it is possible to perform a training of a joint SOM where the weighting vectors are defined with $\boldsymbol{w}_{\boldsymbol{z}} = (\boldsymbol{w}_{\boldsymbol{x}}, \boldsymbol{w}_{\boldsymbol{y}}) \in \mathbb{R}^{p+q}$. Nevertheless, this joint processing of our heterogeneous data does not take into account, at this stage, the reliability of the sources. Optical and SAR data interfere in the same manner in the location of each class center on the map (the class center of winning neurons $\boldsymbol{w}_{\mathcal{C}_k}$, $k \in \{1, 2, \ldots, K\}$), while SAR data is much less reliable than optical data in the land cover classification accuracy (but in case of missing optical information due to cloud occultation of course). Then, instead of a joint processing, an enslaved processing is set up to perform SOM training and a SOM-based BBA of the SAR data only.

Enslaved SOM training starts with a classical SOM training of the optical data only, which generates a SOM_{MS}. Then, the neurons of SOM_{MS} are concatenated by the $q = 6$ texture-based SAR feature components to fit the $\mathbb{R}^{p+q}$ of the joint processing. The training of this hybrid map begins, but only the last q-components (dedicated to the SAR data) are modified. In this case, the optical part is preserved, while the SAR part follows the optical part in the location of classes on the map (locations of the winning neurons $\boldsymbol{w}_{\mathcal{C}_k}$). This defines $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ from which the BBA $m_{\text{SAR}}(\cdot)$ is built for each pixel of the SAR image. The scheme of fusion of reliable[4] multispectral optical and SAR image is given in Fig. 2.

B. Dealing with missing optical data

When the optical sensor acquires a scene in the presence of clouds, two kinds of missing data must be considered: the parts of the data that are hidden by the clouds themselves, and the parts that are affected by the shadow of the clouds. In this case, a binary mask allows the training of Kohonen's map using valid (reliable) optical data only. The BBA of pixels located in occulted areas could be simply modeled by a vacuous BBA, that is by $m_{\text{MS}}(\Theta) = 1$. However this standpoint ignores the joint observation of optical and radar sensors. In fact, the optical pixel may be recovered by using the joint Kohonen's map $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ which reflects the links between optical and radar parts in the observation. Based on this remark, when a pixel $\boldsymbol{x}_{\text{MS}}$ is considered missing in the optical image due to the presence of clouds or shadow, the co-located radar observation $\boldsymbol{y}_{\text{SAR}}$ is considered. Its winning neuron in the radar restriction of $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ allows us to consider the optical part of Kohonen's map. This spectral signature is substituted for $\boldsymbol{x}_{\text{MS}}$ to recover the missing information.

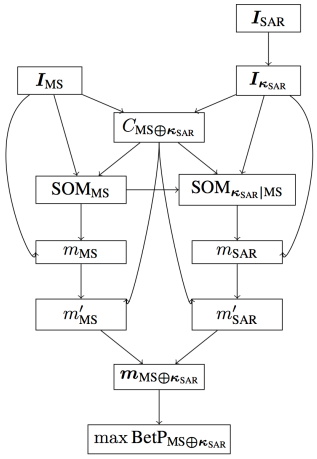


Figure 2. Fusion framework between a reliable optical multispectral image and a SAR image.

IV. Optical and radar joint classification

For each pixel of the *fused image*, the fusion process consists in combining the BBAs $m_{\text{MS}}(\cdot)$ and $m_{\text{SAR}}(\cdot)$ built by the method presented in the previous section. In order to take into account the reliability of the optical and radar sources in the fusion process, we apply a different discounting on the BBA $m_{\text{MS}}(\cdot)$ and on $m_{\text{SAR}}(\cdot)$ before their combination. We apply a contextual discounting [13] of $m_{\text{MS}}(\cdot)$ and an importance discounting [14] of $m_{\text{SAR}}(\cdot)$ to better fuse the two heterogeneous sources of information that do not yield the same kind of physical information with a different level of discrimination efficientcy. More precisely, before making the combination of the two BBAs, we discount them as follows:

- Contextual discounting of $m_{\text{MS}}(\cdot)$

$$m'_{\text{MS}}(X) = \sum_{Y_1, Y_2 \in 2^\Theta | Y_1 \cup Y_2 = X} m_{\text{MS}}(Y_1) m^{\lambda_i}_{\theta_i}(Y_2) \tag{7}$$

where $m^{\lambda_i}_{\theta_i}(\cdot)$ represents the BBA attributed to $\theta_i \in \Theta$ which enters in the contextual discounting method. $\lambda_i \in [0, 1]$ is a tuning parameter associated with the class θ_i. This BBA is defined by $m^{\lambda_i}_{\theta_i}(\emptyset) = \lambda_i$ and $m^{\lambda_i}_{\theta_i}(\theta_i) = 1 - \lambda_i$.

[4]Here we assume that the information at a given pixel in the optical image is not missing, i.e. there is no cloud occultation effect at this pixel.

- Importance discounting of $m_{\text{SAR}}(\cdot)$

$$\begin{cases} m'_{\text{SAR}}(\emptyset) = 1 - \beta \\ m'_{\text{SAR}}(X) = \beta \cdot m_{\text{SAR}}(X), \text{if } X \in 2^{\Theta} \setminus \{\emptyset\} \end{cases} \quad (8)$$

where β is a chosen importance factor in $[0, 1]$.

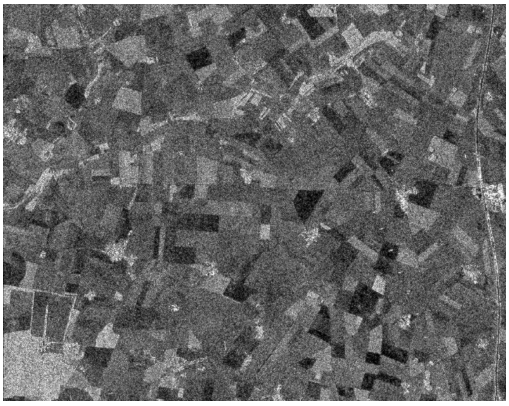
The fusion is achieved by combining the discounted BBAs $m'_{\text{MS}}(\cdot)$ and $m'_{\text{SAR}}(\cdot)$ with PCR6 rule [20], followed by a normalization step as explained in [14]. The justification of PCR6 rule comes from the fact this rule makes a rational redistribution of partial conflict proportionally only to the masses of propositions that are involved in the conflict. Hence, PCR6 rule offers the preservation of the specificity of the information in the fusion process. PCR6 is preferred to Dempster's rule because Dempster's rule provides counter-intuitive results [18], and it does not respond to importance discounting as shown in [14]. The result of the combination of $m'_{\text{MS}}(\cdot)$ and $m'_{\text{SAR}}(\cdot)$ with PCR6 is denoted symbolically as $m_{PCR6}(\cdot) = [m'_{\text{MS}} \oplus m'_{\text{SAR}}](\cdot)$. The normalization step after applying PCR6 is necessary in our application because $m'_{\text{MS}}(\emptyset) > 0$ implies $m_{PCR6}(\emptyset) > 0$. Therefore, after applying PCR6 rule we normalize all masses values $m_{PCR6}(X)$ for $X \neq \emptyset$ by $m_{PCR6}(X)/(1 - m_{PCR6}(\emptyset)) \rightarrow m_{PCR6}(X)$, and we set $m_{PCR6}(\emptyset) = 0$ in order to get a fused and normalized BBA from which the final decision will be made. In this work, we adopt the classical maximum of betting probability, denoted $BetP(\cdot)$, decision-making strategy introduced in [26]. We use this decision-making strategy because it is simple to implement, and it offers a compromise between the pessimistic (max of belief) and the optimistic (max of plausibility) decisional attitudes. More precisely, we classify the pixel under analysis in the class $\theta_i \in \Theta$, if $BetP(\theta_i) = \max_j\{BetP(\theta_j)\}$.

V. Experimental results

We present the results of joint classification obtained with this new fusion method applied on real SPOT-5 and RADARSAT-2 images taken at almost same period of time (April 2015). Our study covers a $11.5 \times 9\text{km}^2$ area of the flat agricultural Beauce region, located in the south-west of Paris, France which is characterized by very large fields dominated by rape and cereal (wheat, barley, corn) crops. In this experience, the multispectral image I_{MS} is acquired by the SPOT-5 French satellite, and the SAR iamage I_{SAR} is acquired by the RADARSAT-2 Canadian satellite in Ultra-Fine mode. The SPOT-5 image is characterized by a size of 1145×903 pixels, with a spatial resolution of 10m, and has four bands (Green (G), Red (R), Near InfraRed (NIR) and Medium InfraRed (MIR)). The RADARSAT-2 image is composed of 3850×3010 pixels, with a spatial resolution of 3m. Both HH and HV polarization channels are available in the SAR image, but we did use only HH polarization because it interacts more efficiently than HV-polarization with agricultural crops. Fig. 3-(a) and Fig. 3-(b) show, respectively, the false color composite of the SPOT image and its registered SAR image. We distinguish five different land cover types (classes) in Fig. 3(a):

- C_1: The brown class C_1 corresponds to wooded areas. C_1 appears in light grey on SAR image of Fig. 3(b).
- C_2: The dark red field class C_2 corresponds mainly to durum wheat (planted in winter). This class has no explicit signature in the SAR image.

(a) SPOT5/Take5 data acquired on April 20th, 2015. False color composite: RGB = (NIR, R, G) ©CNES



(b) Registered RADARSAT-2 HH F5 mode Ascending acquired on April 23rd, 2015. RADARSAT-2 Data ans Products ©MacDONALD, DETTWILER and ASSOCIATES LTD – All Rights Reserved

Figure 3. Multispectal (a) and radar (b) used in the experiment (Beauce region, France).

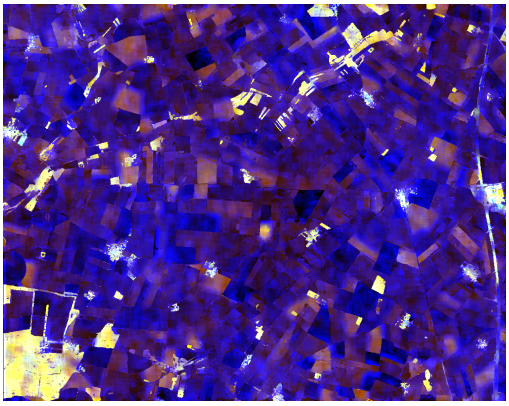


Figure 4. $\boldsymbol{I}_{\boldsymbol{\kappa}_{\text{SAR}}}$ image corresponding to the SAR texture information with false color composition: RGB=(μ, σ, β_1)

- C_3: The light red fields class of Fig. 3 (a), corresponds mainly to rape. Class C_3 is in light grey on Fig. 3 (b).
- C_4: The cyan fields class C_4 corresponds to bare soils. C_4 appears as dark in 3(b). In fact, they correspond mainly to corn and cereal seedling.
- C_5: The grey fields class C_5 corresponds mainly to barley (planted in early spring). This class has no significant signature in the radar image.

A. Preprocessing of SAR image

The fusion process is achieved at the coarser resolution of both images, that is at the resolution of SPOT5 image (i.e. 10m). For this, the RADARSAT-2 image is first processed in order to extract the local statistical parameters $(\mu, \sigma, \beta_1, \beta_2, f_6, f_5)$); the processing is done through a sliding window of 51×51 for $(\mu, \sigma, \beta_1, \beta_2)$ and 15×15 for the Haralick texture parameters (f_6, f_5) estimation. In order to prevent bordering effects between parcels, a naive map extracted from the multispectral image serves as a mask in the local parameter estimation of radar data. This guarantees a parameter estimation on effective homogeneous areas and preserves the borders of each parcel. The choice of analysis window size is based on our object dimensions of interest. Therefore, our analysis windows size is proportional to those field dimensions. The 3m-resolution SAR feature image is then downsampled to a 10m-resolution SAR feature image, and then registered to the SPOT geometry. Fig. 4 shows a false color composition of the radar information at a 10m resolution. The color composition is shown with RGB=(μ, σ, β_1). In comparing visually Fig. (4) with (3)(a), one sees that C_1 class (wooded area) appears in yellow in Fig. (4), C_3 class (rape) appears in light brown and C_4 appears in dark magenta.

B. Fusion & joint classification results without missing data

Here we consider the simplest case where there is no missing data in the optical image when there is no cloud nor shadow effect. In order to merge the belief degrees associated with each pixel from the two input images, a unified FoD is required. The simple classes of this FoD are defined using the K-means unsupervised classifier [28], where the parameter K is set to 5. One applies the K-means classifier to a stack image collecting the spectral information of the SPOT image I_{MS} and the texture information of the (downsampled and registered) SAR image $I_{\kappa_{\text{SAR}}}$. We have taken the cross-calibration factor $\alpha = 5.10^{-3}$ in (6) because it corresponds to the average ratio between the mean value of the optical and radar data. The ground truth was collected in July 2015, while the data were acquired in April, and as a result, any ambiguity between different kinds of crops could not be resolved, as many fields were still in the seedling state. Hence, it was decided that only five classes could be discriminated. The results of the joint classification are shown in Fig. 5. This classification is used as reference image (the ground truth of classification result) for the performance analysis of our fusion method.

Fig. 6 presents the Kohonen's maps SOM_{MS} and $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ trained with the optical and the SAR information.

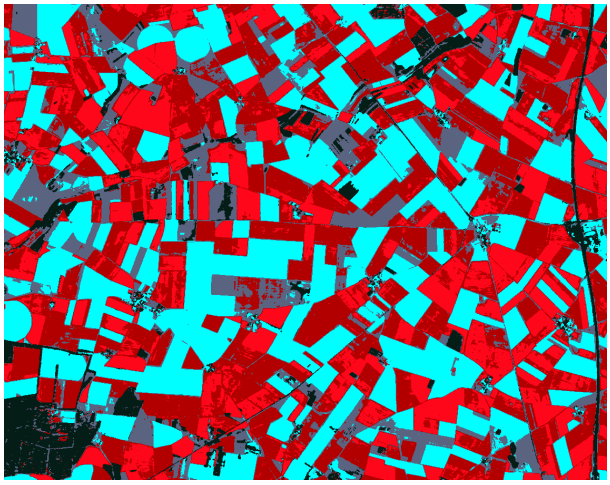
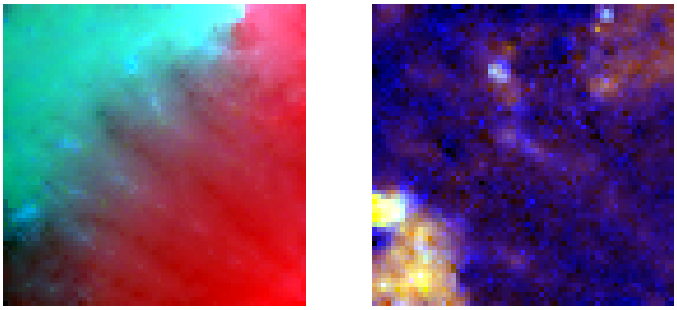


Figure 5. Reference image built from Unsupervised K-means classification (with $K = 5$ classes) applied jointly on multispectral and SAR information

These 65×65 neuron maps were trained with 5000 samples per class. The initial learning rate and neighborhood size were set respectively to 1 and 60.



(a) SOM_{MS} RGB=(NIR, R, G)

(b) $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ RGB=(μ, σ, β_1)

Figure 6. SOM maps of size 65×65 neurons. The map (a) has been trained with the optical data only, while the map (b) has been enslaved to map (a) and trained with radar data. Hence, co-located neurons bring the same ground information between the 2 maps

SOM_{MS} in Fig. 6-(a), shows essentially the distribution of spectral signatures representing cover soils (in red) and bare soils (in cyan) in this farming area. In the marginal dark zone between these two types of spectral signatures, there is a location dedicated to man-made structures (buildings and roads) and a forest area (darker, at the bottom left area of SOM_{MS}). The enslaved map $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ of Fig. 6-(b), shows the same kind of neurons at the same location on the map, viewed by the textural parameters extracted from SAR data.

One sees that the purple area in the middle of $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ map, appears quite homogeneous, while we have two very different areas in SOM_{MS} map. This illustrates the fact that the optical sensor is mainly sensitive to the presence of chlorophyll in this farming area, while the radar is sensitive to the surface roughness. However, surface roughness may appear similar, from a radar observation, in bare soil and also in cover fields, depending on the plantation. Nevertheless, the wooded area

is clearly discriminated from SAR sensor. The wooded area appears in brown-yellow in Fig. 4 and at the bottom left of $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ in Fig. 6-(b). The SAR sensor does not help in the discrimination between bare soil and cover soil; nevertheless, it easily discriminates the 2 kinds of cover fields that appear in red from a SPOT point of view (e.g., Fig. 5), as an area at the top right of the $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ map appears in brown. It helps to do discrimine between corn seed in winter and in early spring. The use of the enslaved radar part of $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$ allows the Credal classification to tackle this ambiguity in order to improve the final joint classification.

From Kohonen's maps SOM_{MS} and $\text{SOM}_{\kappa_{\text{SAR}}|\text{MS}}$, the estimation of BBAs is then performed. The contextual discounting factors λ_i involved in (7) were calculated using results given by the confusion matrix [27] derived from the cross decisions given by the optical source and the decisions of the reference data only. Each λ_i is estimated using the percentage of correct classifications of the target class θ_i. In our application, the contextual reliability factors are: $\lambda_1 = 0.85$, $\lambda_2 = 0.86$, $\lambda_3 = 0.52$, $\lambda_4 = 0.9$, and $\lambda_5 = 0.87$. Based on the subjective attribute of SAR source, we have taken $\beta = 0.4$ as importance discounting factor for discounting the BBA $m_{\text{SAR}}(\cdot)$ with (8). It maximizes class discrimination of the overall process.

Fig. 7 shows the classification result drawn from the fusion of optical and SAR information based on PCR6 rule and max of $BetP(\cdot)$ decision-making strategy. The measure of performance (i.e. confusion matrix) is established by comparing the result of Fig. 7 with the classification ground truth of Fig. (5). A Correct Classification Rate (CCR) of 77.25% has been evaluated from Table I, and a Kappa Index of 0.74 which is considered to a good agreement in [29].

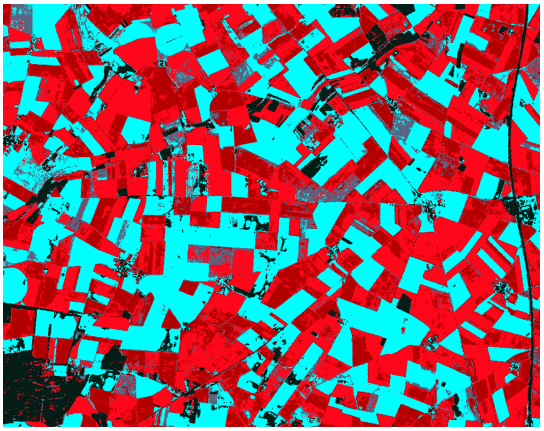


Figure 7. Joint classification result with PCR6 and max of $BetP$ strategy.

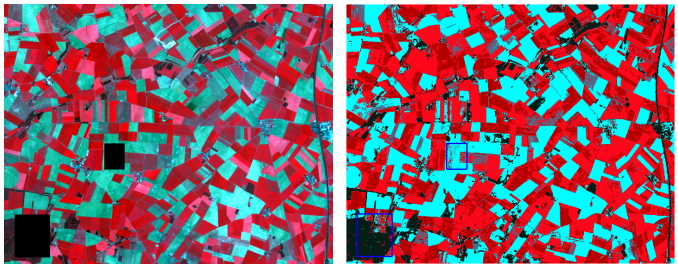
The accounting for source reliability for each class allows the decision making to mitigate ambiguous fields such as those which are slightly covered, but may still be considered as bare soils, or the seedling fields which may not have the same surface roughness. This impacts mainly the class C_5 (fields mainly barley), which can be classified as C_2 or C_3, depending on the roughness signature and density of chlorophyll.

Table I
QUANTITATIVE RESULTS OBTAINED USING CONFUSION MATRIX

	C_1	C_2	C_3	C_4	C_5
C_1	95.35	0.00	1.94	2.69	0.02
C_2	0.02	60.65	34.83	0.11	4.39
C_3	0.00	4.87	95.13	0.00	0.00
C_4	4.59	0.00	0.04	95.36	0.01
C_5	1.75	9.04	17.88	31.55	39.78

C. Fusion & joint classification results with missing data

The same analysis has been done with missing data in the optical image. For this, we have applied artificial occultation masks on SPOT5 image to simulate the presence of clouds or shadows. Two masked regions, namely, zone 1 and zone 2, were selected in order to hide different kinds of ground cover. The mask of zone 1 mainly covers a non-agricultural area composed of 18,056 pixels, while the mask of zone 2 covers some covered and some bare soils composed of 6,300 pixels, as shown in Fig. 8(a). The joint classification result is shown in Fig. 8(b).



(a) The original SPOT image with missing data at the location of the mask.

(b) Classification of the heterogeneous data with missing data.

Figure 8. Results for joint classification with simulated cloud cover. The decision is performed by the maximum of pignistic probability over all simples hypotheses

A quantitative analysis gives an overall classification accuracy of CCR = 73.94% which is very close to CCR = 77.25% obtained when processing the cloud-free SPOT5 image with SAR image.

VI. CONCLUSION

A new method has been proposed for combining complementary information drawn from optical and radar remote sensing data for land cover classification in agricultural areas. The proposed approach refers to the direct combination of heterogeneous data at the pixel level in the features space, and considers the image registration problems as resolved. This new approach is based on hybrid training of Kohonen's map using heterogeneous data for mass functions estimation. This step helps to deal with the heterogeneity of data sources by representing those in the same semantic meaning through co-located observations. The methodology benefits from this joint training of heterogeneous data to restore missing parts of

optical data. It is worth noting that the enslaved processing is relevant when one of two sources of information is considered to be more accurate than the other. When the cloud coverage is too significant, the related source of information can no longer be considered as sufficiently accurate. In that case, a joint processing should be preferred in order to recover missing data. Our experimental tests on real images showed the benefit of heterogeneous joint processing in the analysis of a complex farming area, even in the case of missing data. A deeper analysis should be performed to better link the heterogeneous data fusion process with Credal theory and the hybrid training of Kohonen map when joint training data is available.

APPENDIX - SAR TEXTURE DESCRIPTORS

Let $p(i)$, $i = 0, 1, ..., N_g - 1$ be the pmf of the intensity levels occurrence, where N_g is the number of gray levels. $p(i)$ is calculated by dividing the histogram value $h(i)$ by the total number of pixels of the odd sliding window of size $w \times w$, that is $p(i) \triangleq h(i)/w.w$. We use the following SAR texture descriptors:

- The mean μ: It is the average level of intensity of the region of texture

$$\mu = E[i] = \sum_{i=0}^{N_g-1} i \times p(i)$$

- The standard deviation σ: it describes the variation of intensity around the mean

$$\sigma = \sqrt{E[(i-\mu)^2]} = [\sum_{i=0}^{N_g-1} (i-\mu)^2 \times p(i)]^{-1/2}$$

- The skewness β_1: It describes how symmetric the intensity distribution is around the mean

$$\beta_1 = E\Big[\big(\frac{i-\mu}{\sigma}\big)^3\Big] = \sum_{i=0}^{N_g-1} \frac{(i-\mu)^3}{\sigma^3} \times p(i)$$

- The normalized kurtosis β_2: It measures the flatness excess of the intensity distribution

$$\beta_2 = E\Big[\big(\frac{i-\mu}{\sigma}\big)^4\Big] = \sum_{i=0}^{N_g-1} \frac{(i-\mu)^4}{\sigma^4} \times p(i) - 3$$

- Haralick's descriptor f_5: It is the inverse difference moment which measures local similarities in the image

$$f_5 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} \times p(i,j)$$

where $p(i,j)$ is the value of the Gray Level Cooccurrence Matrix (GLCM) at the cell (i,j) [9].

- Haralick's descriptor f_6: It is the sum average descriptor defined as

$$f_6 = \sum_{i=2}^{2N_g} \Big[i \times \sum_{\substack{k=0 \\ k+\ell=i}}^{N_g-1} \sum_{\ell=0}^{N_g-1} p(k,\ell) \Big]$$

REFERENCES

[1] C. Pohl, J. van Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications, IJRS, Vol. 19(5), pp. 823–854, 1998.

[2] Z. Jixian, Multi-source remote sensing data fusion: status and trends, Int. J. of Image and Data Fusion, Vol. 1(1), pp. 5–24, 2010.

[3] A. Dromigny-Badin, Image fusion using evidence theory: applications to medical and industrial images, Ph.D. Thesis, INSA, Lyon, France, 1998.

[4] J. Dong, et al., Survey of multispectral image fusion techniques in remote sensing applications, Chinese Academy of Sciences Journal, Beijing, China, 2011.

[5] T. Kohonen, The self-organizing map. Proceedings of the IEEE, Vol. 78(9), pp. 1464–1480, 1990.

[6] M.A. Kraaijveld, et al., A nonlinear projection method based on Kohonen's topology preserving maps. IEEE Trans. on Neural Networks, Vol. 6(3), pp. 548–559, 1995.

[7] C. Oliver, S. Quegan, Understanding synthetic aperture radar images, Artech house, USA, 1998.

[8] A. Materka, M. Strzelecki, Texture analysis methods – a review, Inst. of Electronics, Techn. Univ. of Lodz, Tech. Rep., 1998.

[9] R. Haralick, et al., Texture features for image classification, IEEE Trans. on SMC, Vol. 3(6), 1973.

[10] G. Shafer, A Mathematical Theory of Evidence, Princeton Press, 1976.

[11] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. of Math. Stat., (38):325-339, 1967.

[12] A. Dempster, A generalization of Bayesian inference, J . of Royal Stat. Soc., (B30):205-247,1968.

[13] D. Mercier, et al., Contextual discounting of belief functions, Proc. of ECSQARU'05, pp. 552–562, Springer-Verlag, 2005.

[14] F. Smarandache, et al., Fusion of sources of evidence with different importances and reliabilities, Proc. of Fusion 2010, UK, July 2010.

[15] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, Why Dempster's fusion rule is not a generalization of Bayes fusion rule, Fusion 2013 Proc.

[16] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, Why Dempster's rule doesn't behave as Bayes rule with informative priors, INISTA 2013.

[17] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, Int. J. of Intell. Syst., (29):223-252, 2014.

[18] J. Dezert, P. Wang, A. Tchamova, On the validity of Dempster-Shafer theory, Fusion 2012 Proc., Singapore, July 9–12, 2012.

[19] J. Dezert, A. Tchamova, D. Han, Total Belief Theorem and Generalized Bayes' Theorem, Fusion 2018 Proc., Cambridge, July 2018.

[20] F. Smarandache, J. Dezert, Advances and applications of DSmT for information fusion, Vol.3, Chap. 4, American Research Press, 2009. https://www.onera.fr/staff/jean-dezert/references

[21] J. Dezert, et al., Decision-Making with Belief Interval Distance, Proc. of Belief 2016 Int. Conf., Prague, CZ, Sept. 21-23, 2016.

[22] M.A. Kraaijveld, et al., A nonlinear projection method based on Kohonen's topology preserving maps, IEEE Trans. on NN, Vol. 6(3), pp. 548–559, May 1995.

[23] I. Hammami, et al., The Kohonen map for Credal Classification of Large Multispectral Images, Proc. of IGARSS?2014, pp. 3706–3709, July 2014.

[24] I. Hammami, et al., Kohonen's map approach for the belief mass modeling, IEEE Trans. on NN & Learning Systems, Vol. 27(10), pp. 2060–2071, Oct. 2016.

[25] J. Inglada, G. Mercier, A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and its Extension to Multiscale Change Analysis, IEEE Trans. on Geoscience and Remote Sensing, Vol. 45(5), pp. 1432–1446, May 2007.

[26] P. Smets, R. Kennes, The transferable belief model, Artif. Int., Vol. 66, pp. 191–234, 1994.

[27] Z. Elouedi, et al., Discountings of a belief function using a confusion matrix, Proc. of Int. Conf. on Tools with AI, Vol. 1, pp. 287–294, Oct. 2010.

[28] J. A. Hartigan, M. A. Wong. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, Series C. vol 28 (1): pp. 100–108. 1979.

[29] Cohen, Jacob. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. vol 20 (1): pp 37–46, 1960.