# MULTISOURCE CLUSTERING OF REMOTE SENSING IMAGES WITH ENTROPY-BASED DEMPSTER-SHAFER FUSION

*S. Ranoeliarivao, F. de Morsier*[⋆]*, D. Tuia*[‡]*, S. Rakotoniaina, M. Borgeaud*[°]*, J.-Ph. Thiran*[⋆] *and S. Rakotondraompiana*

Institute and Observatory of Geophysics Antananarivo (IOGA), University of Antananarivo, Madagascar
[⋆] Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne, Switzerland
[‡] LaSIG, École Polytechnique Fédérale de Lausanne, Switzerland
[°] European Space Agency, ESRIN, Frascati, Italy

## ABSTRACT

In this paper, we propose a strategy for fusing clustering maps obtained with different remote sensing sources. Dempster-Shafer (DS) Theory is a powerful fusion method that allows to combine classifications from different sources and handles ignorance, imprecision and conflict between them. To do so, it attributes evidences (weights) to different hypothesis representing single or unions of classes. We introduce a fully unsupervised evidence assignment strategy exploiting the entropy among cluster memberships. Ambiguous pixels get stronger evidences for union of classes to better represent the uncertainty among them. On two multisource experiments, the proposed Entropy-based Dempster-Shafer (EDS) performs best along the different fusion methods with VHR images, when the single class accuracies from each source are complementary and one of the sources shows low overall accuracy.

***Index Terms*—** Dempster-Shafer, multisource fusion, unsupervised, entropy, fuzzy C-Means, remote sensing

## 1. INTRODUCTION

Classification in remote sensing can provide thematic information on large regions (e.g. spatial distribution of forest against urban areas). Many sources exist nowadays, so that images from different sensors are often available over the same region at a similar time period. In the optical domain, the images are redundant, but show complementarities that can be exploited. In the last decade, the fusion of remote sensing images has been intensely investigated, in particular under the umbrella of *pansharpening*, which is the fusion of high spatial resolution images (panchromatic) with multispectral or hyperspectral images [1]. Fusion of different classifiers [2] or of classifications from multiple sources [3] have also received wide attention in remote sensing. The complementarity between images with different spectral bands and sensitivity can lead to better class discrimination and more accurate classification maps after their fusion.

Multisource fusion can be performed at different levels, either at the pixel or features level or as a decision fusion tool, thus combining several single-source classification maps [4]. The Dempster-Shafer (DS) fusion methodology is of the latter type. DS theory of evidence is a powerful fusion method having the advantage of handling ignorance (imprecision and uncertainty) and conflict between the sources [5], [6]. DS theory attributes *evidences* (weights) to a set of different hypothesis, being single classes or unions of classes. In case of ambiguity, higher weights are given to the union of several classes than on a single class. This potentially allows to unveil the ambiguity after the combination with another source. The DS theory has been successfully applied to the segmentation fusion of biomedical images [7] and the classification fusion of remote sensing images [4], [8].

While the DS theory for fusion is well established, the way of deriving the evidences from the classification varies. Some methods are theoretically founded such as those proposed by Dubois and Prade [9] to derive evidences from likelihood and probability according to the least Commitment Principle, and by Bloch [10] to allocate evidences from mathematical operators, and that of Mercier et al [11] to discount evidences in a contextual way. For instance, evidences can also be assigned using a validation set if an accurate ground truth is available [12].

When such ground reference is not available, one must recour to unsupervised assignment. In this case, the assignment of evidences to the different hypothesis can be derived from cluster memberships [13, 4]. The ambiguity among the two highest memberships [14] and a threshold separating high from low ambiguity situations [7] have been considered in the past to assign evidences to unions of classes. However, and for more than two classes, the ambiguity should be defined in a continuous way and not only by considering two situations (low or high ambiguity). Moreover, when using the current approaches, it is difficult to set a threshold which is common to all pixels, because ambiguity can also vary locally.

In this paper, we tackle the aforementioned weaknesses in the context of unsupervised classification, by considering local assignments of DS evidences. The ambiguity threshold is replaced by an ambiguity factor weighting the evidences, based on the entropy among the fuzzy memberships. This provides a natural multiclass measure to assess uncertainty among several classes.

## 2. FUZZY CLUSTERING

Let us consider the pixels $\mathbf{x}^s(j) \in \mathbb{R}^{d^s}$ of two co-registered images from different sensors $s$ with associated labels $y(j) = C_i, 1 \leq i \leq N$ corresponding to the $N$ different classes. Fuzzy cluster memberships $\mu_{ij} \in [0, 1]$ are obtained from the Fuzzy C-Means (FCM) algorithm [13]. The clusters centers are matched before the fusion in order to remain coherent among the sources. We observed that the FCM clustering solutions were very stable regarding the initialization in our experiments.

## 3. DEMPSTER-SHAFER THEORY FOR DATA FUSION

The Dempster-Shafer theory has the advantage of considering imprecision and conflict between multisource information. The different hypothesis considered in the DS theory can even be union of classes. This can be seen as the ability of representing "mixed" pixel, resulting from a source unable to distinguish certain classes. Since we aim at distinguishing only the $N$ classes of interest, the unions of classes are not considered as potential final classes; however, the union of classes are considered in the fusion scheme to sort the ambiguous cases, that may be solved by the information brought by another source.

The evidences are the weights assigned to the hypothesis $H_i$, corresponding to the single class and union of classes. The evidences $m^s(H_i)$ for a source $s$ follow these conditions:

$$0 \leq m^s(H_i) \leq 1,\ m^s(\emptyset) = 0,\ \sum_{i=1}^{N} m^s(H_i) = 1 \qquad (1)$$

which are similar to standard probabilities properties. The fusion of the set of evidences $m^s(H_i)$ from each source (here $s = 1, 2$) is realized by the normalized orthogonal sum of the masses (2).

$$m(H_i) = \frac{1}{1-K} \sum_{H_p \cap H_q = H_i} m^1(H_p) \cdot m^2(H_q) \qquad (2)$$

$$K = \sum_{H_p \cap H_q = \emptyset} m^1(H_p) \cdot m^2(H_q) \qquad (3)$$

In other words, all the product of evidences between the sources contributing to an hypothesis ($H_i$) are summed. The union of classes are not considered as a possible final class. $K$ represents the conflicts between the sources.

After fusion of the evidences, a decision rule is used to define the final class assignment. In the literature, there are several decision rules such as maximum of plausability or credibility [8], [6]. The maximum of credibility, equivalent to the maximum of evidence, is chosen to result in labels being the most probable class [15].

$$y(j) = \arg\max_{Ci} \{m(C_i)(j), 1 \leq i \leq N\} \qquad (4)$$

## 4. UNSUPERVISED EVIDENCE ASSIGNMENT

The unsupervised determination of evidences is based on the fuzzy membership values without requiring any groundtruth information. The union of classes allows to represent a certain ambiguity between classes. Depending on the ambiguity among the membership values, a proportional weight will be set on the evidences representing the corresponding union of classes.

Previously, in order to weight accordingly the evidence representing the union of the two classes $k$ and $l$, the ambiguity has been defined as the absolute difference among the two highest membership $k$ and $l$: $\left|\mu_{kj} - \mu_{lj}\right|$ [7, 14]. This was justified by the use of only gray level intensities of the images, reducing the problem of ambiguity to a problem involving maximally two clusters. The ambiguity was characterized as high or low with respect to a user-defined threshold $\varepsilon$ [7].

### 4.1. Ambiguity factor based on entropy

It is difficult to set a threshold on the ambiguity that is valid for all pixels: such a threshold would not account for situations where the ambiguity is among several memberships and be sensitive for ambiguity around $\epsilon$. We propose an ambiguity factor weighting the different evidences and avoiding the recourse to a binary threshold. The ambiguity is based on the entropy of the memberships distribution. The entropy reflects the internal organization of the memberships: a flat distribution will have a high entropy, a peaked distribution will have a low entropy. The ambiguity factor $\rho(j)$ is defined as the normalized entropy for pixel $j$:

$$\rho\,(j) = \frac{\sum_{i=1}^{N} \mu_{ij} \ln(\mu_{ij})}{\rho_{\max}} \qquad (5)$$

where the maximal entropy is $\rho_{\max} = \ln(N)$, met when all the memberships are equal: $\mu_{ij} = 1/N$.

### 4.2. Smooth evidence assignment

Inspired by Boudraa et al. [7] on the evidence assignment from the membership distribution, the ambiguity factor $\rho(j)$ is inserted to weight the different evidences and to avoid having a global binary threshold. The different evidences for the single classes – Eq. (6) – and for the union of classes –

Eqs. (7) and (9) – are weighted by the entropy-based ambiguity factor $\rho(j)$.
The evidences for the single classes are defined as

$$m(C_i)(j) = [1 - m_{\bar{k}}(j) - m_{kl}(j) - m_{\bar{kl}}(j)] \cdot \mu_{ij} \quad (6)$$

When it is not ambiguous, one membership is significantly more important and the others can be grouped in $m_{\bar{k}}(j)$, which is the evidence of the union of all memberships except the highest membership defined as

$$m_{\bar{k}}(j) = m(\bigcup_{\substack{i=1\\i\neq k}}^{N} C_i)(j) = (1-\rho(j)) \cdot \sum_{\substack{i=1\\i\neq k}}^{N} \mu_{ij} \cdot (\beta - \mu_{ij}) \quad (7)$$

with $\beta = \max_{1\leq k\leq N}(\mu_{kj})$. This evidence will be down-weighted in case of *high ambiguity* among the class memberships since it would become less relevant.

$m_{kl}(j)$ and $m_{\bar{kl}}(j)$ are the evidences related to the two highest memberships, corresponding to the classes $C_k$ and $C_l$. They are defined as follows

$$m_{kl}(j) = m(C_k \cup C_l)(j) = \rho(j) \cdot \alpha \cdot (\mu_{kj} + \mu_{lj})_{k\neq l} \quad (8)$$

$$m_{\bar{kl}}(j) = m(\bigcup_{\substack{i=1\\i\neq k, i\neq l}}^{N} C_i)(j) = \rho(j) \cdot \alpha \cdot \sum_{\substack{i=1\\i\neq k, i\neq l}}^{N} \mu_{ij} \quad (9)$$

where $\alpha = \beta - \min_{1\leq k\leq N}(\mu_{kj})$. These evidences are made proportional to the ambiguity factor $\rho(j)$. They will be down-weighted in the case of *low (or no) ambiguity* between memberships, and become significant when ambiguous.

Finally, the evidence conditions of Eq. (1) are respected through adequate normalization of the single class evidences $m(C_i)$ in Eq. (6). The mass of the union of all the classes is null under the hypothesis that the clustering function avoids creating regions where all classes overlap. The union of more than the two highest memberships could be taken into account, but this would lead to an extensive number of evidences with very low weights. This assignment of evidences boils down to the one presented in [7] if $\rho(j)$ is set to 1 or 0, in case of high or low ambiguity respectively.

## 5. EXPERIMENTS

The proposed Entropy-based Dempster-Shafer (EDS) fusion scheme has been tested on two optical remote sensing cases:

The *Geneva* dataset consists in a SPOT image with a spatial resolution of 20m and 3 spectral bands (Near infrared (NIR), Red (R) and Green (G) spectral channels), and a Landsat TM image with a spatial resolution of 30m and 6 spectral bands (from 450 nm to 2350 nm), acquired over Geneva, Switzerland, in 1990 and 1988 respectively. Images are 512×512 pixels after co-registration on a 20m pixel grid. A ground truth of 45311 samples is divided in four classes: 1. Water, 2. Farming, 3. Vegetation, 4. Urban (buildings, ground). First row of Fig. 3 illustrates the images.

|  | LAND | SPOT | STAC | SUM | PROD | ENT | ADS | EDS | SDS |
|---|---|---|---|---|---|---|---|---|---|
| OA | 76.83 | 74.18 | 77.59 | 78.41 | **78.85** | 78.17 | 78.49 | 78.83 | 78.90 |
| $\kappa$ | 0.622 | 0.584 | 0.637 | 0.645 | **0.650** | 0.642 | 0.645 | 0.647 | 0.644 |
| Water | **0.856** | 0.763 | 0.841 | 0.829 | 0.832 | 0.827 | 0.831 | 0.830 | 0.824 |
| Farm. | 0.199 | 0.190 | 0.187 | 0.223 | **0.226** | 0.220 | 0.224 | 0.221 | 0.230 |
| Veg. | 0.760 | 0.805 | 0.813 | 0.811 | **0.814** | 0.810 | 0.811 | 0.811 | 0.769 |
| Urban | 0.809 | 0.779 | 0.814 | 0.820 | **0.824** | 0.817 | 0.820 | **0.824** | 0.828 |

**Table 1**. Average results for 3-fold cross-validation of the "Geneva" images

|  | IKON | SPOT | STAC | SUM | PROD | ENT | ADS | EDS | SDS |
|---|---|---|---|---|---|---|---|---|---|
| OA | 52.52 | 71.15 | 72.61 | 72.07 | 71.92 | 72.16 | 73.17 | **73.69** | 74.96 |
| $\kappa$ | 0.385 | 0.622 | 0.643 | 0.633 | 0.632 | 0.635 | 0.648 | **0.656** | 0.662 |
| Water | 0.154 | **0.940** | 0.661 | 0.886 | 0.870 | 0.880 | 0.897 | 0.874 | 0.767 |
| Veg. | 0.519 | 0.446 | 0.482 | **0.536** | 0.508 | 0.534 | **0.536** | 0.524 | 0.631 |
| Farm. | 0.130 | 0.006 | **0.395** | 0.071 | 0.142 | 0.122 | 0.170 | 0.193 | 0.009 |
| Urban | 0.765 | 0.922 | **0.952** | 0.905 | 0.909 | 0.905 | 0.907 | 0.904 | 0.880 |
| Rice. | 0.603 | 0.507 | **0.728** | 0.536 | 0.584 | 0.554 | 0.560 | 0.639 | 0.699 |

**Table 2**. Average results for 3-fold cross-validation of the "Tana" images

The *Tana* dataset consists in a SPOT image with a spatial resolution of 2m50 and 3 spectral bands (NIR-R-G), and an Ikonos image with a spatial resolution of 1m and the 3 visible spectral bands (R-G-B), both taken in july 2006 over Antananarivo, Madagascar. Images are 2000×2000 pixels after co-registration on a 2m50 pixel grid. A ground truth with 4243 samples collected and divided in five classes: 1. Water, 2. Ricefield, 3. Vegetation (mango, eucalyptus, fir), 4. Farming, 5. Urban (buildings, ground). First row of Fig. 4 illustrates the images.

The proposed EDS fusion method is compared with 1) the Ambiguity threshold ($\epsilon = 0.15$) Dempster-Shafer (ADS) [7], 2) the Supervised Dempster-Shafer (SDS) [12], 3) the clustering of the stacked sources (STAC) and 4) other standard fusion methods including sum (SUM), product (PROD) or entropy-weighted product of fuzzy memberships (ENT). Tables 1 and 2 summarize the overall accuracy (OA), Cohen's Kappa ($\kappa$) and per class accuracies for the two datasets averaged over a three-fold validation, one-fold for validation of SDS and the remaining two-folds for testing the final classification. Figures 1 and 2 report the standard deviations for the two datasets. Classification maps before and after fusion are shown in Figs. 3 and 4.

In the *Geneva* case study, the classification results on Landsat and SPOT images show few complementarity between class accuracies. The spatial difference (30m and 20m of spatial resolutions respectively) does not bring much complementarity. The standard fusion methods show the same values of accuracy than the unsupervised DS fusions, with the

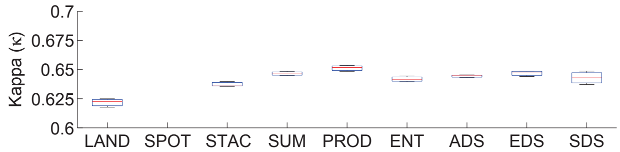



**Fig. 1**. $\kappa$ results (3-fold cross-validation) for *Geneva*.

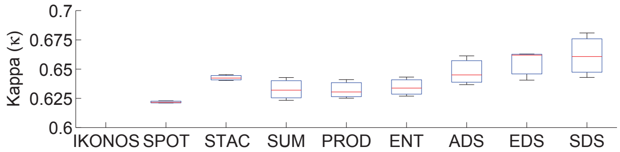



**Fig. 2**. $\kappa$ results (3-fold cross-validation) for *Tana*.

standard product fusion giving slightly better accuracies. The lack of complementarity between the source classes could not allow the unsupervised DS entropy-based fusion to produce better results but still achieves better $\kappa$ than single source clusterings and stacked sources clustering.

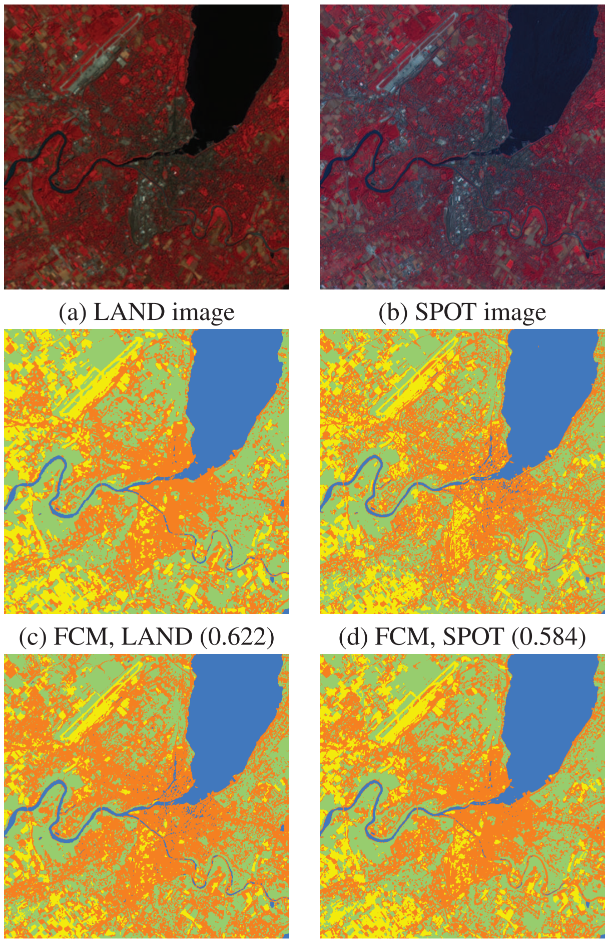
In *Tana*, complementarity is observed: the classification result on the Ikonos image is much lower than on the SPOT image. The Ikonos image has wavy water surfaces which makes the clustering harder for this source and lead to poor results. In both images the accuracy obtained for the class "Farming" is very low, inherently to its diversity (contains spectral signatures corresponding to different types of crops).

The standard fusion methods are simply finding a trade-off between the two clustering and result in slightly lower accuracies compared to the clustering on the stacked spectral bands.

The proposed EDS outperforms all the other fusion methods and reaches the accuracy of the supervised fusion (SDS). This shows the great potential with Very-High Resolution (VHR) images. We remember that contrarily to the ambiguity threshold approach (ADS) [7], EDS does not need any pre-defined parameters. The Supervised DS fusion method exagerates the tendencies, i.e SDS pushes up the classes having a high value of accuracy in one of the images, leaving aside some other classes with low accuracies. Whereas the methods ADS and EDS give more homogeneous accuracies through the classes. This difference is mostly due to the global assignment of evidences (per class) for SDS, in contrary to the local assignment of evidences (per pixel) of the proposed method. EDS allows to handle very complementary sources and exploit their respective ambiguities.

## 6. CONCLUSION

We propose an approach for the fusion of clustering from multiples sources, based on an unsupervised assignment of Dempster-Shafer evidences, generalizing the treatment of ambiguity. Instead of a user-defined ambiguity threshold [7] to define if a sample is highly ambiguous or not, our method defines an ambiguity factor based on the local entropy among the cluster memberships. Ambiguous pixels, representing the mixture of classes, are given stronger evidences representing the union of classes. This increases the possibility of unveiling the ambiguity among similar classes thanks to the other source. Our Entropy-based Dempster-Shafer fusion method (EDS) performs better than standard fusion methods and competes with supervised Dempster-Shafer fusion schemes. This confirms the benefits of our fully unsupervised method, recalling that no user-defined parameters are set. It shows particularly better performance with Very-High Resolution images where sources present more diversity and complementarity among the classes. Further perspectives are on developing more grounded evidence assignments and exploring other types of clustering [16] and the application of such fusion to different modalities such as Synthetic Aperture Radar (SAR) images or hyperspectral images.



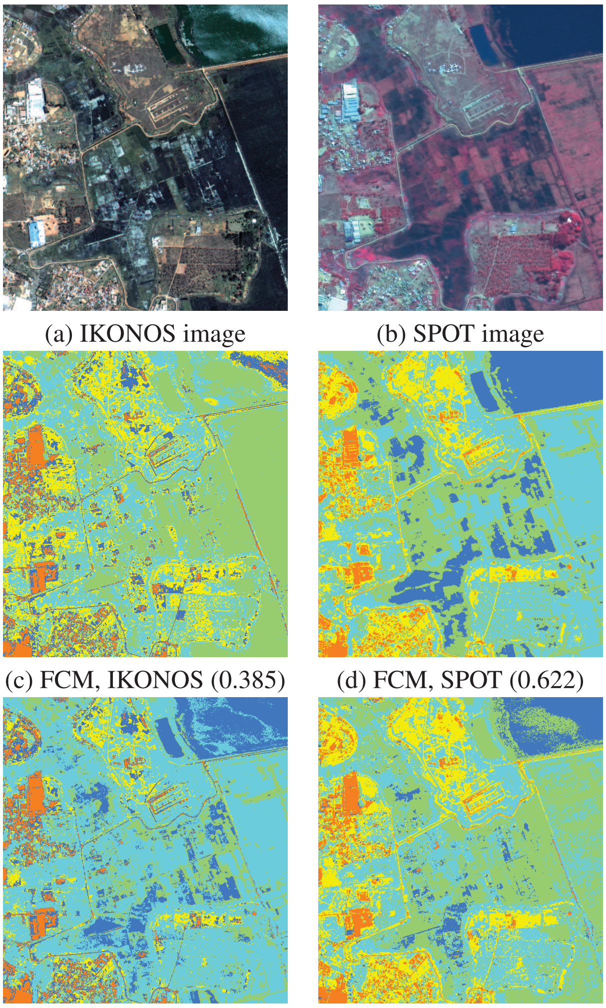
(a) LAND image (b) SPOT image

(c) FCM, LAND (0.622) (d) FCM, SPOT (0.584)

(e) SDS (0.644) (f) Proposed EDS (0.647)

ground truth: water vegetation farming urban ricefield

**Fig. 3**. Images and classification maps for the *Geneva* dataset



(a) IKONOS image (b) SPOT image

(c) FCM, IKONOS (0.385) (d) FCM, SPOT (0.622)

(e) SDS (0.662) (f) Proposed EDS (0.656)

**Fig. 4**. Images and classification maps for the *Tana* dataset (a 600 × 600 subset only is shown)

## 7. REFERENCES


[1] L. Wald, "Some terms of reference in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, 1999.

[2] J. Meynet and J.-Ph. Thiran, "Information theoretic combination of pattern classifiers," *Pattern Recog.*, vol. 43, no. 10, pp. 3412–3421, 2010.

[3] M. Fauvel, J. Chanussot, and J.A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, 2006.

[4] L. Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 1018–1031, 2002.

[5] G. Shafer, *A mathematical theory of evidence*, vol. 76, Princeton university press Princeton, 1976.

[6] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," in *Uncertainty in artificial intelligence*, 1990, vol. 5, pp. 29–39.

[7] A.O. Boudraa, A. Bentabet, and F. Salzenstein, "Dempster-shafer's basic probability assignment based on fuzzy membership functions," *Electron. Lett. Comp. Vis. Im. Anal.*, vol. 4, no. 1, pp. 1–10, 2004.

[8] S. Chitroub, "Classifier combination and score level fusion: concepts and practical aspects," *Int. J. Image Data Fusion*, vol. 1, no. 2, pp. 113–135, 2010.

[9] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, no. 3, pp. 244–264, 1988.

[10] I. Bloch, "Defining belief functions using mathematical morphology - application to image fusion under imprecision," *Int. J. Approx. Reason.*, vol. 48, no. 2, pp. 437 – 465, 2008, ¡ce:title¿In Memory of Philippe Smets (19382005)¡/ce:title¿.

[11] D. Mercier, B. Quost, and T. Denux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Information Fusion*, vol. 9, no. 2, pp. 246 – 258, 2008.

[12] A.B. Aissa, N.E. El Faouzi, and E. Lefevre, "Classification multisource via la théorie des croyances," in *Workshop Fouilles de données complexes dans un processus d'extraction de connaissances, EGC*, 2004, pp. 31–44.

[13] J.C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[14] L. Bentabet, Y.M. Zhu, O. Dupuis, V. Kaftandjian, D. Babot, and M. Rombaut, "Use of fuzzy clustering for determining mass functions dempster-shafer theory," in *Proc. WCCC-ICSP*. IEEE, 2002, vol. 3, pp. 1462–1470.

[15] L. Xu, A. Krzyzak, and C.Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 418–435, 1992.

[16] F. de Morsier, D. Tuia, M. Borgeaud, V. Gass, and J.-Ph. Thiran, "Unsupervised change detection via hierarchical support vector clustering," in *IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*. IEEE, 2012, pp. 1–4.