A TWO-STEP DECISION FUSION STRATEGY: APPLICATION TO HYPERSPECTRAL AND MULTISPECTRAL IMAGES FOR URBAN CLASSIFICATION

Walid Ouerghemmi[a], Arnaud Le Bris[a], Nesrine Chehata[b], Clément Mallet[a]

[a] Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mandé, France - walid.ouerg@gmail.com, (arnaud.le-bris, clement.mallet)@ign.fr
[b] EA 4592 Géoressources & Environnement, Bordeaux-INP/Université Bordeaux Montaigne - nesrine.chehata@ensegid.fr

KEY WORDS: Decision fusion, Optimization, Graphical model, Hyperspectral, Multispectral, Urban classification

ABSTRACT:

Very high spatial resolution multispectral images and lower spatial resolution hyperspectral images are complementary sources for urban object classification. The first enables a fine delineation of objects, while the second can better discriminate classes and consider richer land cover semantics. This paper presents a decision fusion scheme taking advantage of both sources classification maps, to produce a better classification map. The proposed method aims at dealing with both semantic and spatial uncertainties and consists in two steps. First, class membership maps are merged at pixel level. Several fusion rules are considered and compared in this study. Secondly, classification is obtained from a global regularization of a graphical model, involving a fit-to-data term related to class membership measures and an image based contrast sensitive regularization term. Results are presented on three datasets. The classification accuracy is improved up to 5%, with comparison to the best single source classification accuracy.

1. INTRODUCTION

Mapping urban environments requires very high spatial resolution (VHR) optical images (<5 m). Indeed, such spatial resolution is necessary to individualize and precisely delineate urban objects and to consider sharper geometrical details (e.g. Herold et al, 2003; Cleve et al, 2008). However, high spatial resolution sensors have generally a poor spectral configuration (i.e. usually three or four bands, RGB/RGB-NIR), limiting their ability to discriminate fine classes and limiting classification performances (Thomas et al, 2003; Carleer, 2005; Yu et al, 2006) compared to superspectral or hyperspectral (HS) sensors. Unfortunately, these latter generally exhibit a lower spatial resolution. To overcome the weakness of both sensors, multispectral (MS) and HS imageries could be jointly integrated to benefit from their complementary characteristics, so as to have; 1) rich geometrical and textural details to finely delineate objects, and 2) rich spectral information to efficiently separate the classes. Thus, the fusion of such sensors should enhance the classification performance at the highest spatial resolution. However, such fusion scheme would have to cope with both **spatial** and **semantic** uncertainties.

Fusion of heterogeneous datasets (i.e. different spatial and spectral resolution), have been widely investigated in the literature (e.g. Pohl and Genderen, 1998; Schmitt and Zhu, 2016). The fusion procedure can be carried at three distinct levels:

1) **observation level**: the most popular method for that purpose is the Pan-sharpening. It consists in merging high resolution panchromatic (PAN) image with low resolution MS one to produce high resolution MS image. A review of such methods can be found in (Loncan, 2015).

2) **feature level**: (e.g. Fauvel et al, 2007; Wegner et al., 2011; Ban and Jacob, 2013); consisting in applying a single classification using features extracted from both sources. For instance, Wegner et al. (2011) proposed a conditional random field (CRF) model for building detection using InSAR and orthophoto features.

3) **decision level:** (e.g. Djafari, 2003; Fauvel et al, 2006); consisting in merging different classification maps that are calculated from heterogeneous datasets.

This paper aims at designing a generic fusion method that could be applied to other sources than HS/MS, involving poor spatial resolution but rich semantic information sensor and very high spatial MS sensor. Thus, the fusion at observation level is not conceivable. Only the fusion at decision level is investigated in this paper. The challenge here is to merge complementary single's expert classifications to enhance them, in a case where training datasets are quite poor.

Fusion methods operating at the decision level can be applied in two situations, considering either they try to merge multiple classifiers applied to a same dataset, or multiple datasets. Concerning the second group, the aim is to take advantage of the information provided by multiple data sources. Fusion methods consisting in using all probabilistic outputs of several single source classifications as a new feature set for a classifier as in (Ceamanos et al., 2010), must be avoided here because of their strong requirements in terms of training dataset to prevent overfitting. Thus, this study will rather consider decision fusion rules, such as the ones based on probabilistic, fuzzy, possibilistic fusion or evidence theory. Benediktsson and Kanellopoulos (1999) combined neuronal and statistical Maximum Likelihood classifiers using several consensus theory rules (i.e. majority voting, complete agreement, CONSNN-NN) to classify MS and HS images. A characterization of the spatial organization of SAR images elements is investigated by Tupin et al. (1999) by merging the responses of multiple low-level detectors applied to the same image. The fusion is carried by Dempster-Shafer (Shafer, 1979) evidence theory rules. Fauvel et al. (2006) investigated the use of fuzzy decision rules to combine the classification results of a conjugate gradient neural network and a fuzzy classifier over an IKONOS image. Paisitkriangkrai et al (2015) combined convolutional neural networks and random forest classifiers using a multiplication scheme, the final result was regularized by a conditional random field (CRF). Jeon (1999), investigated the fusion of multi-temporal thematic mapper images, using decision fusion based methods (i.e. Jointly likelihood and weighted majority fusion). Waske and Benediktsson (2007) applied SVM to classify SAR and MS independently, the probabilities maps were then fused using different strategies (i.e. SVM's, majority voting, and absolute maximum rule).

Hervieu et al. (2016) combined the classification results of HS and MS images, within a graphical energetic CRF model optimized using a graph-cut algorithm.

This paper proposes a fusion scheme at the decision level to merge a low resolution LR-HS and a VHR-MS images. The method aims at dealing with both semantic and spatial uncertainties, and is based on two steps; decision fusion at pixel level and classification optimization through a global regularization framework. Several decision fusion rules are tested including fuzzy, Bayesian, margin and Dempster-Shafer based rules. The fusion is optimized in as second step by a graph-cut algorithm including a contrast sensitive spatial regularization term.

2. PROPOSED METHOD

The method is based on three main steps: a) classification of a HS and a MS images and generation of the posterior probabilities, b) fusion of the posterior probabilities at the decision level, and c) classification optimization (Figure 1). First, two input images are classified. A gaussian kernel SVM classifier (Vapnik, 1999) was used here, but other supervised classifiers could be used (e.g. random forest). The posterior probabilities are retrieved with the Platt's technique (Platt, 2000). At the end, a posterior class probabilities map is generated for each classification map.
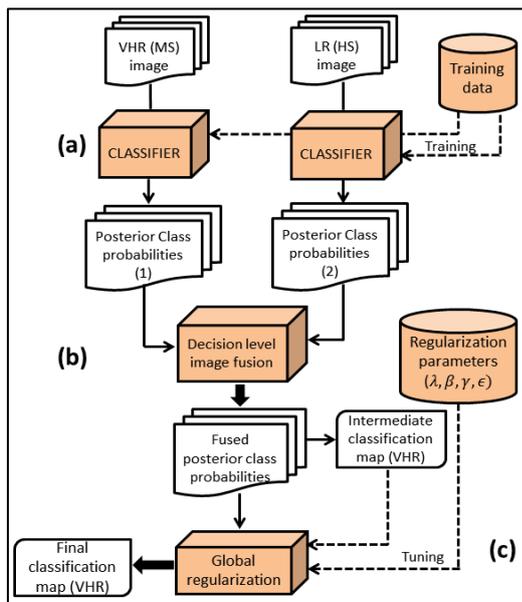
Figure 1. Two-step strategy for multi-sources data fusion.

Secondly, a decision fusion is applied to these posterior class probability maps. A variety of rules were tested. These rules could be divided into 4 classes; fuzzy decision rules (i.e. *Min*, *Max*, *Compromise, Prioritized, Accuracy dependent*), Bayesian combination (i.e. *Sum* and *Product* based rules), evidence theory (i.e. *Dempster-Shafer* rule), margin theory (i.e. *Margin-Max* rule). These decision fusion rules will be detailed hereafter. They enable to combine two class probability maps into a more accurate one at the highest resolution. The last step consists in performing a global regularization of the classification map obtained at step 2, so as to deal with spatial uncertainties between both sources. A graphical model was used, involving a fit-to-data term and a contrast sensitive regularization term. This formulation has been used successfully for many purposes related to image fusion (e.g. Kolmogorov and Zabih, 2004).

3. DECISION FUSION

The decision fusion rules used in this paper are exclusively based on the class membership probabilities generated at pixel level by the classifier. The fusion is carried out pixel by pixel, combining the class membership probabilities of each source. For this study, ten decision rules have been tested; these rules are derived from different probabilistic, possibilistic and evidential theories.

3.1. Fuzzy rules

Introduced first by Zadeh (1965), fuzzy rules become popular tools which were often used to process uncertain data (e.g. Dubois, 1992; Fauvel, 2006).

3.1.1 Theory and general properties: Let us consider a reference set $\mathcal{L}$ of classes, then a fuzzy set A in $\mathcal{L}$, is a set of ordered pairs:

$$A = [(x, P_A(x)|x \in \mathcal{L})], \tag{1}$$

where $P_A : \mathcal{L} \to [0,1]$ is called the membership probability (also called membership function) of A in $\mathcal{L}$, it ranges a nonnegative real numbers, whose supremum is finite. The membership probability of the fuzzy set is a crisp function (i.e. real valued). The intersection of two fuzzy sets P_A and P_B is given by the minimum of their membership probabilities:

$$\forall x \in \mathcal{L} \quad (P_A \cap P_B)(x) = \text{Min}(P_A(x), P_B(x)). \tag{2}$$

The union of two fuzzy sets P_A and P_B is given by the maximum of their membership probabilities:

$$\forall x \in \mathcal{L} \quad (P_A \cup P_B)(x) = \text{Max}(P_A(x), P_B(x)). \tag{3}$$

The complement of a fuzzy set P_A is given by:

$$\forall x \in \mathcal{L} \quad P_{\bar{A}}(x) = 1 - P_A(x). \tag{4}$$

3.1.2 Measure of conflict between two sources; Let us consider two sources A and B, with the corresponding membership probabilities $P_A(x)$ and $P_B(x)$, the conflict between these sources is quantified using Dubois & Prade measure (Dubois and Prade, 1994) (1- K), where:

$$K = \text{Sup}_{x} \text{Min}(P_A(x), P_B(x)). \tag{5}$$

3.1.3 Confidence measure: To reduce the influence of unreliable information within each fuzzy set, a weight w_i called pointwise accuracy as proposed by Fauvel et al. (2006). Considering a multi-image classification case, let us consider the fuzzy set $P_i(x)$, with i the number of source/classification images and x a pixel belonging to a source i. Rationally, a classifier is considered reliable if regarding a pixel x, one class has high membership and the others are low, conversely, if more than one class presents a high membership, the fuzzy set will present a high degree of fuzziness, and the classifier will be considered unreliable regarding the pixel x. Starting from this assumption, one can weight each fuzzy set by w_i to reduce the influence of unreliable information as follows:

$$w_i = \frac{\sum_{k=0, k \neq i}^{n} H_{\alpha QE}(P_k)}{(n-1) \sum_{k=0}^{n} H_{\alpha QE}(P_k)}, \tag{6}$$

where n is the number of sources, and $H_{\alpha QE}(P_k)$ the fuzziness degree of source k. $H_{\alpha QE}$ is a measure fuzziness called the α-quadratic entropy (QE) (Pal and Bezdek, 1994).

Let us consider two sources A and B with class membership probabilities P_A and P_B. The pointwise measure w_i was

integrated in the fusion rules calculation, to favor the most reliable source; each membership is multiplied by the pointwise measure as follows: $P_A(x) = w_A . \tilde{P}_A(x)$, $P_B(x) = w_B . \tilde{P}_B(x)$, where $\tilde{P}_A$, $\tilde{P}_B$ are the original membership probabilities, and w_A, w_B are respectively the corresponding pointwise measures.

3.1.4 Fuzzy rules:

Five fuzzy operators were considered for fusion

1) A conjunctive T-norm *Min* operator:

$$P_{fusion}(x) = \text{Min}(P_A(x), P_B(x)). \qquad (7)$$

2) A disjunctive T-norm *Max* operator:

$$P_{fusion}(x) = \text{Max}(P_A(x), P_B(x)). \qquad (8)$$

3) A *Compromise* operator (as proposed by Dubois and Prade, 1992) following equation (9):

$$P_{fusion}(x) = \begin{cases} \text{Max}\left(T_1, \text{Min}(T_2, (1 - K))\right) & if \ (1 - \text{K}) \neq 0 \\ \text{Max}(P_A(x), P_B(x)) & if \ (1 - \text{K}) = 1 \end{cases},$$
$$(9)$$

where $T_1 = \dfrac{\text{Min}(P_A(x), P_B(x))}{K}$, $T_2 = \text{Max}(P_A(x), P_B(x))$.

-When the conflict between A and B is low (i.e. $(1 - \text{K}) \approx 0$), the operator behavior is conjunctive.
-When the conflict between A and B is high (i.e. $(1 - \text{K}) \approx 1$), the operator behavior is disjunctive.
-When the conflict is partial (i.e. $0 < (1 - \text{K}) < 1$), the operator behaves in a compromise way.

Let us consider the *Compromise* rule (9), where T_1 is normalized to a membership belonging to [0, 1] (i.e. division by K), unlike "$\text{Min}(T_2, (1 - K))$". Indeed, T_1 will be favored at $P_{fusion}(x)$ level. To deal with this, the next alternative *Compromise* rule was proposed. Let us consider $Cbest1 = argmax_{c \in \mathcal{L}} P_{fusion}(x)$, and $Cbest2 = argmax_{c \in \mathcal{L} \setminus Cbest1} P_{fusion}(x)$, the intra-membership conflict f_c is measured as follow: $f_c = abs(Cbest1 - Cbest2)$, a threshold $t_c = 0.25$ is proposed for the decision process:

$if \ f_c < tc \ \#existence \ of \ intra - membership \ conflict \ \#$
$do \ \{P_{fusion}(x) = \text{Max}(P_A(x), P_B(x))\}$
$else \ \#no \ intra - membership \ conflict \ \#$
$do \ \{P_{fusion}(x) \ following \ equation \ (9)\}$

---

4) Prioritized operators (Dubois and Prade, 1994) *Prior 1* (Equation 10) and *Prior 2* (Equation 11) rules are considered:

$$P_{fusion}(x) = \text{Max}(P_A(x), \text{Min}(P_B(x), K)). \qquad (10)$$

$$P_{fusion}(x) = \text{Min}\left(P_A(x), \text{Max}(P_B(x), (1 - K))\right). \qquad (11)$$

For both operators, when the conflict between A and B is high (i.e. $K \approx 0$), P_A contradicts P_B, and only the information provided by P_A is taken into consideration, P_B is considered as a specific piece of information. Contextual dependent operators are more adequate to deal with conflictual situations than conjunctive and disjunctives operators (i.e. Bloch, 1996; Fauvel, 2006) which are ill suited to handle the conflict.

5) An *Accuracy Dependent* (*AD*) operator (Fauvel, 2006), integrating both local and global confidence measurements:

$$P_{fusion}^j(x) = \text{Max}\left(\text{Min}\left(w_i P_i^j(x), f_i^j(x)\right), i \in [1, n]\right), (12)$$

where f_i^j is the global confidence of source i regarding class j, P_i^j a class membership of source i, and w_i a normalization factor (see section 3.1.3). This operator ensures that only reliable sources are taken into consideration for each class, via the predefined coefficients f_i^j. The idea seems interesting, nevertheless, the final result will be dependent on the classifier reliability and also on the ground truth availability since it is necessary to generate f_i^j term.

3.2 Bayesian combination

In addition to fuzzy rules we used simpler Bayesian *Sum* and *Product* combinations of membership probabilities (e.g. Bloch, 2006). Each membership is multiplied by the pointwise measure as for (see section 3.1.3). This permits to evaluate the limit of such operators compared to more complex combinations. The fusion is carried using a Bayesian *Sum* (Equation 13) and product (Equation 14) operators as follows:

$$P_{fusion}(x) = P_A(x) + P_B(x). \qquad (13)$$

$$P_{fusion}(x) = P_A(x) . P_B(x). \qquad (14)$$

3.3 Margin-based rule (Margin-Max)

Let's consider two sources A and B, where $\mathcal{S} = \{A, B\}$, and different classes $\mathcal{L} = \{c_i\}_{i \in [1,n]}$. Let $P_s^{(c)}(x)$, be the pointwise membership probability of pixel x to a class c, according to a source s. The margin of source s in x is:

$$margin^{(s)}(x) = P_s^{(cbest1)}(x) - P_s^{(cbest2)}(x), \qquad (15)$$

where $cbest1 = argmax_{c \in \mathcal{L}} P_s^{(c)}(x)$ and $(cbest2) = argmax_{(c \in \mathcal{L}) \setminus \{cbest1\}} P_s^{(c)}(x)$.

Once the margin defined, different fusion rules could be defined based on it. For this study, a *Max-Margin* fusion method is tested. The aim is to combine the sources, conserving in each pixel the most confident source. To calculate the combined membership probabilities of two sources A and B, where $\mathcal{S} = \{A, B\}$, and different classes $\mathcal{L} = \{c_i\}_{i \in [1,n]}$.
$\forall x, \forall c \in \mathcal{L}$,

$$P_{fusion}^{(c)}(x) = P_{S_{best}}^{(c)}(x), \qquad (16)$$

where $S_{best} = argmax_{s \in c} margin^{(s)}(x)$.

3.4 Dempster-shafer (DS) evidence theory based rule

In *DS* formalism, the information from a source s for a class c is represented by a mass function $m_c | m_c \in [0, 1]$ (Shafer, 1976). The evidence theory implies the use of simple classes $c \in \mathcal{L}$ as well as compound classes (i.e. $c' = c_k \cup .... \cup c_n$). Composed classes were here limited to the union of -at most- 2 simple classes (i.e. $c' = c_k \cup c_n$), let's note $\mathcal{L}'$ the new classes ensemble. Masses are associated to each class as:

$-m(\emptyset) = 0$
-Simple classes: $\forall c \in \mathcal{L}$, $\forall pixel \ x$, and $\forall s \in \mathcal{S}$, $m_s^{(c)}(x) = P_s^{(c)}(x)$, where m is the mass affected to class c by source s, and P is a pointwise membership probability of the considered class.
-Compound classes:
The compound classes masses were here generated as follows: $\forall c1, c2 \in \mathcal{L}$, $\forall$ pixel x, and $\forall s \in \mathcal{S}$

$$m_s^{(c1 \cup c2)}(x) = \left(P_s^{(c1)}(x) + P_s^{(c2)}(x) \right) \times \quad (17)$$

$$\left(1 - \text{Max}\left(P_s^{(c1)}(x), P_s^{(c2)}(x) \right) \right) + \text{Min}\left(P_s^{(c1)}(x), P_s^{(c2)}(x) \right).$$

The masses are then normalized as: $\sum_{c \in \mathcal{L}'}^{(s)} m_c^{(s)}(x) = 1$. DS conflict measure between two sources A and B, is computed as follow:

$$K(x) = \sum_{\substack{c,d \in \mathcal{L}' \\ c \cap d = \emptyset}}^{(s)} m_A^{(c)}(x)\, m_B^{(d)}(x), \quad (18)$$

where $c, d \in \mathcal{L}'$, are compound classes with $c \cap d = \emptyset$.
At the end, probability masses are merged as follows:

$$m_{fusion}^{(c)}(x) = \frac{1}{1 - k(x)} \sum_{c1,c2 \in \mathcal{L}}^{(s)} m_A^{(c1)}(x)\, m_B^{(c2)}(x), \quad (19)$$

3.5 Global regularization

This section is dedicated to the global regularization model used as a post-processing step to enhance the classification fusion performance. The problem is expressed using an energetic graphical model and solved as a min-cut problem. The global energy is minimized using a graph-cut method named quadratic pseudo-boolean optimization method (QPBO) (Boykov and Kolmogorov, 2004) associated with, an α-expansion routine to deal with the multi-class problem.

3.5.1   Model definition: The energy term is composed of a data term E_{data} and a regularization term E_{reg}, the model from Hervieu et al. (2016) was adapted to deal only with classification rectification instead of fusion. It uses a graphical model, where the energy model is a probabilistic function of the posterior probability P_{fusion}. For a classification map C, the energy term is written following Equation (20):

$$E(P_{fusion}, C_{fusion}, C) = \sum_{x \in I_{MS}} E_{data}(C(x)) + \lambda \sum_{\substack{x,y \in N \\ x \neq y}} E_{reg}(C(x), C(y)), \quad (20)$$

where :
$E_{data}(C(x)) = f\left(P_{fusion}(C(x)) \right),$
$E_{reg}(C(x) = C(y)) = g(P_{fusion}(C(x)), C_{fusion}),$
$E_{reg}(C(x) \neq C(y)) = h(P_{fusion}(C(x)), C_{fusion}),$

$\lambda \in [0, \infty[$ is a tradeoff parameter between data and regularization terms, N is the 8 connexity neighbors.

E_{data} is a fit-to-data attachment term, function of the probability map P_{fusion} which models the result of the classification fusion, and defined by the function f, as follows:

$$f(t) = -\log(t) \text{ with } t \in [0, 1]. \quad (21)$$

The function f ensures that if the probability for a pixel x to belong to class $C(x)$ is close to 1, E_{data} will be small, and will not impact the total energy E. Conversely, if the probability for a pixel x to belong to class $C(x)$ is low, E_{data} will be near its maximum, and will penalize such configuration. E_{reg} is a regularization term defining the interactions between a pixel x and its 8 neighbors, and enabling to smooth the initial classification map C_{fusion}, by favoring the neighboring pixels to belong to the same class (i.e. minimizing energy). A Potts model (Schindler, 2012) is a typical configuration defined as follows:

$$E_{reg}(C(x) = C(y)) = 0, \quad (22)$$
$$E_{reg}(C(x) \neq C(y)) = 1.$$

In this study, a slightly enhanced Potts model is used. The model integrates contrast information (Hervieu et al, 2016) that is included in the MS image I_{MS} and verifying:

$$E_{reg}(C(x) = C(y)) = 0, \quad (23)$$
$$E_{reg}(C(x) \neq C(y)) =$$
$$(1 - \gamma)\left(1 - P_{fusion}\left(C_{fusion}(x) \right)^{\beta} \right) + \gamma\, V(x, y, \epsilon),$$

Where, $\beta \in [0, \infty[$ is a tradeoff parameter between the smoothing criterion and the importance of C_{fusion} in the model, V is a contrast measure, γ is a tradeoff between the basic model led by the decision fusion classification C_{fusion}, and the integrated contrast term $V(x, y, \epsilon)$, and ϵ a parameter modifying the standard deviation in the exponential term. The contrast term V (Rother et al, 2004) is calculated as follows:

$$V(x, y, \epsilon) = \frac{1}{n} \sum_{i \in [0, dim]} V_i(x, y)^{\epsilon} \text{ with } \epsilon \in [0, \infty], \quad (24)$$

where $V_i(x, y) = \exp\left(\frac{-(I_i(x) - I_i(y))^2}{2(I_i(x) - I_i(y))^2} \right)$, n the dimension of image I_{MS}, I_i the intensity for pixel x in the MS image.

The re-writing of the Potts model regarding the regularization term handles more efficiently the smoothing procedure. Indeed, when $C(x) \neq C(y)$, $E_{reg}(C(x) \neq C(y))$ becomes a function of P_{fusion} and V. If $P_{fusion}\left(C_{fusion}(x) \right)$ is close to 1, decision fusion gives a high confidence to pixel x to belong to class C_{fusion}, then E_{reg} becomes only dependent of V, which will decide whether the configuration C_{fusion} is favored or not. Conversely, if $P_{fusion}\left(C_{fusion}(x) \right)$ is close to zero, E_{reg} is high, and the configuration C_{fusion} is prone to be rejected. A Potts model is obtained if $\gamma = 0$ and $\beta = 0$.

3.5.2   Parameters setup: Given the energy term E (see Equation 20), 4 parameters controlling the regularization degree are used; λ, γ, β, and ϵ. Each of these term is attached to a particular sub-terms of E. $\lambda \in [0, \infty[$ is a tradeoff parameter between the terms E_{data} and E_{reg}. The more λ increases the more is the regularization effect, the choice of this parameter will depend on the distribution of the decision fusion map to be optimized. $\gamma \in [0,1]$ is a tradeoff parameter between the basic energy model and the rectified model integrating the contrast measure. Last, $\epsilon \in [0, \infty[$ is a parameter controlling the influence of the contrast measure in the energy term. A Potts model is obtained using the following parametrization:

$$\gamma = 0 \text{ and } \beta \to +\infty \text{ (or) } \gamma = 1 \text{ and } \beta \to +\infty \text{ and } \epsilon = 0.$$

The parameters were fixed otherwise to a Potts configuration which over-smooths the decision fusion classification: $\gamma = 0.5, \beta = 1$, and $\epsilon = 1$. For λ, two configurations were tuned: $\lambda = 0.1$ for the *Min* and *Dempster-Shafer* rules, and $\lambda = 10$ for the *Compromise* rule. This configuration was a good tradeoff and gave the results presented further in this paper.

4.   EXPERIMENTS AND RESULTS

4.1   Data

Three datasets were used over the cities of Pavia (Italy) and Toulouse (France) (Figures 2.a-c). For all the datasets, a SVM classifier was trained using 50 samples per class extracted from the images. Concerning Pavia city, two datasets called "Pavia University" and "Pavia Center" were used; these datasets have respectively 103 and 102 spectral

bands from 430 to 860 nm. Pavia University is a 335 x 605 image, Pavia Center is a 715 x 1096 image, and both have a ground sample distance (GSD) of 1.3 m. Both Pavia images are composed of 9 land cover classes (Figures 2.e, 2.f): *Asphalt, Meadows, Gravel ,Trees, Painted Metal Sheets, Bare Soil, Bitumen, Self-Blocking Bricks, Shadows* for Pavia University and *Water, Trees, Meadows, Self-Blocking Bricks, Bare soil, Asphalt, Bitumen roofing, Tiles roofing, Shadows* for Pavia Center. For both datasets, MS and HS images were generated. MS images were generated using Pleiades satellites spectral configuration (limited to three RGB bands), with a GSD of 1.3 m, while HS images were resampled at a lower spatial resolution of 7.8 m and at the full original spectral range (i.e. 103 and 102 bands). To sum up, both datasets include a MS image (RGB with 1.3 m GSD), and a

HS image (Full spectral range with 7.8 m GSD).

The last dataset called "Toulouse Center", is a scene over Toulouse city in France. It has 405 spectral bands ranging from 400 to 2500 nm, and an initial GSD of 1.6m. Its associated land cover is composed of 15 classes (Figure 2.d): *Slate roofing, Asphalt , Cement, Water, Pavements, Bare soil, Gravel roofing, Metal roofing1, Metal roofing 2,Tiles roofing, Grass, Trees, Railway tracks, Rubber roofing, Shadows*. MS and HS images were created for the fusion purpose; a MS image using Pleiades satellite spectral configuration (four RVB - NIR bands), with a GSD of 1.6 m, and a HS image which is a resampled version of the original image at a lower spatial resolution of 8 m. To sum up, Toulouse dataset has a MS image (RGB with 1.6 m GSD), and a HS image (Full spectral range with 8 m GSD).
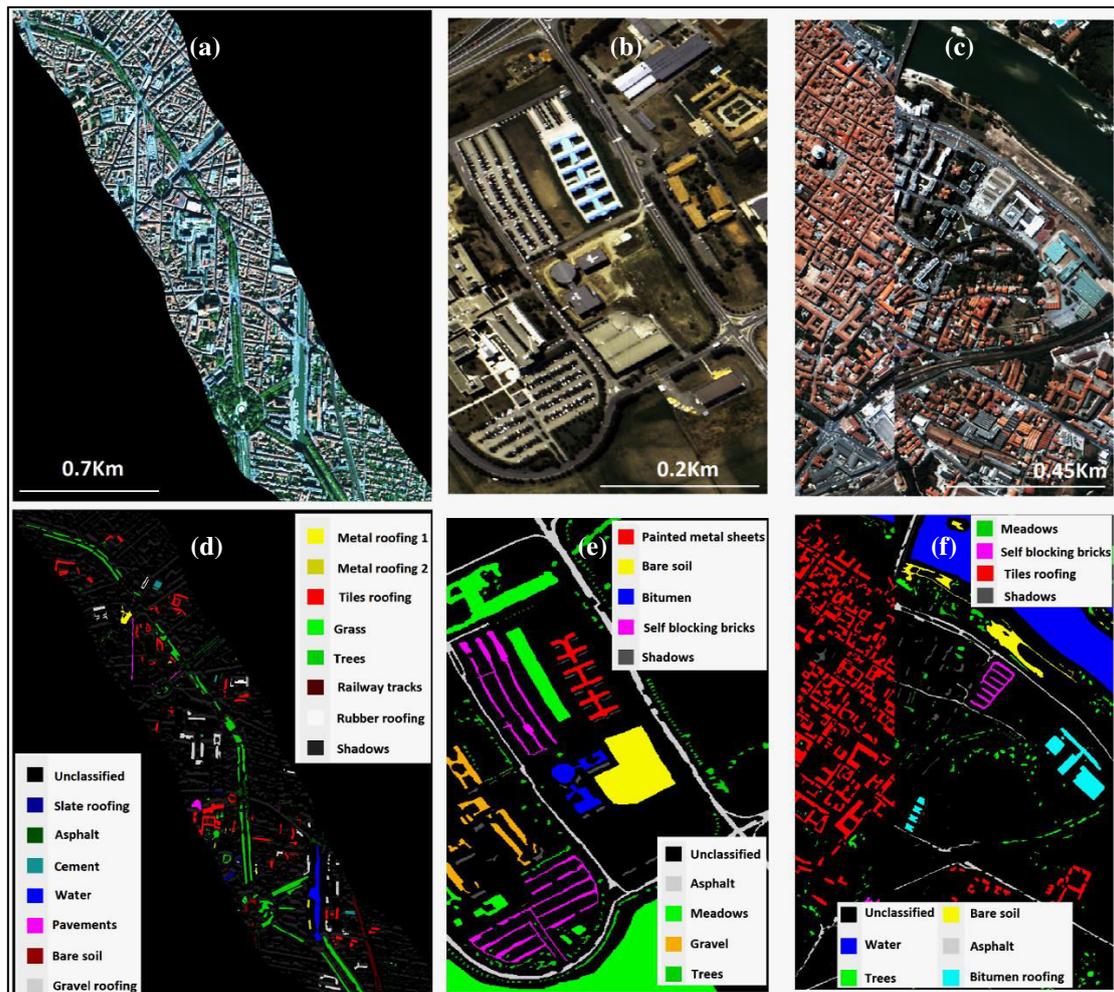
Figure 2. RGB Datasets and corresponding ground truth; (a),(d) Toulouse center, (b),(e) Pavia University, (c),(f) Pavia Center.

4.2 Results and discussion

4.2.1. Sources comparison: MS image is characterized by a high spatial resolution and a few bands, while the HS one has a low spatial resolution and a hundred(s) of bands. The SVM classifier was applied over these images leading to 1) a sharp objects delineation in the MS image due to its good spatial resolution, but also a lot of artefacts (Figures 3.b, 4.b, 5.b), and 2) a good discrimination of the different classes in the HS image but also a blurry objects delineation due to its low spatial resolution (Figures 3.a, 4.a, 5.a). The corresponding classification accuracies are listed in Table 2: the accuracies are better using HS image.

4.2.2. Decision fusion classification: 10 different decision fusion rules were first tested and compared over the 3 datasets, the quantitative results of Table 1 let us consider that the *Compromise*, *Bayesian Product*, *Margin-Max* and *Dempster-Shafer* rules are the most efficient. The comparison must also take into consideration the visual inspection of the results, as ground truth data remains very limited on these datasets. For Pavia University, four of the best accuracies were reached for *Min*, *compromise*, *Bayesian Product*, and *Dempster-Shafer* rules. In practice, the *Min/Compromise* rules, gives the best classification rendering, especially regarding class *Self-Blocking Bricks* which is a conflictual class (Figures 3.a, 3.b, Magenta color class), the two other rules seem to overestimate this class, and are more

considering the HS classification map in the fusion process, this explains their better accuracy (Table 1). The *Min* rule acts in cautious way when taking the best of the lowest memberships, the *Compromise* rule acts depending on the degree of conflict between sources. The *Bayesian Product* rule is a good and simple tradeoff if the initial classification maps are not highly conflictual, otherwise, the result will be degraded by wrong information. Concerning Pavia Center, all the rules seem accurate (Figure 4.c, example: *Dempster-Shafer* rule), with an overall accuracy superior to 98% (Table 1). When inspecting the classification images, all rules gave similar good results excepting *Prior 1*, showing a result guided by the HS classification map rather than the MS one.

Toulouse dataset, is the largest one, with up to 15 classes. Thus, the accuracies are lower. The best results were given by the *Max, Prior 2, Bayesian Sum* and *Dempster-Shafer* rules. In practice the *Max, Prior* and *Sum* rules seem to overestimate certain classes, especially tile roofing's and vegetation, the best rendering in terms of classification fusion, is given by the *Min, compromise* and *Dempster-Shafer* rules. Despite a reasonable global accuracy, the *AD* rule presents many misclassifications regarding *tiles roofing* (i.e. underestimation), *metal roofing 1* (i.e. overestimation), and a bad detection of the gravel roofing, this is mainly due to the global accuracy measure which is included in the rule and calculated thanks to the ground truth data.

As a conclusion, the quantitative accuracies doesn't necessarily transcribe the real potential of the fusion rules,

and the best ones from a quantitative and practical qualitative point of view are; the *Compromise*, the *Bayesian Product* and the *Dempster-Shafer* rules.

4.2.3 Global classification regularization: Once the decision fusion done, a global regularization procedure was applied to enhance the classification results and eliminate the artefacts. Table 2 presents the optimization results for the best rules per dataset, the optimization procedure permits to enhance further the classification. Quantitatively, global regularization process permits to slightly enhance the decision fusion classification (by 1-2%) but offers a better visual rendering with an elimination of the artefacts, a better decimation of the classes borders, and a regularization of the scattered pixels (Figures 3.d, 4.d, 5.d). These optimized maps seem better modeling the real scene. The optimization effect is more visible over Pavia University and Toulouse Center, concerning Pavia Center, the decision fusion gives already good results and thus, the optimized maps are only slightly improved (Table 2). Quantitative accuracy results obtained over the Pavia datasets are comparable to other studies (e.g. Fauvel et al, 2007). For Pavia University; the painted metal sheets are better recovered and no mismatches with the surrounding road are noticeable. The proposed method permits to extract some bitumen buildings that were difficult to differentiate from roads (i.e. up right and down right, Figure 2.b), the gravel buildings could be better refined. For Pavia center, the global rendering is enhanced with a minimization of the classification artefacts.
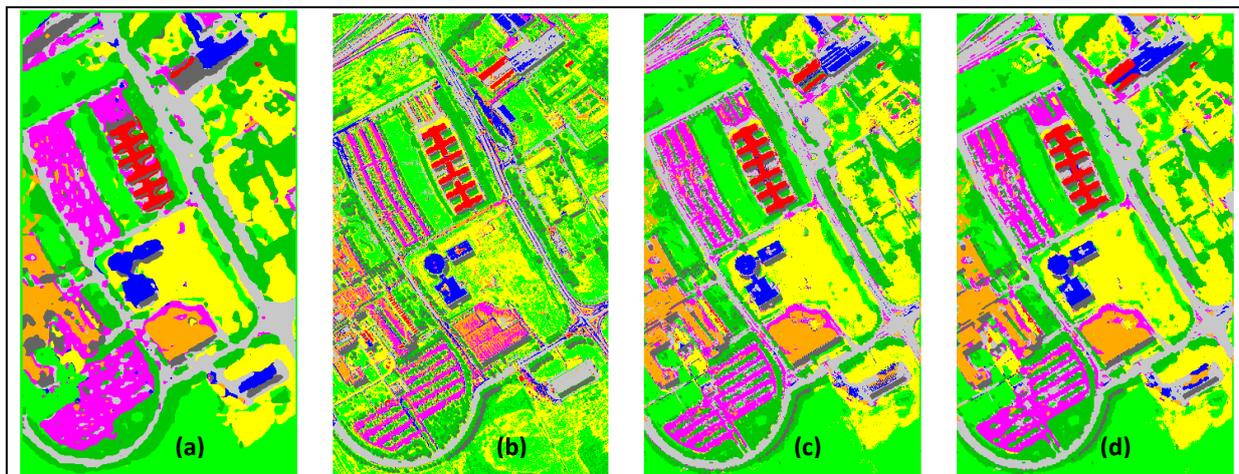
Figure 3. Pavia University classification results using the best decision fusion rule; (a) SVM classification of HS image, (b) SVM classification of MS image, (c) Classification fusion by *Min* rule, (d) Global classification regularization.
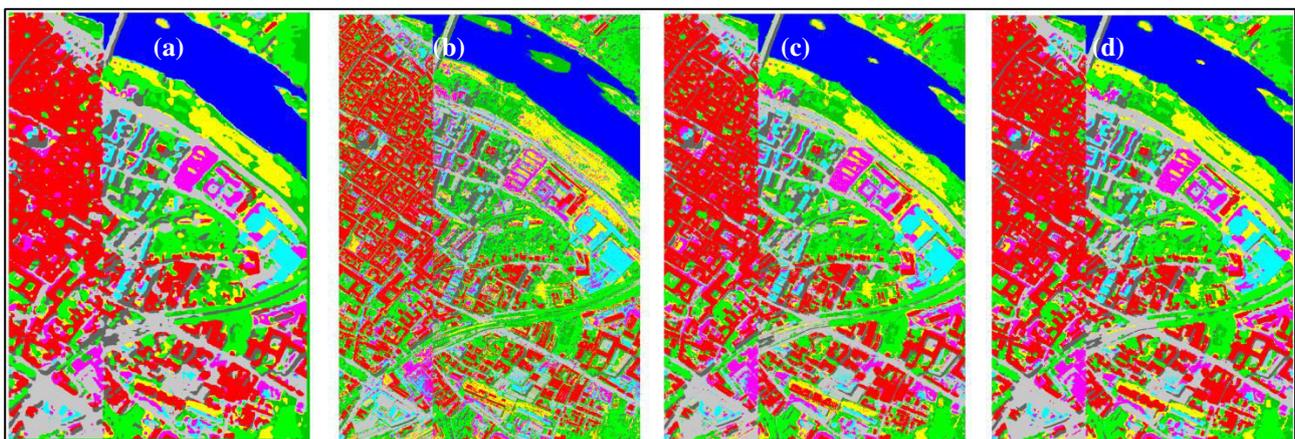
Figure 4. Pavia Center classification results using the best decision fusion rule; (a) SVM classification of HS image, (b) SVM classification of MS image, (c) Classification fusion by *Dempster-Shafer* rule, (d) Global classification regularization.
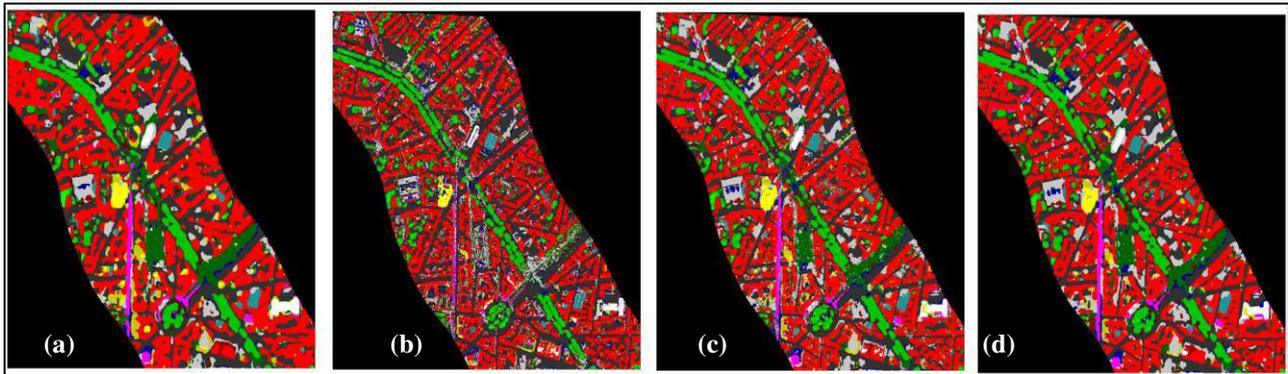
Figure 5. Toulouse Center classification results using the best decision fusion rule; (a) SVM classification of HS image, (b) SVM classification of MS image, (c) Classification fusion by *Compromise* rule, (d) Global classification regularization.

Data	Accuracy	Max	Min	Compromise	Prior 1	Prior 2	AD	Sum Bayes	Prod Bayes	Margin -Max	Dempster -Shafer
Pavia University	OA (%)	92.8	**96.1**	**96.1**	94.7	92.8	95.0	95.0	**96.6**	94.0	**96.4**
	Kappa (%)	90.7	**94.9**	**95.0**	93.1	90.7	93.5	93.5	**95.5**	92.2	**95.4**
	F-score (%)	90.6	**95.1**	**95.0**	93.4	90.6	93.5	93.2	**95.6**	92.0	**95.3**
Pavia Center	OA (%)	98.5	98.6	**98.8**	98.2	98.5	**99.0**	98.7	**99.0**	98.8	98.9
	Kappa (%)	97.8	98.0	**98.3**	97.5	97.8	**98.7**	98.1	**98.6**	98.3	98.5
	F-score (%)	96.0	96.3	**96.7**	95.3	96.0	**97.7**	96.5	**97.2**	96.6	97.1
Toulouse Center	OA (%)	**75.6**	72.2	73.6	71.3	**75.6**	75.8	**75.7**	74.5	**75.6**	74.6
	Kappa (%)	**62.4**	58.7	60.2	57.7	**62.4**	58.1	**62.7**	61.4	**62.5**	61.5
	F-score (%)	**69.8**	65.8	68.0	65.5	**69.8**	28.3	**70.5**	69.8	69.6	69.8

Table 1. Classification accuracy after fusion procedure, 10 fusion rules at decision level are compared.

Data	Accuracy	Image HS classification	Image MS classification	Decision fusion	After regularization
Pavia University (**Min rule**)	OA (%)	94.7	68.8	96.1	97.0
	Kappa (%)	93.1	61.6	94.9	96.1
	F-score (%)	93.4	72.8	95.1	96.3
Pavia Center (**Dempster-shafer rule**)	OA (%)	98.2	92.0	98.9	99.3
	Kappa (%)	97.5	89.0	98.5	99.0
	F-score (%)	95.3	83.5	97.1	98.0
Toulouse Center (**Compromise rule**)	OA (%)	71.2	69.2	73.5	74.6
	Kappa (%)	57.6	53.8	60.2	61.5
	F-score (%)	65.4	55.9	68.0	70.9

Table 2. Classification accuracy of images HS and MS separately, after decision fusion, and after global regularization.

5. CONCLUSION

This paper proposes a two-step method dealing with multi-source data fusion and global regularization. Several decision fusion methods were tested and compared, including *fuzzy* rules, Bayesian combinations, *margin-based* rule, and *Dempster-Shafer* based rule. Among the *fuzzy* rules, the *Min* and *Compromise* rules are the most efficient. The *Max* rule is often affected by misclassifications due to the fact it pays more confidence to the highest membership. The *prioritized* rules favor a source rather than the other, indeed, the reliability is not ensured as noticed for *Prior 1*, which gives confidence to the less reliable source. *AD* rule accuracy is dependent on the ground truth reliability; the rule gives encouraging results for Pavia datasets, the accuracy was not sufficient for Toulouse dataset. The Bayesian *Sum* and *Product* rules could be interesting in case of low conflict between sources, since they give acceptable results over Pavia Center and Toulouse. Concerning the proposed *margin-based* rule, it performs well over Pavia center, and correctly over Toulouse, but it is not enough sufficient over Pavia University. Finally the *Dempster-Shafer* rule gives homogeneous performance over the three datasets leading to interesting results.

Even if the decision fusion enables to increase the classification accuracy compared to the initial classification maps, the results are affected by classification artefacts, and unclear borders, the final maps are either guided by one of the initial maps or by both, the final result is, therefore, a better version of the initial maps. The final step is a global regularization optimization of the decision fusion results, for classification enhancement. The principle is based on regularization regarding each pixel membership and its neighborhood, and regarding an image contrast measure when comparing neighboring pixels. The optimization procedure gives encouraging results, with clear borders among the different classes, and artefacts elimination.

The method also has the possibility to integrate other decision rules in a fully tunable way, the optimization model is simple and flexible and could be tuned depending on the used dataset. Further work will investigate the explicit use of conflict measures from fusion step within the regularization framework. At the moment the optimization parameters selection is rather manual, some automation could be included, and other contrast measures could be tested to improve the accuracy. The proposed fusion scheme is quite generic and could also be applied to other similar configurations, such as low spatial resolution times series and VHR-MS data. For this study, the HS image was generated from the VHR-MS one leading thus to quite optimistic

results. Real cases will be tested, including temporal differences between two images.

ACKNOWLEDGMENTS

REFERENCES

Ban, Y., and Jacob, A., 2013. Object-Based Fusion of Multitemporal Multiangle ENVISAT ASAR and HJ-1B Multispectral Data for Urban Land-Cover Mapping. In: *IEEE Trans. on Geoscience and Remote Sensing*, 51(4), 1998-2006.

Benediktsson, J.A. and Kanellopoulos, I., 1999. Decision fusion methods in classification of multisource and hyperdimensional data. *IEEE Trans. on Geoscience and Remote Sensing*, 37(3), 1367-1377.

Boykov, Y. and Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(9), 1124-1137.

Carleer, A. P., Debeir, O. and Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogram. Eng. Rem. Sen.*, 71(11), 1285-1294.

Ceamanos, X., Waske, B., Benediktsson, J.A., Chanussot, J., Fauvel, M. and Sveinsson, J.R., 2010. A classifier ensemble based on fusion of Support Vector Machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4), 293–307.

Cleve, C., Kelly, M., Kearns, F.R., Moritz, M., Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Computers, Environment and Urban Systems*, 32(4), 317-326.

Dubois, D., and Prade, H., 1992. Combination of fuzzy information in the framework of possibility theory. In: *Data Fusion in Robotics and Machine Intelligence (M.A. Abidi, R.C. Gonzalez, eds.)*, 481-505. Academic Press, New York.

Dubois, D., and Prade, H., 1994. Possibility theory and data fusion in poorly informed environments. *Control Engineering Practice*, 2 (5), 811-823.

Fauvel, M., Chanussot, J. and Benediktsson, J.A., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Trans. on Geoscience and Remote Sensing*, 44(10), 2828-2838.

Fauvel, M., Benediktsson, J., Sveinsson, J., and Chanussot, J., 2007. Spectral and spatial classification of hyperspectral data using SVMS and morphological profiles. In: *Proc. IGARSS*, pp. 4834–4837.

Herold, M., Liu, X.H., and Clarke, K.C., 2003. Spatial metrics and image texture for mapping urban land-use. *Photogrammetric Engineering & Remote Sensing*, 69(9), 991-1001.

Hervieu, A., Le Bris, A., and Mallet, C., 2016. Fusion of hyperspectral and VHR multispectral image classifications. In: urban areas, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. III–3, 457-464.

Jeon, B., 1999. Decision fusion approach for multitemporal classification. *IEEE Trans. on Geoscience and Remote Sensing*, 37 (3), 1227-1233.

Kolmogorov V. and Zabih, R., 2004. What energy functions can be minimized via Graph Cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2), 65–81.

Loncan, L., Almeida, L.B., Bioucas-Dias, J.M., Liao, W., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G., Simoes, M., Tourneret, J.-Y., Veganzones, M., Vivone, G., Wei, Q., and Yokoya, N., 2015. Hyperspectral pansharpening: a review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 27-46.

Mohammad-Djafari, A., 2003. A Bayesian approach for data and image fusion. *American Institute of Physics Conference Proceedings*, 659, 386-408.

Paisitkriangkrai, S., Sherrah, J., Janney P., and Van-Den Hengel, A., 2015. Effective semantic pixel labelling with convolutional networks and Conditional Random Fields. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, 36-43

Pal, N.R., and Bezdek, J., C., 1994. Measuring fuzzy uncertainty. *IEEE Trans. Fuzzy Syst.*, 107-118.

Pohl, C., Van Genderen, J.L., 1998. Multi-sensor image fusion in remote sensing: Concepts methods and applications", *Int. J. Remote Sens.*, 19(5), 823-854.

Rother, C., Kolmogorov, V. and Blake, A., 2004. "Grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics*, 23(3), 309-314.

Schmitt, M., and Zhu, X.X., 2016. Data Fusion and Remote Sensing: An ever-growing relationship, In: *IEEE Geoscience and Remote Sensing Magazine*, 4(4), 6-23.

Shafer, G., 1976. A Mathematical Theory of Evidence. *Princeton University Press*.

Thomas, N., Hendrix, C., and Congalton, R.G., 2003. A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering & Remote Sensing*, 69(9), 963-972.

Tupin, F., Bloch, I., and Maitre, H., 1999. A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors. *IEEE Trans. on Geoscience and Remote Sensing*, 37(3), 1327-343.

Vapnik, V. N., 1999. An overview of statistical learning theory. *IEEE Trans. of Neural Networks*, 10, pp. 988-999.

Waske, B. and Benediktsson, J.A., 2007. Fusion of support vector machines for classification of multisensor data. *IEEE Trans. on Geoscience and Remote Sensing*, 45, 3858–3866.

Wegner, J. D., Hänsch, R., Thiele, A., and Soergel, U., 2011. Building detection from one orthophoto and high-resolution InSAR data using conditional random fields. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, 83-91.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., and Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high resolution remote sensing imagery. *Photogrammetric Engineering & Remote Sensing*, 72(7), 799-811.

Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8(3), 338–353.