Decision Fusion for the Classification of Urban Remote Sensing Images

Mathieu Fauvel, *Student Member, IEEE,* Jocelyn Chanussot, *Senior Member, IEEE,* and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—The classification of very high resolution remote sensing images from urban areas is addressed by considering the fusion of multiple classifiers which provide redundant or complementary results. The proposed fusion approach is in two steps. In a first step, data are processed by each classifier separately, and the algorithms provide for each pixel membership degrees for the considered classes. Then, in a second step, a fuzzy decision rule is used to aggregate the results provided by the algorithms according to the classifiers' capabilities. In this paper, a general framework for combining information from several individual classifiers in multiclass classification is proposed. It is based on the definition of two measures of accuracy. The first one is a pointwise measure which estimates for each pixel the reliability of the information provided by each classifier. By modeling the output of a classifier as a fuzzy set, this pointwise reliability is defined as the degree of uncertainty of the fuzzy set. The second measure estimates the global accuracy of each classifier. It is defined *a priori* by the user. Finally, the results are aggregated with an adaptive fuzzy operator ruled by these two accuracy measures. The method is tested and validated with two classifiers on IKONOS images from urban areas. The proposed method improves the classification results when compared with the separate use of the different classifiers. The approach is also compared with several other fuzzy fusion schemes.**

***Index Terms*—Classification, data fusion, decision fusion, fuzzy logic, fuzzy set theory, remote sensing.**

I. Introduction

VERY HIGH resolution commercial satellite images from urban areas are now available. Information provided by these images is both spectral and spatial. Several spectral bands are currently available with very high spatial resolution, and by using these data, it is possible to identify small structures such as small houses or roundabouts in dense urban areas. To integrate these data in urban development planning, emergency response, or Earth survey, structures present in the scene should be classified, and there is a strong need for automated or semiautomated classification algorithms.

Manuscript received September 30, 2005; revised December 14, 2005. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne Program of the French and Icelandic governments (PAI EGIDE).

M. Fauvel is with the Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble, 38402 Saint Martin d'Hères, France, and also with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: mathieu.fauvel@lis.inpg.fr).

J. Chanussot is with the Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble, 38402 Saint Martin d'Hères, France (e-mail: jocelyn.chanussot@lis.inpg.fr).

J. A. Benediktsson is with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Digital Object Identifier 10.1109/TGRS.2006.876708

Several urban classification methods have been proposed in the remote sensing literature. These methods are usually based on a feature extraction step followed by a classification algorithm where feature extraction can be, e.g., band selection or band combination in the multispectral case [1]. In the case of panchromatic images, the morphological profile has been used in [2] to extract information about the size and the relative contrast of the structures, thus providing a multidimensional feature vector. In [3], textural information was used as a feature for the classification process. Classification algorithms, based on a statistical approach such as maximum likelihood [1] or a neural network have been frequently used [2], [4], [5]. In [6] and [7], possibilistic models and fuzzy logic were used to design a fuzzy classifier.

All these methods have their own characteristics and advantages; none of them is strictly outperforming all the others. The neural network approach has the advantage that no prior information about the distribution of the input data is needed. However, if an accurate multivariate statistical model can be determined, statistical methods should provide better classification accuracies than neural networks. Classifiers based on possibilistic models do not need any training, and class definitions can be done with linguistic variables [8]. Furthermore, the computation time is usually shorter with statistical approaches than neural methods.

Usually, for a given data set, performance in terms of "global" and "by class" classification accuracies depends on the considered classes, i.e., on their spectral and spatial characteristics. For instance, methods based on morphological filtering are well suited to classify structures with a typical spatial shape, like manmade constructions. On the contrary, algorithms based on spectral information (such as statistical approaches, Gaussian mixture models, or neural networks [1]) perform better for the classification of vegetation and soils. Therefore, we propose to use several approaches and try to take advantage of the strengths of each algorithm. This concept is called "decision fusion" [9]. Decision fusion can be defined as the process of fusing information from several individual data sources after each data source has undergone a preliminary classification. For instance, Benediktsson and Kanellopoulos [9] proposed a multisource classifier based on a combination of several neural/statistical classifiers. The samples are first classified by two classifiers (i.e., a neural network and a multisource classifier); every sample with agreeing results is assigned to the corresponding class. Where there is a conflict between the classifiers, a second neural network is used to classify the remaining samples. The main limitation of this method is the need of large training sets to train the different classifiers. In [10], Jeon and Landgrebe used two decision fusion rules to classify multitemporal Thematic Mapper data. Recently, Lisini *et al.* [11] proposed to combine sources according to their class accuracies. In this study, the decision fusion rule is modeled with fuzzy data fusion rules. Fuzzy-based fusion techniques have already been applied for various decision fusion schemes. For instance, Tupin *et al.* [12] combined several structure detectors to classify SAR images using Dempster–Shafer theory. Chanussot *et al.* [13] proposed several strategies to combine the output of a line detector applied to multitemporal images. Also dealing with multitemporal SAR images, Amici *et al.* [14] investigate the usefulness of fuzzy and neurofuzzy techniques to fuse the multitemporal information for the monitoring of flooded areas.

In this paper, we propose to aggregate the results of different classifiers. Conflicting situations, where the different classifiers disagree, are solved by estimating the pointwise accuracy and modeling the global reliability for each algorithm [15]. This leads to the definition of an adaptive fusion scheme ruled by these reliability measures. The proposed algorithm is based on fuzzy sets and possibility theory.

The framework of the addressed problem is modeled as follows. For a given data set, n classes are considered, and m classifiers are assumed to be available. For an individual pixel, each algorithm provides as an output a membership degree for each of the considered classes. The set of these membership values is then modeled as a fuzzy set, and the corresponding degree of fuzziness determines the pointwise reliability of the algorithm. The global accuracy is manually defined for each class after a statistical study of the results obtained with each separately used classifier. Hence, fusion is performed by aggregating the different fuzzy sets provided by the different classifiers. It is adaptively ruled by the reliability information and does not require any further training. The decision is postponed to the end of the fusion process to take advantage of each algorithm and enable more accurate results in conflicting situations.

The paper is organized as follows. Fuzzy set theory and measures of fuzziness are briefly presented in Section II-A. Section II-B presents the model for each classifier's output in terms of a fuzzy set. Then, the problem of information fusion is discussed in Section III. The proposed fusion scheme is detailed in Section IV, and experimental results are presented in Section V. Finally, conclusions are drawn.

II. Fuzzy Set Theory

Traditional mathematics assigns a membership value of 1 to elements that are members of a set, and 0 to those which are not, thus defining "crisp sets." On the contrary, "fuzzy set" theory handles the concept of partial membership to a set, with real-valued membership degrees ranging from 0 to 1. Fuzzy set theory was introduced in 1965 by Zadeh [16] as a mean to model the vagueness and ambiguity in complex systems. It is now widely used to process unprecise or uncertain data [17], [18]. In particular, it is an appropriate framework to handle the output of a given classifier for further processing. The output is usually not in a binary form and includes some ambiguity. In this section, we first recall general definitions and properties of fuzzy sets. Then, we detail the model used for the representation of the classifiers output.

A. Fuzzy Set Theory

1) Definitions:

Definition 1 (Fuzzy subset): A fuzzy subset[1] F of a reference set U is a set of ordered pairs $F = \{(x, \mu_F(x)) | x \in U\}$, where $\mu_F : U \rightarrow [0, 1]$ is the membership function of F in U.

Definition 2 (Normality): A fuzzy set is said to be "normal" if and only if $\max \mu_F(x) = 1$.

Definition 3 (Support): The support of a fuzzy set F is

$$\mathrm{Supp}(F) = \{x \in U | \mu_F(x) > 0\}.$$

Definition 4 (Core): The core of a fuzzy set is the (crisp) set containing the points with the largest membership value ($\mu_F(x) = 1$). It is empty if the set is nonnormal.

2) Logical Operations: Classical Boolean operations extend to fuzzy sets [16]. With F and G as two fuzzy sets, classical extensions are defined as follows.

1) *Union:* The union of two fuzzy sets is defined by the maximum of their membership functions, i.e.,

$$\forall x \in U, \quad (\mu_F \cup \mu_G)(x) = \max\left\{\mu_F(x), \mu_G(x)\right\}. \tag{1}$$

2) *Intersection:* The intersection of two fuzzy sets is defined by the minimum of their membership functions, i.e.,

$$\forall x \in U, \quad (\mu_F \cap \mu_G)(x) = \min\left\{\mu_F(x), \mu_G(x)\right\}. \tag{2}$$

3) *Complement:* The complement of a fuzzy set F is defined by

$$\forall x \in U, \quad \mu_{\overline{F}}(x) = 1 - \mu_F(x). \tag{3}$$

3) Measures of Fuzziness: Fuzziness is an intrinsic property of fuzzy sets. To measure how fuzzy a fuzzy set is, and thus estimate the ambiguity of the fuzzy set, several definitions have been proposed [19], [20]. Ebanks [21] proposed to define the degree of fuzziness as a function f with the following properties.

1) $\forall F \subset U$, if $f(\mu_F) = 0$, then F is a crisp set.
2) $f(\mu_F)$ is maximum if and only if $\forall x \in U$, $\mu_F(x) = 0.5$.
3) $\forall (\mu_F, \mu_G) \in U^2$, $f(\mu_F) \geq f(\mu_G)$ if

$$\forall x \in U \begin{cases} \mu_G(x) \geq \mu_F(x), & \text{if } \mu_F(x) \geq 0.5 \\ \mu_G(x) \leq \mu_F(x), & \text{if } \mu_F(x) \leq 0.5. \end{cases}$$

4) $\forall F \in U$, $f(\mu_F) = f(\mu_{\overline{F}})$. A set and its complement have the same degree of fuzziness.
5) $\forall (\mu_F, \mu_G) \in U^2$, $f(\mu_F \cup \mu_G) + f(\mu_F \cap \mu_G) = f(\mu_F) + f(\mu_G)$.

[1]For convenience, we will use the term "fuzzy set" instead of "fuzzy subset" in the following, where a fuzzy set F is described by its membership function μ_F.

Pal and Bezdek [22] proposed a measure of fuzziness based on the multiplicative class.

Definition 5 (Multiplicative class): The multiplicative class is defined as

$$H_*(\mu_F) = K \sum_{i=1}^{n} g\left(\mu_F(x_i)\right), \quad K \in R^+ \tag{4}$$

where $g(\mu_F)$ is defined as

$$\begin{cases} g(t) = \widetilde{g}(t) - \min\limits_{0 \le t \le 1} \widetilde{g}(t) \\ \widetilde{g}(t) = h(t)h(1-t) \end{cases} \tag{5}$$

and h is a concave increasing function on [0, 1], i.e.,

$$h : [0,1] \to R^2, \quad \forall x \in [0,1] \; h'(x) > 0 \text{ and } h''(x) < 0. \tag{6}$$

The multiplicative class allows the definition of various fuzziness measures, where different choices of g lead to different behaviors. For instance, let $h : [0,1] \to R^+$ be $h(t) = t^\alpha$, $0 < \alpha < 1$. The function h satisfies the required conditions for the multiplicative class, and the function

$$H_{\alpha\mathrm{QE}}(\mu_F) = \frac{1}{n 2^{-2\alpha}} \sum_{i=1}^{n} \mu_F(x_i)^\alpha \left(1 - \mu_F(x_i)\right)^\alpha \tag{7}$$

is a measure of fuzziness, namely, the α-quadratic entropy (QE). Rewriting (7) as

$$\begin{cases} H_{\alpha\mathrm{QE}}(\mu_F) = \dfrac{1}{n} \sum\limits_{i=1}^{n} S_{\alpha\mathrm{QE}}\left(\mu_F(x_i)\right) \\ S_{\alpha\mathrm{QE}}\left(\mu_F(x_i)\right) = \dfrac{\mu_F(x_i)^\alpha \left(1 - \mu_F(x_i)\right)^\alpha}{2^{-2\alpha}} \end{cases} \tag{8}$$

we can analyze the influence of parameter α (see Fig. 1): the measure becomes more and more selective as α increases from 0 to 1. With α close to 0, all the fuzzy sets have approximately the same degree of fuzziness, and the measure is not sensitive to changes in μ_F, whereas with α close to 1, the measure is highly selective, with the degree of fuzziness quickly decreasing when the fuzzy set differs from $\mu_F = 0.5$. Therefore, in this paper, we chose $\alpha = 0.5$ as a good tradeoff [22].

B. Class Representation

An n-class classification problem is considered for which m different classifiers are available. For a given pixel x, the output of classifier i is the set of numerical values, i.e.,

$$\left\{\mu_i^1(x), \mu_i^2(x), \ldots, \mu_i^j(x), \ldots, \mu_i^n(x)\right\} \tag{9}$$

where $\mu_i^j(x) \in [0,1]$ (after a normalization, if required) is the membership degree of pixel x to class j according to classifier i. The higher this value, the more likely it is that the pixel belongs to class j. Depending on the classifier, $\mu_i^j(x)$ can be of a different nature: probability, posterior probability at the output of a neural network, membership degree at the output of a fuzzy classifier, etc. In any case, the set $\pi_i(x) = \{\mu_i^j(x),\ j = 1, \ldots, n\}$ can be considered as a fuzzy set.

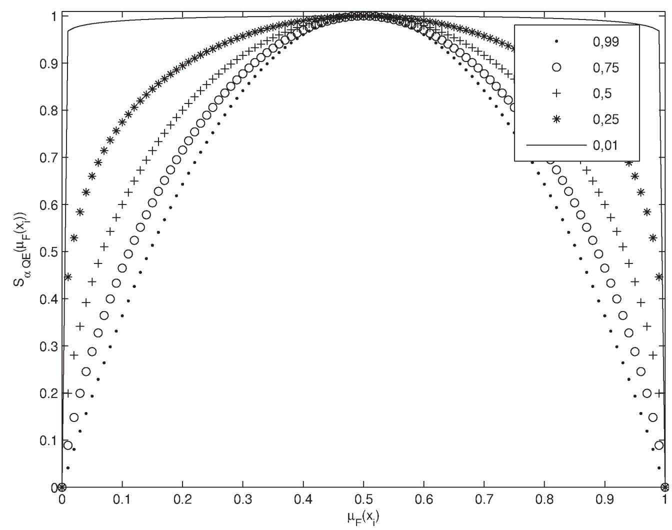


Fig. 1. Influence of α on $S_{\alpha\mathrm{QE}}$.

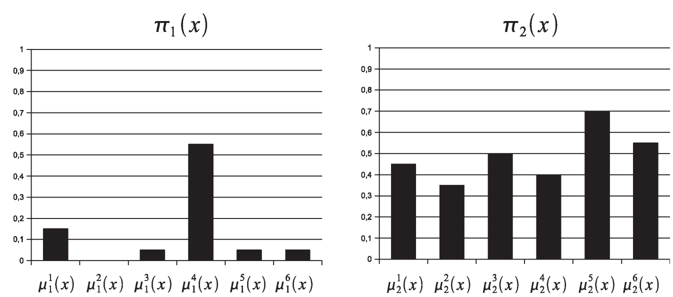


Fig. 2. Example of two conflicting sets π for a given pixel x.

As a conclusion, for every pixel, m fuzzy sets are computed, one by each classifier. They constitute the input for the fusion process, i.e.,

$$\{\pi_1(x), \pi_2(x), \ldots, \pi_i(x), \ldots, \pi_m(x)\}. \tag{10}$$

In Fig. 2, two conflicting sets are represented. In fact, the fusion of the nonconflicting results is of little interest in our case. Although it might increase our belief in the corresponding result, it certainly will not change the final decision and, thus, will not increase the classification performances. On the contrary, in the case of conflicting results, at least one classifier is wrong, and the fusion gives a chance to correct this and increase the classification performances. Fuzzy set theory provides various combination operators to aggregate these fuzzy sets. Such combination operators are discussed in the next section.

For visual inspection, membership maps can be plotted. A membership map represents the $\mu_i^j(x)$ for all pixels for class i. For instance, Fig. 6(a)–(d) shows the membership degrees to the class "buildings" and "houses" obtained for each pixel in an IKONOS image by two different classifiers, respectively.

III. Information Fusion

After briefly reviewing the basics of data fusion, we discuss in this section the problem of measuring the confidence of individual classifiers. Finally, we propose an adaptive fusion operator. In the following, we denote the fuzzy set i by π_i and the number of sources by m.

A. Introduction

Data fusion consists in combining information from several sources to improve the decision [23]. Of course, the most challenging issue is to solve conflicting situations where the sources disagree. Numerous combination operators have been proposed in the literature. They can be classified into three different kinds, depending on their behavior [24].

1) *Conjunctive combination:* This corresponds to a "severe" behavior. The resulting fuzzy set is necessarily smaller than the initial sets, and the core is included in the initial cores (it can only decrease). The largest conjunctive operator is the fuzzy intersection (2). T-norms are conjunctive operators, leading to a fuzzy set $\pi_\wedge(x) = \bigcap_{i=1}^{N} \pi_i(x)$. They are commutative, associative, increasing, and with $\pi_i(x) = 1$ as a neutral element [i.e., if $\pi_2(x) = 1$, then $\pi_\wedge(x) = \pi_1(x) \cap \pi_2(x) = \pi_1(x)$]. They satisfy the following property:

$$\pi_\wedge(x) \leq \min_{i \in [1,m]} \pi_i(x). \tag{11}$$

2) *Disjunctive combination:* This corresponds to an "indulgent" behavior. The resulting fuzzy set is necessarily larger than the initial sets, and the core contains the initial cores (it can only increase). The smallest disjunctive operator is the fuzzy union (1). T-conorms are disjunctive operators, leading to a fuzzy set $\pi_\vee(x) = \bigcup_{i=1}^{N} \pi_i(x)$. They are commutative, associative, increasing, and with $\pi_i(x) = 0$ as neutral element. They satisfy the following property:

$$\pi_\vee(x) \geq \max_{i \in [1,m]} \pi_i(x). \tag{12}$$

3) *Compromise combination:* This corresponds to intermediate "cautious" behaviors. $T(a, b)$ is a compromise combination if it satisfies

$$\min(a, b) < T(a, b) < \max(a, b). \tag{13}$$

For purpose of illustration, we can consider the following toy problem. To estimate how old a person is, two estimates are available, each one modeled by a fuzzy set. These fuzzy sets are represented in Fig. 3(a)—note that they are highly conflicting. From these two information sources, we want to classify a person into one of three classes, i.e., 1) young (under 30); 2) middle aged (between 30 and 65); or 3) old (above 65). To illustrate the three possible modes of combination, we aggregate the information with the min operator (T-norm), the max operator (T-conorm), and the three different compromise operators. Results are presented in Fig. 3. The decision is taken by selecting the class corresponding to the maximum membership.

1) *Conjunctive combination:* Fig. 3(b) presents the result obtained with the min operator, i.e., the less severe conjunctive operator. It is a unimodal fuzzy set. This fuzzy set is subnormalized, but this problem could be solved using $\pi'_\wedge(x) = \pi_\wedge(x)/\sup_x(\pi_\wedge(x))$, but this would not change the shape of the result. In this case, the decision would be "middle aged," which is not compatible with any of the initial sources. In this case, the sources strongly disagree, and the conjunctive fusion does not help in the classification. As a conclusion, conjunctive operators are not suited for conflicting situations.
2) *Disjunctive combination:* Fig. 3(c) presents the result obtained with the max operator, i.e., the less indulgent disjunctive operator. The resulting membership function is multimodal, and each maximum is of equal amplitude. Again, no satisfactory decision can be made.
3) *Compromise combination:* Three different such operators are discussed. They are all based on the measure of the conflict between sources. The conflict may be defined as $1 - C$ with

$$C(\pi_1, \pi_2) = \sup_x \min\left(\pi_1(x), \pi_2(x)\right). \tag{14}$$

The first compromise combination operators were proposed by Prade and Dubois [25]. Bloch has classified these operators as contextual dependent (CD) operators [26]. Note that the context can be characterized in different ways, i.e., knowledge regarding the potential conflict between the sources, knowledge about the reliability of one given source, or introduction of some spatial information. These operators have been proposed in possibility theory [27], but they can also be used in fuzzy set theory for combining membership functions [26]. Being able to adapt to the context, these operators are more flexible and thus provide interesting results. The first considered operator (as shown in (15), at the bottom of the page) adapts its behavior as a function of the conflict between the sources.

- It is conjunctive if the sources have low conflict.
- It is disjunctive if the sources have high conflict.
- It behaves in a compromise way in case of partial conflict.

Fig. 3(d) presents the obtained result using operator (15). Corresponding decision (middle aged) is still not satisfactory.

In this case, some information on source reliability must be included, and the most reliable source(s) should be privileged in the fusion process. Different situations can be considered.

- It is possible to assign a numerical degree of reliability to each source.

$$\pi(x) = \begin{cases} \max\left(\frac{\min(\pi_1(x),\pi_2(x))}{C(\pi_1,\pi_2)}, \min\left(\max\left(\pi_1(x), \pi_2(x)\right), 1 - C(\pi_1, \pi_2)\right)\right), & \text{if } C(\pi_1, \pi_2) \neq 0 \\ \max\left(\pi_1(x), \pi_2(x)\right), & \text{if } C(\pi_1, \pi_2) = 0 \end{cases} \tag{15}$$

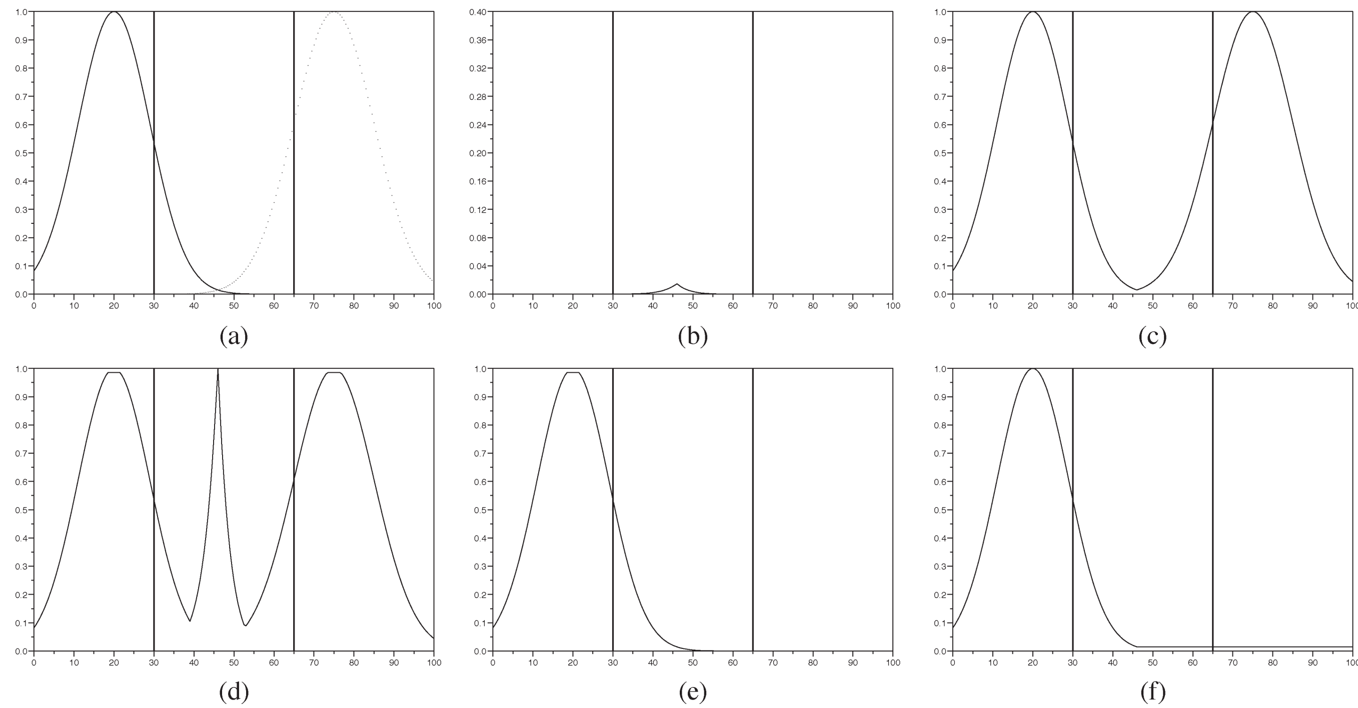


Fig. 3. Examples of combination operator. (a) Two possibility distributions. (b) and (c) Result of the min and the max operators, respectively. (d)–(f) Result of the three compromise operators presented in (15)–(17), respectively.

- A subset of sources is reliable, but we do not know which one(s).
- The relative reliability of the sources are known, but with no quantitative values. However, priorities can be defined between the sources.

The two following adaptive operators are examples of "prioritized fusion operator" [25]:

$$\pi(x) = \min\left(\pi_1(x), \max\left(\pi_2(x), 1 - C(\pi_1, \pi_2)\right)\right) \quad (16)$$

$$\pi(x) = \max\left(\pi_1(x), \min\left(\pi_2(x), C(\pi_1, \pi_2)\right)\right). \quad (17)$$

For both operators, when $C(\pi_1, \pi_2) = 0$, π_2 contradicts π_1, and the only information provided by π_1 is retained. In this case, π_2 is considered as a specific piece of information, whereas π_1 is viewed as a fuzzy default value. Assuming π_1 is more accurate than π_2, we get the result presented in Fig. 3(e) and (f), enabling a satisfactory decision.

As a conclusion, conjunctive and disjunctive combination operators are ill suited to handle conflicting situations. These situations should be solved by CD operators, incorporating reliability information.

B. Measure of Confidence

1) Pointwise Accuracy: For a given pixel and a given classifier, we propose to interpret the fuzziness of the fuzzy set $\pi_i(x)$ defined in (9) as a pointwise measure of the accuracy of the method. We intuitively consider that the classifier is "reliable" if one class has a high membership value, whereas all the others have a membership value close to zero. On the contrary, when no membership value is significantly higher than the others, the classifier is "unreliable," and the results it provides should not be taken too much into account in the final decision. In other words, uncertain results are obtained when the fuzzy set $\pi_i(x)$ has a high fuzziness degree, the highest degree being reached for uniformly distributed membership values.

To reduce the influence of unreliable information and thus enhance the relative weight of reliable information, we weight each fuzzy set by

$$\begin{cases} w_i = \dfrac{\sum\limits_{k=0, k\neq i}^{m} H_{\alpha \mathrm{QE}}(\pi_k)}{(m-1)\sum\limits_{k=0}^{m} H_{\alpha \mathrm{QE}}(\pi_k)} \\ \sum_{i=0}^{m} w_i = 1 \end{cases} \quad (18)$$

where $\alpha = 0.5$, $H_{\alpha \mathrm{QE}}(\pi_k)$ is the fuzziness degree of source k, and n is the number of sources. When a source has a low fuzziness degree, w_i is close to 1, and it only slightly affects corresponding fuzzy set. Fig. 4 illustrates the effects of this normalization.

2) Global Accuracy: Beyond the adaptation to the local context described in the previous paragraph, we can also use prior knowledge regarding the performances of each classifier. This knowledge is modeled for each classifier i and for each class j by a parameter f_i^j. Such global accuracy can be determined by a separate statistical study on each of the used classifiers. If, for a given class j, the user considers that the results provided by classifier i are satisfactory, parameter f_i^j is set to 1. Otherwise, it is set to 0. In as much as this decision is binary, we assume for each class that there is at least one method ensuring a satisfactory global reliability.

3) Combination Operator: Many combination rules have been proposed in the literature, from simple conjunctive or

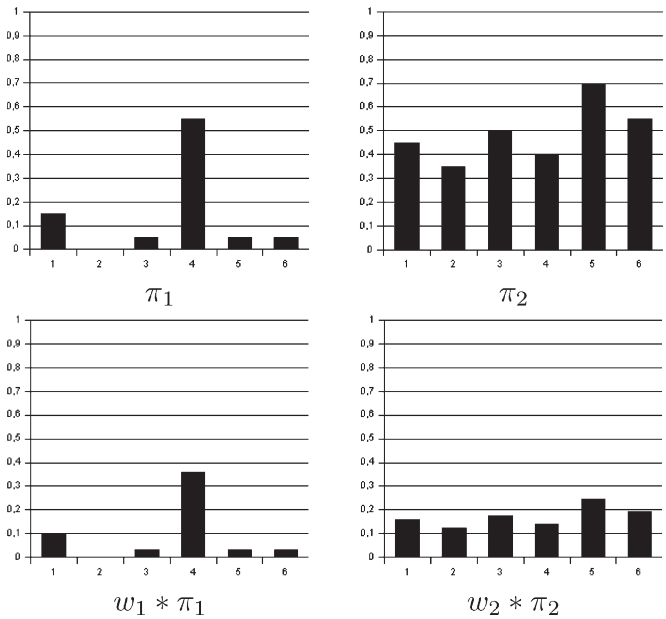


Fig. 4. Normalization effects. This figure shows two fuzzy sets (π_1 and π_2) with different fuzziness ($H_{\alpha\mathrm{QE}}(\pi_1) = 0.51$, $H_{\alpha\mathrm{QE}}(\pi_2) = 0.97$, $w_1 = 0.65$, and $w_2 = 0.35$). The normalization effect is shown on the lower line. Influence of classifier 2 is more reduced by w_2 than classifier 1 is reduced by w_1.

disjunctive rules, such as min or max operators, to more elaborated CD operators, such as defined by (16) and (17), where the relative reliability of each source is used. However, with these operators, sources have always the same hierarchy, and the fusion scheme does not adapt to the local context. In this paper, we propose the following extension:

$$\mu_f^j(x) = \max\left(\min\left(w_i \mu_i^j(x), f_i^j(x)\right), \quad i \in [1, m]\right) \quad (19)$$

where f_i^j is the global confidence of source i for class j, w_i is the normalization factor defined in (18), and μ_i^j is an element of the fuzzy set π_i defined in (9). This combination rule ensures that only reliable sources are taken into account for each class (predefined coefficients f_i^j) and that the fusion also automatically adapts to the local context by favoring the source that is locally the most reliable (weighting coefficients w_i).

IV. Fusion Scheme

We present here the complete proposed fusion scheme. In a first step, each classifier is applied separately (but no decision is taken). Then, the results provided by the different algorithms are aggregated. The final decision is taken by selecting the class with the largest resulting membership value.

The fusion step is organized as follows. For each pixel

1) Separately build fuzzy sets for the classes in each source.
2) Compute the fuzziness degree of each fuzzy set.
3) Normalize data with w_i defined in (18).
4) Apply operator (19).
5) Select the class corresponding to the highest resulting membership degree.

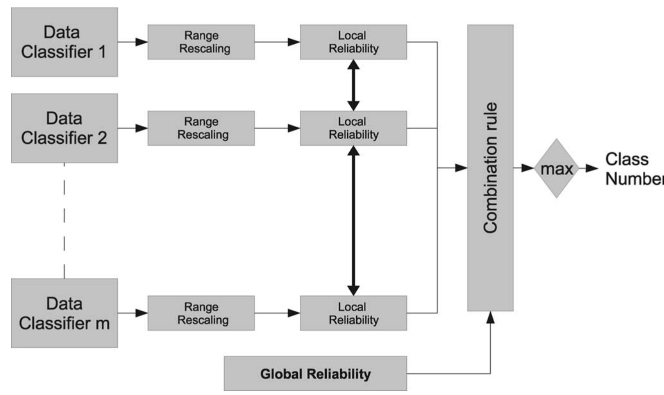


Fig. 5. Block diagram of the fusion method.

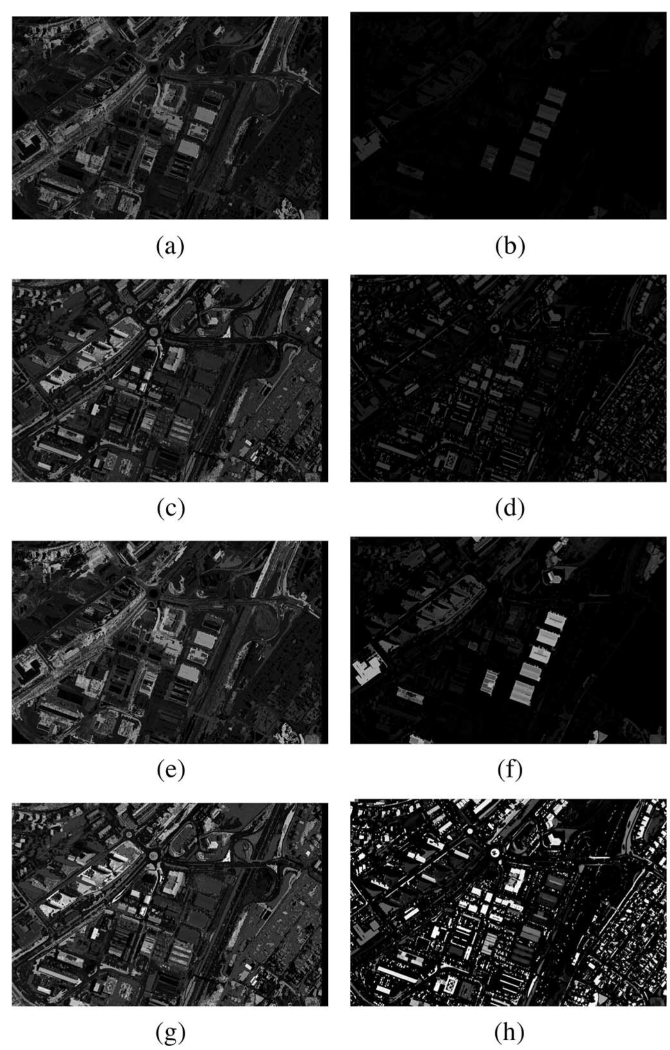


Fig. 6. Possibility maps. (a) and (c) Membership maps given by the neural network for the buildings and houses classes, respectively. (b) and (d) Membership maps given by the fuzzy classifier for the classes buildings and houses, respectively. (e)–(h) Stretched versions of the four images with the algorithm given Section IV.

The block diagram of the fusion process is given in Fig. 5. Note that in Fig. 5, the range of the fuzzy sets is rescaled before the fusion step to combine data with the same range. The rescaling has to save the order relation in the classifier's outputs for a

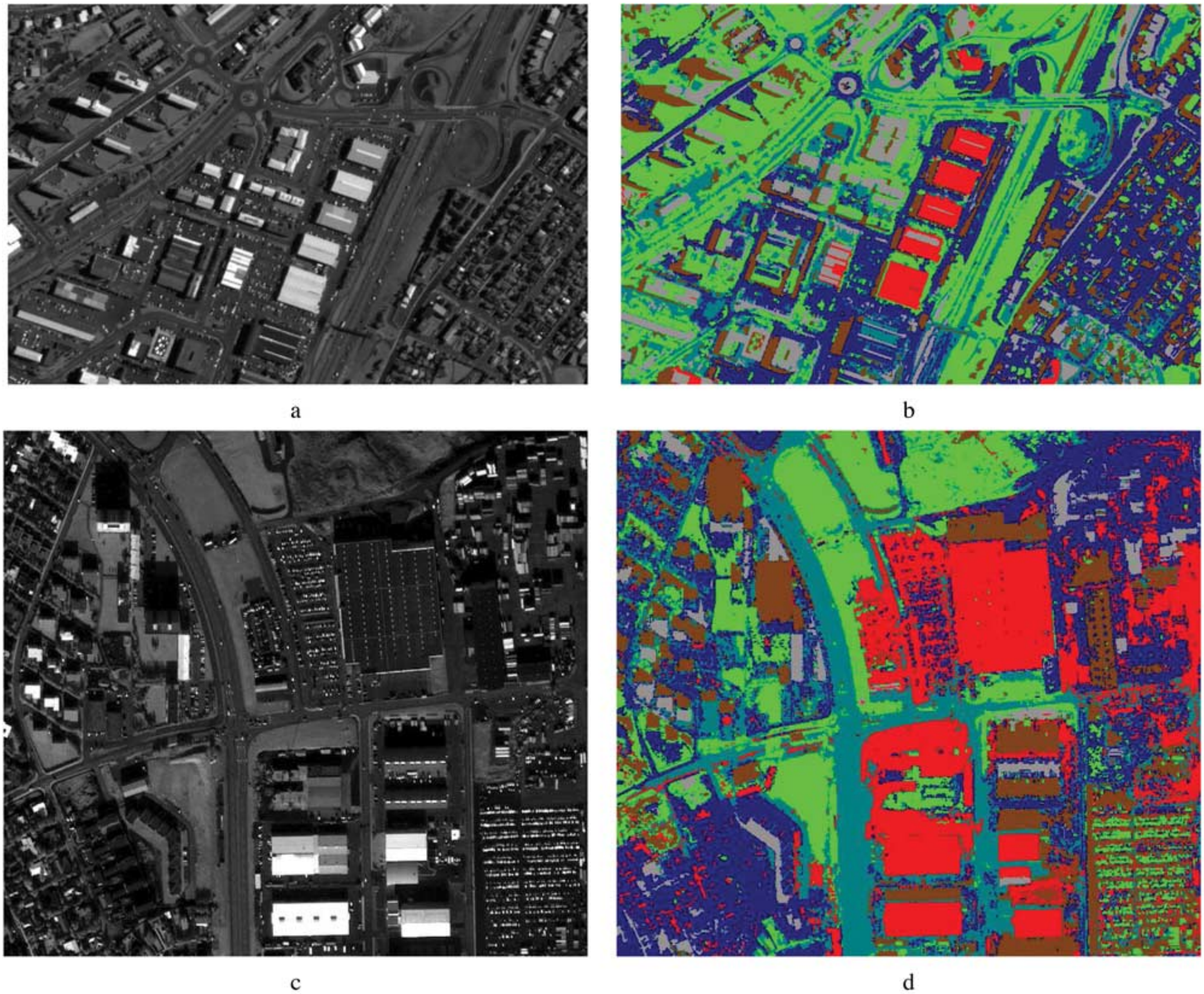


Fig. 7. Test images and results. (a) Original IKONOS image 1. (b) Image 1 classification results. (c) Original IKONOS image 2. (d) Image 2 classification results. (Red) Large buildings. (Gray) Small houses. (Dark blue) Streets. (Blue-gray) Large road. (Green) Open areas. (Brown) Shadows.

given pixel. This is achieved with the following range stretching algorithm:

- for all $\pi_i(x) = \{\mu_i^1(x), \ldots, \mu_i^j(x), \ldots, \mu_i^n(x)\}$, compute:
  - $M = \max_{j,x}[\mu_i^j(x)]$,
  - $m = \min_{j,x}[\mu_i^j(x)]$,
  - for all $\mu_i^j(x)$, compute:

$$* \; \mu_i^j(x) = \frac{\mu_i^j(x) - m}{M - m}.$$

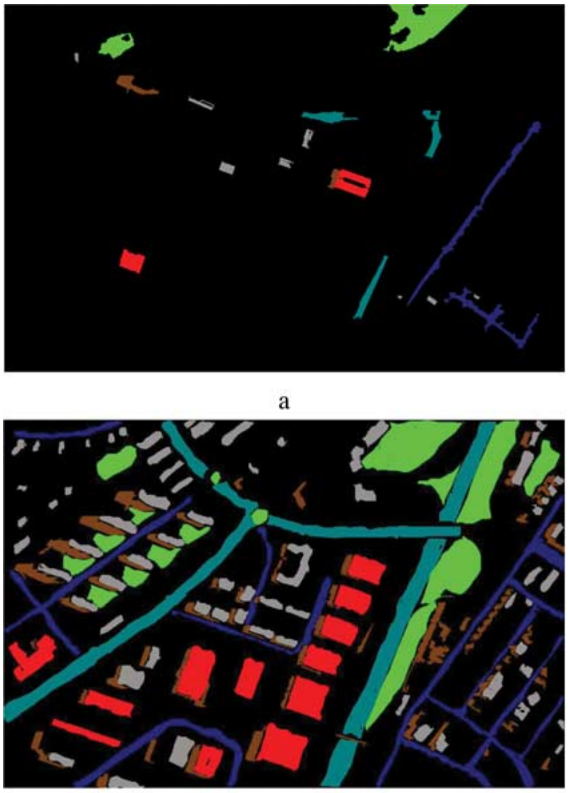
V. Experimental Results

In this section, we present the application of the proposed general fusion scheme to the improvement of classification results using remote sensing images from urban areas. The proposed approach was applied to two very high resolution IKONOS panchromatic images from Reykjavik, Iceland. Six classes were considered in each case, namely: 1) large buildings; 2) houses; 3) large roads; 4) streets; 5) open areas; and 6) shadows. Each image consists of a single channel with 1-m resolution.

Two classification algorithms were used, i.e., 1) a conjugate gradient neural network [2] and 2) a fuzzy classifier [8]. Both are composed of two steps. The first step is the feature extraction by morphological filters, and the second step is the actual classification, using either a neural network or a fuzzy possibilistic model. The classification accuracies for the different classifiers were compared to determine the global confidence in the fusion process. The inputs for the fusion process were the posterior probabilities from the outputs of the neural network and the membership degrees for the fuzzy classifier. These inputs are displayed as images in Fig. 6.

A. *First Test Image*

The first test image (976 × 640 pixels) is shown in Fig. 7(a). Table II shows the test accuracies for the two classifiers. To test the generalization ability of the classifiers, independent samples



a

b

Fig. 8. (a) Training map. (b) All available reference samples for image 1.

TABLE I
INDEXES OF CONFIDENCE FOR IMAGE 1

	Neural Network	Fuzzy Logic
Large Buildings	0	1
Houses	0	1
Large Roads	1	1
Streets	1	0
Open Areas	0	1
Shadows	0	1

were used for training and testing. The training map and the reference map are shown in Fig. 8(a) and (b), respectively.

Starting from the class classification accuracies, the global reliabilities were set as follows. The neural network classifier gave higher accuracies than the fuzzy classifier for the streets and the large roads classes. However, for the other four classes, the fuzzy classifier outperformed the neural network in terms of accuracies. In the fusion, we defined the indexes of confidence in a binary way according to the accuracies. For a given class, full confidence was given to the best classifier, i.e., the one with the highest classification accuracy. Then, if the accuracy of the other classifier was close to the highest (by 5%), then full confidence was also granted to that classifier. Otherwise, the index of confidence was set to 0. The confidence values are listed in Table I.

The accuracy obtained for the final classification is given Table II. The overall accuracy increased from 40.3% for the neural network and 52.1% for fuzzy the classifier to 59.1% with the fusion. After the fusion, houses and large buildings were classified with similar accuracies as by using the fuzzy classifier alone. However, this classification of streets improved in terms of accuracies from 9.8% for the fuzzy classifier to 55.7% with the fusion. The classification accuracies for shadows and open area also increased from 83.8% and 52.1%, respectively, to 86.6% and 60.9%, respectively. On the other hand, the classification accuracy for large roads decreased from 59.1% to 43.7%. Both the original and classified images are shown in Fig. 7.

TABLE II
TEST ACCURACIES IN PERCENT FOR IMAGE 1

	Neural Network	Fuzzy Logic	Fusion
Large Buildings %	26.2	47.6	47.4
Houses %	33.4	67.8	67.4
Large Roads %	59.1	58.8	43.7
Streets %	55.6	9.8	55.7
Open Areas %	30.9	52.2	60.9
Shadows %	32.7	83.3	86.6
O. A. %	40.3	52.1	59.1
A. A. %	39.7	53.3	60.3

TABLE III
INDEXES OF CONFIDENCE FOR IMAGE 2

	Neural Network	Fuzzy Logic
Large Buildings	1	0
Houses	0	1
Large Roads	1	1
Streets	1	0
Open Areas	1	1
Shadows	0	1

The results of the first experiment illustrate the complementary behaviors of the fuzzy and neural network classifiers. Although the global accuracy is higher with the fuzzy classifier, the neural classifier performs better in terms of accuracies for the large roads and streets classes. Note that these accuracy numbers were obtained using manual ground truth where each pixel in the original image was labeled. In as much as no pre- or postprocessing was done, the accuracies should be interpreted in a relative way rather than in an absolute way.

B. *Second Test Image*

The second test image is 700 $\times$ 630 pixels. Table IV shows the test accuracies for the two classifiers that were used in the second experiment. The global reliability was defined in the same way as in the first experiment. The indexes of confidence are listed in Table III.

The test accuracies for the final classification is given in Table IV. As can be seen in the table, the overall accuracy increased from 57.0% for the neural network and 43.1% for fuzzy classifier to 75.7% after the fusion. With the fusion, classification accuracy for open areas increased from 46.6% to 73.7%. Shadows and large buildings classification accuracies were similar for the fuzzy classifier and the neural network. The biggest improvement after the fusion was achieved in the classification of large roads, where the classification accuracy increased from 0.0% to 94.2%. Furthermore, the overall road classification accuracy increased from 41.5% to 58.6%. However, at the same time, the classification accuracy for streets

TABLE IV
TEST ACCURACIES IN PERCENT FOR IMAGE 2

	Neural Network	Fuzzy Logic	Fusion
Large Buildings %	89.6	26.3	94.8
Houses %	29.9	42.8	33.8
Large Roads %	0	0	94.2
Streets %	83.6	77.4	22.7
Open Areas %	46.5	44.9	73.7
Shadows %	43.7	98.7	90.4
O. A. %	57.0	43.1	75.7
A. A. %	48.9	48.4	68.3

TABLE V
TEST ACCURACIES IN PERCENT FOR DIFFERENT COMBINATION RULES WITHOUT THE POINTWISE ACCURACY MEASURE OF IMAGE 1

	Max	Min	Operator (15)	Operator (16)	Operator (17)
Large Buildings %	31.6	42.8	32.7	47.8	39.6
Houses %	68.2	67.2	65.1	67.6	64.3
Large Roads %	66.4	68.0	66.4	59.4	69.6
Streets %	2.1	5.9	2.1	7.2	4.2
Open Areas %	9.1	9.1	9.1	8.3	13.1
Shadows %	52.8	81.1	52.8	84.4	53.5
O. A. %	37.0	43.0	36.7	42.6	39.5
A. A. %	38.4	45.7	38.0	46.1	40.7

TABLE VI
TEST ACCURACIES IN PERCENT FOR DIFFERENT COMBINATION RULES WITH THE POINTWISE ACCURACY MEASURE OF IMAGE 1

	Max	Min	Operator (15)	Operator (16)	Operator (17)
Large Buildings %	48.4	40.8	48.4	47.8	47.6
Houses %	70.2	55.6	70.2	67.8	67.3
Large Roads %	59.7	71.6	59.7	59.4	59.7
Streets %	6.1	7.2	6.1	7.2	6.8
Open Areas %	8.1	9.7	8.1	8.3	8.4
Shadows %	84.2	70.9	84.2	84.3	84.1
O. A. %	42.9	40.5	42.9	42.7	42.5
A. A. %	46.1	42.6	46.1	46.2	45.7

decreased from 83.6% to 22.7%. Both the original and the classified images are shown in Fig. 7.

C. Comparison With Other Combination Rules

In this section, we compare the results provided by the proposed operators with others combination rules. When possible, we use the accuracy measure previously defined in Section III-B. For the min and max operators, we compute experiments with and without pointwise accuracy information. We do the same for operator (15). Conflict was computed for both cases. For operators (16) and (17), the less accurate classifier was chosen as the less important classifier based on the global test accuracy.

The obtained results are given in Tables V and VI. As can be seen from the tables, our proposed method outperformed the others combination rules in terms of accuracy. It can be seen that the classification accuracy for streets is still not satisfactory. No combination rule was able to use the information provided by the neural network.

For the max operator, the pointwise accuracy information improved the classification accuracy as compared with the fusion with the max operator without pointwise accuracy information. That was due to the "normalization effect": the unreliable information was reduced due to operator (18). Conversely, pointwise accuracy information deteriorated the classification with the min operator. Adaptive operators (15) and (17) seemed to perform better with pointwise accuracy information; the global accuracy for operators (15) and (17), respectively, increased from 36.7% and 39.5% to 42.9% and 42.5%. No significant changes were noted for operator (16). From these experiments, it can be concluded that if there is no available information on source reliability, pointwise accuracy can be used to significantly improve the fusion. However, knowledge about the global reliability of each classifier seems to be more useful. Finally, to investigate the influence of the contextual information, two additional experiments were conducted. In each experiment, we removed one type of the contextual information types and compared the results in terms of classification accuracies with those obtained if two types of contextual information are used. For the global information, if we set its values to 1 for both classifiers and all classes, operator (19) becomes the simple max operator using pointwise accuracy information; the experiment was already done in the previous paragraph. For the pointwise accuracy information, all w_i were put to one, and we kept only the global information. Results are listed in Table VII. From these experiments, it is clear that both contextual information types are needed to achieve a good classification in terms of accuracies.

TABLE VII
TEST ACCURACIES IN PERCENT FOR OPERATOR (19) WITH DIFFERENT TYPES OF CONTEXTUAL INFORMATION

	Point-wise accuracy	Global accuracy	Both accuracies
Large Buildings %	48.4	42.9	47.7
Houses %	70.2	67.2	67.4
Large Roads %	59.7	64.5	43.7
Streets %	6.1	4.9	55.7
Open Areas %	8.1	37.0	60.9
Shadows %	84.2	92.8	86.6
O. A. %	42.9	49.5	59.1
A. A. %	46.1	51.5	60.3

The results of these additional experiments demonstrate the need to control the fusion process. Without information about conflict, accuracy, and confidence, the accuracies are generally worse than before the fusion. Although the pointwise accuracy is easy to compute and is independent of the classifiers, global accuracy is a critical problem of the method. More developments are needed to automate their definition.

VI. CONCLUSION

The fusion of several classifiers has been considered in the classification of panchromatic remote sensing data from urban areas. Starting from a complementary use of different classifiers, the proposed method is based on a fuzzy combination rule. Two measures of accuracy are used in the combination rule: The first one, based on prior knowledge, defines global reliabilities, both for each classifier and each class. The second one automatically estimates the pointwise reliability of the results provided by each classifier and, thus, enables the adaptation of the fusion rule to the local context. The proposed approach does not need any training and only requires about 1 min of computation time for each image using a Pentium IV personal computer. Furthermore, no prior assumptions are needed regarding the modeling of the data (e.g., Bayes theory, possibility theory, etc.) before the data are fused.

The obtained experimental results show that the complementary use of different classifiers leads to significant improvement of global classification accuracies. The overall accuracy was improved by about 7% in the first experiment and 18% in the second experiment.

A key point in the presented framework lies in its generality for "decision-level fusion." Although only two classifiers were used in the paper, additional algorithms could easily be added to the process. For instance, specialized algorithms such as street detectors could be used without increasing errors in building detection. This generalization also holds for the inclusion of multisource data such as multispectral or multitemporal images. One algorithm could be used on each image, and then, the fusion could be done with the results computed on each image.

In this paper, αQE was chosen for the fuzziness evaluation because the sensitivity of that measure can be modified with the value of α. Several other measures could be used, e.g., "fuzzy entropy" [19].

One limitation of the proposed approach is the use of binary values for global confidence. With fuzzy confidence, the combination rule could be rewritten with T-conorm and T-norm, both of which are less indulgent and less severe than max and min. Moreover, the use of T-conorm and T-norm would allow a finer definition of global accuracy.

Our current research is now oriented toward fusion of spectral and spatial classification results. That way, we can integrate much complementary information for the final classification process.

References

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.

[2] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[3] M. Pesaresi, "Textural classification of very high resolution satellite imagery: Empirical estimation of the interaction between window size and detection accuracy in urban environment," in *Proc. IEEE ICIP*, Oct. 1999, vol. 1, pp. 114–118.

[4] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1920–1932, Sep. 2003.

[5] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 981–996, Jul. 1995.

[6] J. Chanussot, J. A. Benediktsson, and M. Vincent, "Classification of remote sensing images from urban areas using a fuzzy model," in *Proc. IGARSS*, Sep. 2004, vol. 1, pp. 556–559.

[7] F. Melgani, B. A. R. al Hashemy, and S. M. R. Taha, "An explicit supervised classification method for multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 287–295, Jan. 2000.

[8] J. Chanussot, J. A. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 40–44, Jan. 2006.

[9] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

[10] B. Jeon and D. A. Landgrebe, "Decision fusion approach for multitemporal classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1227–1233, May 1999.

[11] G. Lisini, F. Dell'Acqua, G. Triani, and P. Gamba, "Comparison and combination of multiband classifiers for Landsat urban land cover mapping," in *Proc. IGARSS*, Jul. 2005, vol. 4, pp. 2823–2826.

[12] F. Tupin, I. Bloch, and H. Maitre, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1327–1343, May 1999.

[13] J. Chanussot, G. Mauris, and P. Lambert, "Fuzzy fusion techniques for linear features detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1292–1305, May 1999.

[14] G. Amici, F. Dell'Acqua, P. Gamba, and G. Pulina, "A comparison of fuzzy and neuro-fuzzy data fusion for flooded area mapping using SAR images," *Int. J. Remote Sens.*, vol. 25, no. 20, pp. 4425–4430, Oct. 2004.

[15] R. R. Yager, "A general approach to the fusion of imprecise information," *Int. J. Intell. Syst.*, vol. 12, no. 1, pp. 1–29, Jan. 1997.

[16] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[17] G. J. Klir and B. Yuan, *Fuzzy Set and Fuzzy Logic: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[18] K. Tanaka, *An Introduction to Fuzzy Logic for Practical Application*. Berlin, Germany: Springer-Verlag, 1996.

[19] A. DeLuca and S. Termini, "A definition of nonprobabilistic entropy in the setting of fuzzy set theory," *Inf. Control*, vol. 20, no. 4, pp. 301–312, May 1972.

[20] L. A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, no. 2, pp. 421–427, Aug. 1968.

[21] B. R. Ebanks, "On measures of fuzziness and their representations," *J. Math. Anal. Appl.*, vol. 94, no. 2, pp. 24–37, Aug. 1983.

[22] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 2, pp. 107–118, May 1994.

[23] I. Bloch, *Fusion d'Informations en Traitement du Signal et des Images*, H. Sciences, Ed. Paris, France: Germes Lavoisier, 2003.

[24] M. Oussalah, "Study of some algebraical properties of adaptive combination rules," *Fuzzy Set Syst.*, vol. 114, no. 3, pp. 391–409, Sep. 2000.

[25] D. Dubois and H. Prade, "Possibility theory in infomation fusion," in *Data Fusion and Perception, CISM Courses and Lectures*, vol. 431, G. Della Riccia, H. Lenz, and R. Kruse, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 53–76.

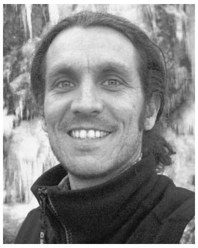
[26] I. Bloch, "Information combination operators for data fusion: A comparative review with classification," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 26, no. 1, pp. 52–67, Jan. 1996.

[27] D. Dubois and H. Prade, *Combination of Information in the Framework of Possibility Theory*, M. Al Albidi *et al.*, Ed. New York: Academic, Jan. 1992, pp. 481–505.

Mathieu Fauvel (S'06) received the degree in electrical engineering and the M.S. degree in image and signal processing from the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 2004. He is currently working toward the Ph.D. degree from conjointly the INPG and the University of Iceland, Reykjavik.

His research interests are remote sensing, data fusion, pattern recognition, and signal and image processing.

Mr. Fauvel is a Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



Jocelyn Chanussot (M'04–SM'04) received the degree in electrical engineering from the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 1995, and the Ph.D. degree from Savoie University, Annecy, France, in 1998.

He was with the Automatics and Industrial Micro-Computer Science Laboratory. In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (French National Defense Department). Since 1999, he has been an Associate Professor of signal and image processing at INPG and has been working at the Laboratoire des Images et des Signaux, Grenoble. He is especially involved in the team working on geophysical and underwater acoustical signal and image processing. His research interests include multicomponent image processing, nonlinear filtering, remote sensing, data fusion, and their applications to natural real bidimensional data (e.g., seismic data, sonar images, and optical remote sensing).

Dr. Chanussot is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is the Co-Chair of the GRS Data Fusion Technical Committee and a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

He is currently a Professor of electrical and computer engineering with the University of Iceland. He has held visiting positions at the Department of Information and Communication Technology, University of Trento, Trento, Italy (2002–present), School of Computing and Information Systems, Kingston University, Kingston upon Thames, U.K. (1999–2004), the Joint Research Centre of the European Commission, Ispra, Italy (1998), Denmark's Technical University (DTU), Lyngby (1998), and the School of Electrical and Computer Engineering, Purdue University (1995). He was a Fellow at the Australian Defence Force Academy, Canberra, A.C.T., Australia, in August 1997. From 1999 to 2004, he was a Chairman of the energy company Metan Ltd. His research interests are in remote sensing, pattern recognition, neural networks, image processing, and signal processing, and he has published extensively in those fields.

Dr. Benediktsson is an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGARS) and an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was an Associate Editor of TGARS from 1999 to 2002. He coedited (with Prof. D. A. Landgrebe) a Special Issue on Data Fusion of TGARS (May 1999). In 2002, he was appointed Vice President of Technical Activities in the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) and (with P. Gamba and G. G. Wilkinson) a Special Issue on Urban Remote Sensing from Satellite (October 2003). From 1996 to 1999, he was the Chairman of GRSS Technical Committee on Data Fusion and was elected to the Administrative Committee of the GRSS for the term 2000 to 2002, and in 2002, he was appointed Vice President of Technical Activities of GRSS. He was the founding Chairman of the IEEE Iceland Section and served as its Chairman from 2000 to 2003. He was the Chairman of the University of Iceland's Science and Research Committee (1999–2005), a member of Iceland's Science and Technology Council (2003–2006), and a member of the Nordic Research Policy Council (2004). He was a member of a NATO Advisory Panel of the Physical and Engineering Science and Technology SubProgramme (2002–2003). He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, and in 2000, he was granted the IEEE Third Millennium Medal. He is a member of Societas Scinetiarum Islandica and Tau Beta Pi.