FUZZY CLASSIFICATION FOR MULTI-MODALITY IMAGE FUSION

Isabelle BLOCH

Télécom Paris, Département Images
46 rue Barrault, F-75634 Paris Cedex 13
Tel: 33 1 45 81 75 85, Fax: 33 1 45 81 37 94, Email: bloch@ima.enst.fr

ABSTRACT

In the framework of fuzzy set theory, we propose: (i) a classification scheme for multi-modality image fusion, where membership degrees to a class issued from several images are combined before taking a decision, (ii) a classification of fusion operators, depending on their behaviour.

1. INTRODUCTION

In a context of multi-modality image fusion, fuzzy set theory appears as a powerful framework since it provides many tools adapted to this task. Its main properties, which will be exploited in this paper, are the following:

- It provides a way to represent imprecise and uncertain information [16], including its characteristics in terms of redundancy and complementarity which are key features in data fusion.
- It is able to manage knowledge about the sources (conflict, reliability).
- It provides a large series of operators for combining information issued from different images, along with decision rules.
- At last, this theory is well adapted to image processing since the natural spatial interpretation of fuzzy sets leads to efficient representations of imprecise structures or classes in pictures [10].

In the next section, we briefly describe a scheme for fuzzy classification from several images where the decision is taken only at the end of the fusion process. In section 3, we propose a classification of fuzzy fusion operators depending on their behaviour and provide some criteria for choosing an operator adapted to the problem at hand. In section 4, some applications will illustrate our purpose.

2. IMAGE FUSION BY FUZZY CLASSIFICATION

The principle of fuzzy classification consists in assigning to each point a membership degree to each class of interest, depending on characteristics of the point and of the classes. Generalizing this principle to multi-modality images avoids to define a metrics on the space of characteristics. Following this idea, we propose a scheme for classification by multi-modality image fusion which consists of 4 steps:

1. define characteristics of points and classes extracted from each image;
2. define a membership function $\mu_i^j(x)$ in [0,1] which describes, for each image j, the membership degree of x to the class i; this step plunges all pixel and class information in [0,1], such that they become comparable, whatever the characteristics or the images they are issued from;
3. fuse the membership functions of all images related to each class i, in order to obtain a global membership function $\mu_i(x)$ to the class i;
4. take a decision in favour of a class i following one or several of following rules (C denoting the number of classes):

$$\mu_i(x) = \max_{j=1}^{C} \mu_j(x), \quad (1)$$

$$\mu_i(x) \geq d, \quad (2)$$

$$\mu_i(x) - \max_{j \neq i} \mu_j(x) \geq \eta. \quad (3)$$

The first rule (equation 1) chooses the class to which x has the highest membership. The two other rules control if the decision is strong enough (equation 2) (d is a decision threshold) or discriminating enough (equation 3) (η is a discrimination threshold).

This scheme avoids to take a decision on each image separately, which could lead to conflictual decisions difficult to manage a posteriori since all numerical information is lost. It represents imprecision and uncertainty through fuzzy membership functions and deals with them until the last step of the process.

In the first step, commonly used characteristics are pixel grey-level from each image, texture indices, response to a detector of particular structures, etc. They are usually extracted by image processing techniques (not necessarily fuzzy): for instance the class "road" in a satellite image can be characterized by the response to a morphological top-hat transform or to a Duda Road Operator.

The definition of μ_i^j can be obtained for example from a fuzzy classification [1], [2] performed on each image separately, without taking any definite decision about the belonging of x to a particular class. Several forms can be used (fuzzy C-means, fuzzy k-nearest neighbours, analytical form whose parameters are determined from the data). It can also be obtained directly by a mapping into [0,1] from the grey level domain, representing the characteristics extracted using some image processing operator.

The fusion operators involved in the third step will be described in the next section.

3. COMBINATION OPERATORS: A REVIEW WITH CLASSIFICATION

Any numerical fusion operator (i.e. a function $F(a,b)$ combining two pieces of information a and b in [0,1]) may have three kinds of behaviour: conjunctive (if $F(a,b) \leq a$ and $F(a,b) \leq b$), disjunctive (if $F(a,b) \geq a$ and $F(a,b) \geq b$), compromise (if $F(a,b)$ is between a and b) [6], [15]. We propose a classification of fusion operators in three classes, depending on their behaviour. Details may be found in [4].

3.1. Context Independent Constant Behaviour (CICB) operators

The first class (called context independent constant behaviour) is composed of operators which have the same behaviour whatever the values of the information to combine, and which can be computed only from the values of a and b. They are ordered independently of the values of a and b (if F and F' are two operators in this class, we have either $\forall(a,b), F(a,b) \leq F'(a,b)$, or $\forall(a,b), F(a,b) \geq F'(a,b)$). Examples of fuzzy operators belonging to this class are:

- triangular norms [9], [11], which generalize set intersection to fuzzy sets (like $\min(a,b)$, ab, $\max(0, a+b-1)$); they always have a conjunctive behaviour;
- triangular conorms [9], [11], which generalize set union to fuzzy sets (like $\max(a,b)$, $a+b-ab$, $\min(1, a+b)$); they always have a disjunctive behaviour;
- means (median, geometrical or arithmetical mean, weighted means, fuzzy integrals, etc.) [6], [14], [15], [8]; they always behave in a compromise way.

Let us mention that some operators used by other data fusion theories belong to the CICB class. Examples are the product used in probabilistic and Bayesian approach or the orthogonal sum of Dempster and Shafer in evidence theory [12].

3.2. Context Independent Variable Behaviour (CIVB) operators

The second class is composed of operators which are context independent like in the first class (i.e. $F(a,b)$ depends only on a and b) but whose behaviour depends on the values of a and b. Examples of fuzzy operators belonging to this class are the associative symmetrical sums (but median) [6], [15]: they behave in a conjunctive way if $\max(a,b) < 1/2$, in a disjunctive way if $\min(a,b) > 1/2$, and in a compromise way if $a < 1/2$ and $b > 1/2$ (or the reverse).

Note that the operators used in MYCIN for combining certainty factors [13] are also CIVB (they are disjunctive if $a > 0$ and $b > 0$, conjunctive if $a < 0$ and $b < 0$, else they have a compromise behaviour).

3.3. Context Dependent (CD) operators

The third class is composed of operators which are context dependent, i.e. which are computed not only from a and b but also depend on a global knowledge or measure on the sources to be fused (like conflict between sources, or reliability of sources). For instance, it is possible to built operators which behave in a conjunctive way if the sources are consonant, in a disjunctive way if they are highly dissonant, and like a compromise if they are partly conflicting (for instance, they may be derived from operators used in possibility theory or artificial intelligence [7]). Such operators are particularly interesting for classification problems, since their adaptive feature makes them able to combine information related to one class in one way, and information related to another class in another way.

3.4. Choice of an operator

The above classification provides an efficient way to choose a class operator, by analysing the problem at hand in terms of expected behaviour in the fusion process. In each class, the choice may then be refined depending on the required properties of the operators, which can be interpreted in terms of data fusion. Let us give some examples.

- Commutativity is generally satisfied by the operators, although human reasoning do not always combine information in a commutative way.
- Associativity, although not always necessary, makes the order of combination not important, and thus facilitates the combination of more than 2 pieces of information.
- The existence of a unit element may be imposed if we want some values will have no influence on the result of the combination.
- On the contrary, the existence of a null element will be imposed if we want a value to determine completely the result of the combination.
- Idempotence expresses a stability of the operators which satisfy this property (like mean operators) if a value occurs twice. It may be desirable in case of strong dependence between sensors. This property is opposed to the Archimedian property which expresses that the information is reinforced or weakened if it occurs twice. For instance, if two independent reliable sensors provide the same information about a fact, we tend to trust this result more than each individual information.
- Excluded middle and non-contradiction have a strong interpretation in terms of reasoning. Examples where excluded middle is not desirable occur in problems where we want to introduce ignorance about an event and its contrary (this is typically one of the key features of Dempster-Shafer evidence theory).

Non-associative operators deserve a special attention when combining more than 2 pieces of information. Given an operator F acting on two variables, two ways may lead to the combination of n pieces of information:

- the first way consists in adding successively pieces of information, according to the following formula:

  $$F(x_1, x_2, ..., x_n) = F[F[...F[F(x_1, x_2), x_3]..], x_n],$$

  for a given order of the x_i's, chosen in an adequate way (for instance, following the occurrence order, or according to some priorities between information, or depending on the conflict between information);

- the second way consists in deriving a combination rule for a given number n of variables, by mimicing the rule for two variables; for instance, the generalization of the arithmetical sum is straightforward and leads to:

$$\frac{x_1 + x_2 + \dots + x_n}{n}.$$

As far as behaviour is concerned, it can be easily proved that the generalization of a CICB operator to n variables is still CICB with the same behaviour, even for non associative operators. For CIVB operators, they remain CIVB but the rules governing their behaviour may be more complex than their equivalent for 2 variables. Examples can be found in [4].

Another criterion for choosing among the operators is their behaviour with respect to decision (dealing with conflicting situations, decisiveness, discrimination power, etc.). Details can be found in [4], and we provide here only a few points. Let us take the example of the combination of 3 variables combined by the arithmetical sum. It can be shown that if all values are of the same order of magnitude (i.e. consensual information), all possible combinations are also of the same order of magnitude. Therefore, the choice of one particular combination (since the operator is not associative) will probably not be crucial. On the contrary, if the information is conflictual, the variations may be more significant. For instance, if x_1 is very low and x_2 and x_3 are both high and close to each other, then combining the consensual values first will provide a significantly smaller result (denoted by R_1) than combining conflictual values first. Moreover, the value $(x_1 + x_2 + x_3)/3$ provided by the second approach is farther from the non conflictual values than R_1. As far as decisiveness is concerned, the problem is to qualify the decision taken from the μ_i's obtained by a fusion operator $\mu_i = F(\mu_i^j, j = 1...l)$ where l is the number of images. A strong indication is provided by the fact that most of the operators can be ordered independently of the values to be combined. For instance, we have for T-norms:

$$\forall(a,b), \max(0, a + b - 1) \leq ab \leq \min(a,b).$$

Extreme operators will generally be less decisive than others. For instance a large T-conorm like $\min(1, a + b)$ (resp. a small T-norm like $\max(0, a + b - 1)$) is likely to be saturated to value 1 (resp. 0) when combining several sources, and thus will not well discriminate among the μ_i's. A last remark concerns the choice of the decision threshold and of the discrimination threshold, which, for analogous reasons, cannot be chosen independently of the operator.

4. APPLICATION IN MULTI-MODALITY IMAGE PROCESSING

These operators are commonly used in artificial intelligence, in particular for pooling expert decisions [7]. However, very few applications are developed until now in image processing. Let us mention an example in artificial vision [5] and one in image fusion [3]. Examples showing the interest of our approach can be found in medical and satellite multi-modality imaging. In both domains, the increase of imaging techniques, which provide different types of information about a phenomenon, underlines the role of data fusion for interpretation and diagnosis help. In medical imaging, applications concern tissue classification, segmentation, disease detection (for instance in multi-echo MRI), anatomical-functional correlation (like TEP + MRI). In satellite imaging, examples can be found for classification and detection of various structures (roads for instance) in multi-spectral SPOT or Landsat images. These applications prove the interest of data fusion for the classification, since a better discrimination between classes results. The idea is to exploit on one hand redundancy between images to make the decision more certain, and on the other complementarity to increase the global information. Fuzzy modelling is a well adapted tool for this task. Also the variety of fuzzy fusion operators allows to take into account, in the same model, several kinds of information and to deal with a lot of different situations. In particular, a priori information, like reliability of a given source when reporting on a phenomenon (class, structure), can be adaptatively introduced in the operators, for instance by means of a weighted mean, or of a CD operator.

We will provide here only a very simple example, which shows that the concepts proposed in the previous section may help for the choice of an operator. It concerns classification from dual echo brain MRI (figure 1). These images provide different contrast between tissues. In the first echo, the cerebrospinal fluid (CSF) in the ventricles and sulci have the highest grey values whereas they have the lowest grey values in the second echo, with a very low contrast to the white matter. Membership functions to the three main classes in the brain (ventricles + CSF, grey matter, white matter) are simply derived from grey values (figure 2). Results are shown on figure 3 for the ventricles (computations are performed within a mask corresponding to the brain). It appears that the posterior part of the ventricles have high membership degrees in the first image; in the second one, parts of the white matter have also quite high membership values, and the anterior part of the ventricles (cornu anterii, CA) are not well delineated. The membership functions resulting from the fusion for T-norms min and product, for T-conorms max and algebraic sum, for arithmetical mean and compensatory operator $\min^{\gamma} \max^{1-\gamma}$ for $\gamma = 0.5$ are shown respectively on figures 4, 5, 6 for the ventricle class. It appears that the posterior part of the ventricles are better differentiated from the other classes with T-norms, whereas the CA have higher membership values with algebraic sum or compensatory operator (see zoom on figure 8). The decision taken by rule 1 leads to the classification shown on figure 7. If we look at the CA, it is clear that this part is very difficult to classify correctly. When combining the membership functions of all classes with the same operator, the CA are classifiied as white matter (see the example of min on the left of figure 7, analogous situations occur with other operators). On the contrary, if the operator can be adapted to the classes, better results can be obtained for the CA, as it appears on the left of figure 7, where the algebraic sum has been used for the ventricles and the min for the other classes (see zoom on figure8).

This proves the interest of CD operators. Of course, a lot of work needs to be done in order to derive the adequate operator automatically.

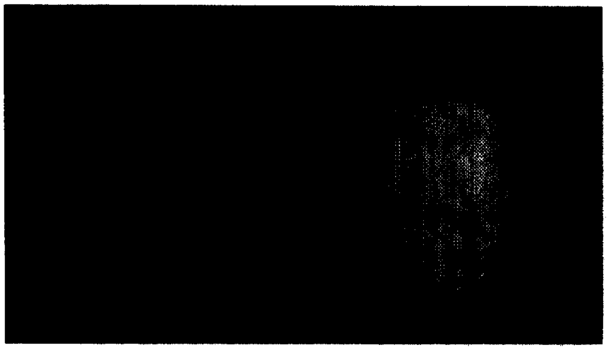

Figure 1: Brain MRI images (2 echos).

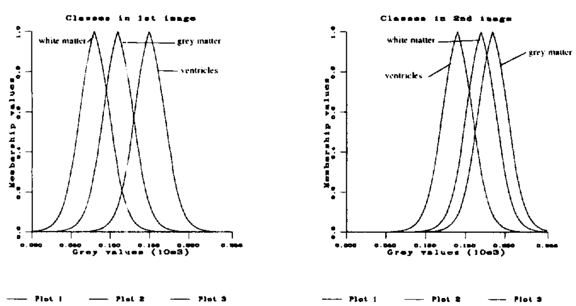

Figure 2: Membership functions to the 3 classes on each image.

Figure 3: Membership function to the class ventricles on each image.

Figure 4: Fusion with T-norms min and product (ventricle class).

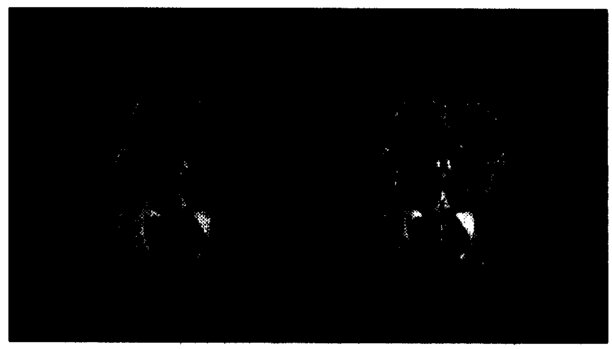

Figure 5: Fusion with T-conorms max and algebraic sum (ventricle class).

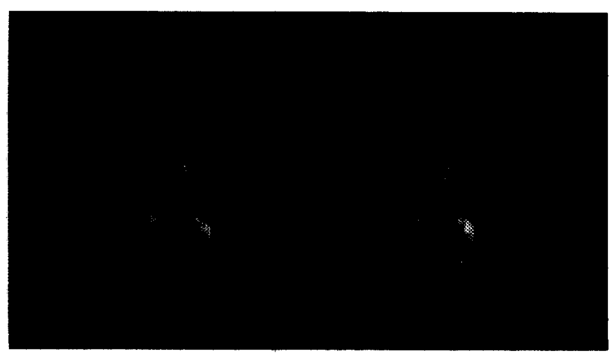

Figure 6: Fusion with arithmetical mean and compensatory operator (ventricle class).

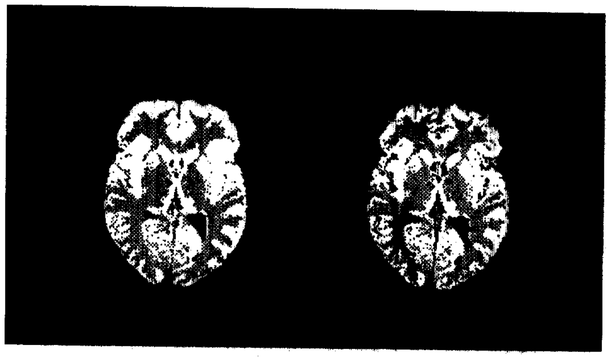

Figure 7: Classification obtained after decision with rule 1 for the min operator (left) and for the algebraic sum for the ventricles and the min for the other classes (right).

5. CONCLUSION

We proposed in this paper a scheme for fuzzy classification for multi-modality images, whose main advantages are:

- the use of fuzzy sets to represent imprecision inherent to the images and to classification problems,
- the rejection of the decision step at the end of the process, which avoids to take premature decision on each image separately (which often leads to conflicts difficult to manage without any additional information).

Then we proposed a classification of the fusion operators and discussed the aspects related to the choice of an operator by given criteria related to decision making. To our opinion, fuzzy operators deserve to be more developed, in particular CD operators, which are still in their infancy but are well adpated to classification problems from heterogeneous sources.

6. REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New-York, 1981.

[2] I. Bloch, *Fuzzy Sets in Image Processing*, ACM Symposium on Applied Computing, Phoenix, Arizona, March 6-8, 175-179, 1994.

[3] I. Bloch, H. Maître, *Fuzzy Mathematical Morphology*, Annals of Mathematics and Artificial Intelligence, Vol. 10, 55-84, 1994.

[4] I. Bloch, *Information Combination Operators for Data Fusion: A Comparative Review with Classification*, Technical Report Télécom Paris 94D013, April 1994.

[5] S. Deveughele, B. Dubuisson, *Using Possibility Theory in Perception : An Application in Artificial Vision*, Second IEEE Int. Conf. on Fuzzy Systems, San Fransisco, California, 821-826, March 28 - April 1, 1993.

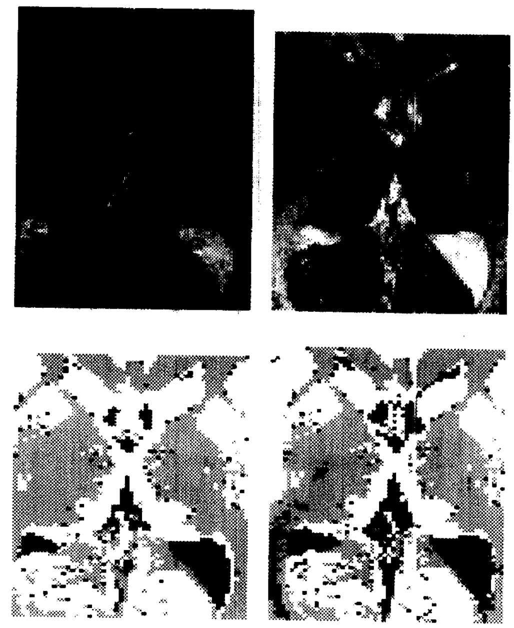

Figure 8: Zoom of figures 4 left, 5 right, 7.

[6] D. Dubois, H. Prade, *A Review of Fuzzy Set Aggregation Connectives*, Information Sciences 36, 85-121, 1985.

[7] D. Dubois, H. Prade, *Combination of Information in the Framework of Possibility Theory*, in Data Fusion in Robotics and Machine Intelligence, M Al Abidi et al. eds, Academic Press, 1992.

[8] M. Grabisch, *On the Use of Fuzzy Integrals as a Fuzzy Connective*, Second IEEE Int. Conf. on Fuzzy Systems, San Fransisco, California, 213-218, March 28 - April 1, 1993.

[9] K. Menger, *Statistical Metrics*, Proc. Nat. Acad. Sci. USA, 28, 535-537, 1942.

[10] A. Rosenfeld, *The Fuzzy Geometry of Image Subsets*, Pattern Recognition Letters 2, 311-317, 1984.

[11] B. Schweizer, A. Sklar, *Probabilistic Metric Spaces*, North Holland, Amsterdam, 1983.

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

[13] E. H. Shortliffe, B. G. Buchanan, *A model of Inexact Reasoning in Medicine*, Mathematical Biosciences 23, 351-379, 1975.

[14] R. R. Yager, *On Ordered Weighted Averaging Aggregation Operators in Multi Criteria Decision Making*, IEEE Trans. SMC, Vol. 18, No-1, 183-190, 1988.

[15] R. R. Yager, *Connectives and Quantifiers in Fuzzy Sets*, Fuzzy Sets and Systems 40, 39-75, 1991.

[16] L. A. Zadeh, *Fuzzy Sets*, Inform. and Control 8, 338-353, 1965.