Information Combination Operators for Data Fusion: A Comparative Review with Classification

Isabelle Bloch

***Abstract*— In most data fusion systems, the information extracted from each sensor (either numerical or symbolic) is represented as a degree of belief in an event with real values, taking in this way into account the imprecise, uncertain, and incomplete nature of the information. The combination of such degrees of belief is performed through numerical fusion operators. A very large variety of such operators has been proposed in the literature. We propose in this paper a classification of these operators issued from the different data fusion theories w.r.t. to their behavior. Three classes are thus defined. This classification provides a guide for choosing an operator in a given problem. This choice can then be refined from the desired properties of the operators, from their decisiveness, and by examining how they deal with conflictual situations.**

I. Introduction

IN most data fusion systems, the information extracted from images or sensors is represented as measures of belief in an event. This information can be either of numerical or of symbolic nature. Its representation as numerical degrees leads to a quantification of its characteristics (uncertain, imprecise, incomplete), which have to be taken into account in a fusion process. These characteristics are of particular importance for data fusion problems in image processing since they are often intrinsic to the images themselves (due to observed phenomena, acquisition device, numerical reconstruction algorithms, post-processing, etc.). Indeed, one of the main tasks of data fusion is to combine information issued from several sources to take a better decision than from one source only, by reducing imprecision and uncertainty and increasing completeness [3].

The events to which degrees of belief are assigned are related to the decision problem at hand. For instance, they may be the presence or absence of a given structure (say a road in a satellite image, a tumor in a medical image), the membership of a point or a set to a class, the detection of an object, etc.

The degrees of belief take generally their values in a real closed interval ([0, 1], $[-1, 1]\cdots$) and are modelized in different ways, depending on the chosen mathematical framework. They are probabilities in data fusion methods based on probability and Bayesian theory, membership degrees to a fuzzy set in fuzzy set theory, possibility distributions, possibility or necessity functions in possibility theory, certainty factors in MYCIN-like systems, mass, belief, or plausibility functions in Dempster-Shafer evidence theory.

Manuscript received June 6, 1994; revised January 19, 1995.
The author is with Département Images, Ecole Nationale Supérieure des Télécommunications, 75634 Paris Cedex 13, France.
Publisher Item Identifier S 1083-4427(96)00059-8.

When pieces of information issued from several sources have to be combined, each of them represented as a degree of belief in a given event, these degrees are combined in the form $F(x_1, x_2, \ldots, x_n)$, where x_i denotes the representation of information issued from sensor i. Then the question is: what information combination operator F should be chosen? A first restriction on F is the closure constraint imposed by the chosen mathematical framework; for instance, when combining probabilities it is desirable (for further combination and comparison purpose) to obtain another probability, i.e., to remain in the same framework. However, this constraint is not very restrictive when working with fuzzy sets. Thus other guidelines are needed.

In this paper, we propose a classification of data fusion operators with respect to their behavior in terms of severity or indulgence, and to the dependence of this behavior w.r.t. to the information to be combined. Section II is dedicated to the description of this classification. In Section III, a review of the commonly used operators (in different data fusion theories) is proposed w.r.t. this classification. In Section IV, the operators are revisited under the light of their properties and their interpretation in terms of data fusion, of their behavior when conflictual situations occur, and of their decisiveness.

II. Classification of Operators Depending on Their Behavior

This section aims at describing the possible behaviors that a fusion operator may have. Using the common sense qualifications of "severe", "indulgent", or "cautious", we propose a classification of the operators in three classes.

We will focus on operators combining two pieces of information. Extension to the combination of three or more will be addressed in Section IV-C.

Let x and y denote two real variables representing the degrees of belief to be combined; they take values into the interval I, which for the sake of simplicity is chosen as [0, 1][1].

Let us consider a function F acting on x and y, defining a combination or fusion operator. Under closure property assumption, $F(x, y)$ also has values in I.

According to the definitions given in [5], [19] for fuzzy operators, we say, for any fusion operator F, that:

- F is conjunctive if $F(x, y) \leq \min(x, y)$ (this corresponds to a severe behavior),

[1] In the case of MYCIN, this interval is $[-1, 1]$.

- F is disjunctive if $F(x,y) \geq \max(x,y)$ (indulgent behavior),
- F behaves like a compromise if $x \leq F(x,y) \leq y$ if $x \leq y$, and $y \leq F(x,y) \leq x$ else (cautious behavior).

We propose now a classification to describe the operators not only as conjunctive or disjunctive ones but also in terms of their behavior with respect to the particular values of the information to be combined.

A. Context Independent Constant Behavior (CICB) Operators

The first class, that we call context independent constant behavior operators (CICB), is composed of operators which have the same behavior whatever the values of the information to combine, and which are computed without any contextual or external information (i.e., only from the values of the information to combine). More formally, a CICB operator F satisfies one of the three following properties, which are exclusive

$$\forall(x,y) \in I^2, F(x,y) \leq \min(x,y), \tag{1}$$

$$\forall(x,y) \in I^2, F(x,y) \geq \max(x,y), \tag{2}$$

$$\forall(x,y) \in I^2, \min(x,y) \leq F(x,y) \leq \max(x,y). \tag{3}$$

B. Context Independent Variable Behavior (CIVB) Operators

The second class is composed of operators which are context independent like in the first class but whose behavior depends on the values of x and y. We call the operators of this class context independent variable behavior (CIVB) operators. For instance, an operator in this class may be severe if both pieces of information are high and indulgent if they are both low (imagine for example the problem of student notation $\cdots$).

C. Context Dependent (CD) Operators

The third class is composed of operators which are context dependent, i.e., which are computed not only from x and y 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 build operators which behave in a conjunctive way if the sources are consonant, in a disjunctive way if they are dissonant, and like a compromise if they are partly conflicting.

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.

III. A Review of Commonly Used Fusion Operators w.r.t. the Proposed Classification

In this section, we will briefly recall the definitions of the main fusion operators used in the different numerical data fusion theories (probability and Bayesian inference, fuzzy sets, possibility theory, MYCIN-like systems, Dempster-Shafer evidence theory), and show that they all fit in the proposed classification.

A. CICB Operators

Examples of operators belonging to this class can be found in several mathematical frameworks (probability and Bayesian fusion, fuzzy sets and possibility theory, Dempster-Shafer evidence theory).

1) Probabilistic and Bayesian Fusion: Let us first consider probabilistic and Bayesian fusion. The degrees of belief are represented by probabilities (a priori, conditional and a posteriori probabilities). Decisions are usually taken from a posteriori probability. Let E be the event to be evaluated, and x_1, x_2 the pieces of information provided by the two sensors. From Bayes theorem, we have

$$\begin{aligned} p(E \mid x_1, x_2) &= \frac{p(x_2 \mid E, x_1)p(E \mid x_1)}{p(x_2 \mid x_1)} \\ &= \frac{p(x_2 \mid E, x_1)p(x_1 \mid E)p(E)}{p(x_2 \mid x_1)p(x_1)} \\ &= p(x_2 \mid E, x_1)p(x_1 \mid E)\frac{p(E)}{p(x_1, x_2)} \\ &= p(x_1 \mid E, x_2)p(x_2 \mid E)\frac{p(E)}{p(x_1, x_2)}. \end{aligned} \tag{4}$$

From this equation, it is clear that the operator involved in the combination is a product of probabilities, which is conjunctive. The term $p(x_1, x_2)$ is a normalization term which is constant for all events (it does not depend on a contextual information) and does not have any influence on the behavior of the operator. Thus Bayesian fusion involves a CICB operator. Note that the above expression reduces to

$$\frac{p(x_2 \mid E)p(x_1 \mid E)p(E)}{p(x_1)p(x_2)} \tag{5}$$

in case of independence between sources.

2) Fuzzy Sets and Possibility Theory: Let us now consider fuzzy sets and possibility theory. Three families of operators used in these theories are CICB: triangular norms (T-norms), triangular conorms (T-conorms) and mean operators.

In the context of stochastic geometry ([10], [11]), a T-norm i is defined as a function of two variables from $[0,1] \times [0,1]$ to $[0,1]$ satisfying several properties: commutativity, associativity, 1 is unit element, increasingness with respect to the two variables. From these properties, limit conditions can be derived: $i(0,1) = i(0,0) = i(1,0) = 0$ and $i(1,1) = 1$, and it is easily shown that 0 is null element ($\forall x \in [0,1], i(x,0) = 0$). A continuity property is often added to these properties.

T-norms generalize intersection to fuzzy sets ([5], [6], [19]). Examples of T-norms are $\min(x,y)$, xy, $\max(0, x+y-1)$.

It is easy to prove the following result: for any T-norm i, the following inequality holds

$$\forall(x,y) \in [0,1]^2, i(x,y) \leq \min(x,y). \tag{6}$$

This shows that the "min" is the greatest T-norm and that any T-norm has a conjunctive behavior, whatever the values of x and y. Therefore, T-norms are CICB operators.

In this framework, a T-conorm is defined as an operation u from $[0,1] \times [0,1]$ to $[0,1]$ such that u is commutative, associative, monotonic, admits 0 as unit element. It verifies limit conditions ($u(0,1) = u(1,1) = u(1,0) = 1$ and

$u(0,0) = 0$), and admits 1 as null element. T-conorms generalize union to fuzzy sets. Examples of T-conorms are $\max(x,y)$, $x+y-xy$, $\min(1,x+y)$. For any T-conorm u, the following inequality holds

$$\forall (x,y) \in [0,1]^2, u(x,y) \geq \max(x,y). \quad (7)$$

This shows that the "max" is the smallest T-conorm and that any T-conorm has a disjunctive behavior, whatever the values of x and y. Therefore, T-conorms are CICB operators.

A mean operator m is defined as a function from $[0,1] \times [0,1]$ in $[0,1]$ such that ([5], [19]) $\min(x,y) \leq m(x,y) \leq \max(x,y)$, $m \neq \min$, $m \neq \max$, $m(x,y) = m(y,x)$, m is increasing w.r.t. both arguments.

Examples of mean operators are:

- the median operators $m_\alpha(x,y) = \text{med}_\alpha(x,y,\alpha)$[2],
- the bisymmetrical[3] continuous strictly monotonous means which have the general form $m(x,y) = k^{-1}[\frac{k(x)+k(y)}{2}]$, with k continuous strictly monotonous (examples in this family are the harmonic mean $\frac{2xy}{x+y}$, the geometrical mean $\sqrt{xy}$, the arithmetical mean $\frac{x+y}{2}$, the quadratic mean $\sqrt{\frac{x^2+y^2}{2}}$),
- the ordered weighted averaging operators [18],
- the fuzzy integrals ([8], [15]).

The behavior of these operators is directly issued from the definition, which shows that they behave always like a compromise. Therefore, mean operators are CICB.

3) Dempster-Shafer Evidence Theory: Let us now consider the Dempster-Shafer evidence theory. In this framework, the information from sensor i is represented by a mass function m_i assigning values in [0, 1] to each subset of the discernment set D (the set of considered events). From the set of mass values for all subsets of D, other functions can be derived (belief, plausibility, communality, doubt). The key point is that the set of values taken by one of these functions can be derived from the set of values given by any other one ([9], [12]).

The combination is performed by the orthogonal sum of Dempster, expressed for n sources as

$$\oplus_{i=1}^{n} m_i(A) = \frac{1}{1-k} \sum_{B_1 \cap B_2 \cdots \cap B_n = A} m_1(B_1) m_2(B_2) \cdots m_n(B_n) \quad (8)$$

where $A, B_1, \ldots, B_n$ are subsets of D, and

$$k = \sum_{B_1 \cap B_2 \cdots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \cdots m_n(B_n). \quad (9)$$

From this expression, the behavior of the orthogonal sum is not very clear. However, let us consider the communality function defined as: $\text{Com}_i(A) = \sum_{A \subset B} m_i(B)$. It can be shown ([9], [12]) that the communality functions are combined using the rule

$$\oplus_{i=1}^{n} \text{Com}_i(A) = \frac{1}{1-k} \text{Com}_1(A) \text{Com}_2(A) \cdots \text{Com}_n(A). \quad (10)$$

Thus, the result on subset A depends only on the communality values given by all sensors for subset A, and the operators is a product, i.e., a conjunctive operator. Here again, like for Bayesian fusion, $1-k$ is a normalization factor (assuring that the result of the combination is a mass function, respectively a communality function) which is constant for all subsets and thus does not influence the behavior of the operator. Therefore, the orthogonal sum of Dempster-Shafer is a CICB operator.

The normalization factor is often interpreted as the conflict since k represents the weight of evidence which would be assigned to the empty set in the absence of normalization. It is interesting to note that k can be computed only from the conflict between two sensors. Let us note k_{12} the conflict between sources 1 and 2, $k_{(12)3}$ the conflict between 3 and the result of the combination of 1 and 2, etc. Then it is easy to show that

$$1 - k = (1 - k_{12})(1 - k_{(12)3})(1 - k_{(1 \cdots n-1)n}). \quad (11)$$

This equation is much more than an efficient way to compute the normalization factor. It shows additionally that $1-k$ is also combined in a conjunctive way (and therefore the conflict has a disjunctive behavior: it increases when the number of sources increases).

B. CIVB Operators

Examples of operators belonging to this class can be found in fuzzy sets and possibility theory, and in MYCIN-like systems.

1) Fuzzy Sets and Possibility Theory: Symmetrical sums are defined as operators σ from $[0,1] \times [0,1]$ in $[0,1]$ such that ([5], [19]): $\sigma(0,0) = 0$, σ is commutative, σ is increasing w.r.t. both arguments, σ is continuous, σ is auto-dual[4]. Moreover, $\sigma(1,1) = 1$.

The general form of symmetrical sums is given by

$$\sigma(x,y) = \frac{g(x,y)}{g(x,y) + g(1-x, 1-y)} \quad (12)$$

with g increasing, positive, continuous, and such that $g(0,0) = 0$. The only symmetrical sum σ which is associative and a mean operator is $\text{med}(x,y,1/2)$. The associative strictly increasing symmetrical sums take the form

$$\sigma(x,y) = \psi^{-1}[\psi(x) + \psi(y)] \quad (13)$$

with ψ strictly monotonous, $\psi(0)$ and $\psi(1)$ nonbounded, $\psi(1-x) + \psi(x) = 0$.

Examples of symmetrical sums are:

- $\sigma_0(x,y) = \frac{xy}{1-x-y+2xy}$, obtained for $g(x,y) = xy$ (it is associative),
- $\sigma_+(x,y) = \frac{x+y-xy}{1+x+y-2xy}$, obtained for $g(x,y) = x+y-xy$ (it is not associative),

[2] Median operators are the only associative means and are defined by: $\text{med}(x,y,\alpha) = x$ if $y < x < \alpha$ or $\alpha < x < y$, $\text{med}(x,y,\alpha) = y$ if $x < y < \alpha$ or $\alpha < y < x$, $\text{med}(x,y,\alpha) = \alpha$ else.

[3] A bisymmetrical operator $m(x,y)$ is such that $m[m(x,y), m(z,t)] = m[m(x,z), m(y,t)]$.

[4] An operator is auto-dual with respect to the complementation to 1 iff $1 - \sigma(x,y) = \sigma(1-x, 1-y)$. This definition is easily extended to any complementation.

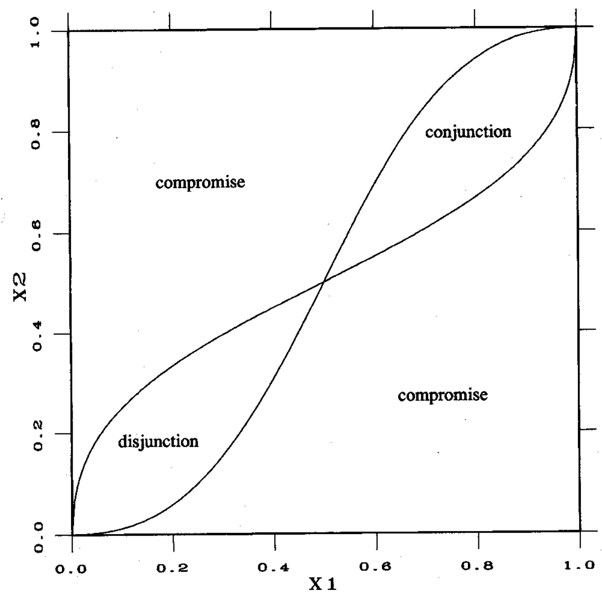


Fig. 1. Behavior of σ_+ (nonassociative symmetrical sum) w.r.t. the relative values of x and y in the $[0,1] \times [0,1]$ square.

- $\sigma_{\min}(x,y) = \frac{\min(x,y)}{1-|x-y|}$, obtained for $g(x,y) = \min(x,y)$ (it is a mean),
- $\sigma_{\max}(x,y) = \frac{\max(x,y)}{1+|x-y|}$, obtained for $g(x,y) = \max(x,y)$ (it is a mean).

Symmetrical sums represent "hybrid" aggregation operators. Let us take the example of $\mathrm{med}(x,y,1/2)$: it is an indulgent operator (with disjunctive behavior) if x and y are low (less than 1/2) resulting in $\max(x,y)$, it is severe (conjunctive behavior) if x and y are high (greater than 1/2) resulting in $\min(x,y)$, and its return value is 1/2 if either x or y is less than 1/2, the other being greater, and in this case it has a compromise behavior. In fact, median operators appear as a limit between CICB and CIVB operators.

More generally, any associative symmetrical sum (but median) has the following behavior [6]:

- conjunctive if $\max(x,y) < 1/2 : \sigma(x,y) \leq \min(x,y)$,
- disjunctive if $\min(x,y) > 1/2 : \sigma(x,y) \geq \max(x,y)$,
- compromise if $x \leq 1/2 \leq y : x \leq \sigma(x,y) \leq y$ and the reverse in+equality holds if $y \leq 1/2 \leq x$.

This illustrates the variable behavior of σ, depending on the values to be combined. Thus such operators are CIVB.

For nonassociative symmetrical sums[5], the three kinds of behavior also occur, but the rules may be more complex. Fig. 1 illustrates the domains where σ_+ is a disjunction, a conjunction or a compromise.

2) MYCIN-Like Systems: Let us now consider systems which use the rules proposed in MYCIN for combining certainty factors [13]. Here, the degrees of belief take their values in $I = [-1,1]$. They do not represent absolute degrees of belief, like in other theories, but rather the amount of belief or disbelief provided by an information to update the belief in an event. These so called "certainty factors" are positive if the information confirms the event, and negative if the information disconfirms the event (the value 0 means that the information says nothing about this event). The rules defined in [13] for combining two pieces of information x and y supporting the same event are:

- $x + y - xy$ if $x \geq 0$ and $y \geq 0$ (i.e., both pieces of information confirm the event),
- $x + y + xy$ if $x \leq 0$ and $y \leq 0$ (i.e., both disconfirm the event),
- $x + y$ if $x \leq 0$ and $y \geq 0$, or $x \geq 0$ and $y \leq 0$ (i.e., one information confirms the event and the other disconfirms it).

It can be easily shown that, $\forall (x,y) \in [-1,1]^2$

$$x \geq 0 \text{ and } y \geq 0 \Rightarrow x + y - xy \geq \max(x,y),$$
$$x \leq 0 \text{ and } y \leq 0 \Rightarrow x + y + xy \leq \min(x,y),$$
$$xy \leq 0 \Rightarrow \min(x,y) \leq x + y \leq \max(x,y).$$

This proves that the MYCIN operator is CIVB and behaves as follows:

- in a disjunctive mode if both x and y are positive, providing a result which confirms the event more than each individual information,
- in a conjunctive mode if both x and y are negative, resulting in a stronger disconfirmation than each individual information,
- like a compromise if one of x and y is negative, the other one being positive; the sign of the result depends on the strength of disconfirmation (respectively, confirmation) of each individual information.

C. CD Operators

This section presents some context dependent operators taking into account some contextual information about the sources. Two examples have been chosen to illustrate the behavior of a CD operator: conflict between the sources and reliability of the sources. This is of course not exhaustive. In particular, as it is well known in image processing, information about spatial context can also be considered and included in the operators[6].

Some operators have been proposed in possibility theory to take into account contextual information [7] and can also be used in fuzzy set theory for combining membership degrees. They allow to take into account several situations.

- The sources may be conflicting when they give information about one event (a class) and consonant when considering another class.
- The sources may have different global reliability.
- A source may be reliable when giving an opinion about one class and not reliable for another class.

1) Dependence on Conflict: Let us first consider operators depending on a measure of conflict. In [7], operators are

[5] The case of nonassociative symmetrical sums which are also mean operators has been treated in Section III-A, since they are CICB.

[6] It should be noted that some methods in image processing which take into account spatial contextual information, like Markov random fields, include this information mainly in the initial probability assignment, and not directly in the fusion operators. Such methods are thus beyond the scope of this paper.

proposed which assume that one source is reliable but we do not know which one. Thus, operators are designed in order that

- they are conjunctive if the sources are consonant (i.e., low conflict): in this case, the sources are necessarily both reliable, and thus the operator can be severe;
- they are disjunctive if the sources are dissonant (i.e., high conflict): a disjunction will automatically favor the reliable source as the one which gives the higher degree of confidence;
- they behave like a compromise in case of partial conflict: this case being the most problematic, the operators are "cautious".

In the following equations, expressing these operators, π_1 and π_2 represent two distributions of possibility to be combined in a global distribution π', $1 - h(\pi_1, \pi_2)$ represents a global measure of conflict between these two distributions, and i denotes a T-norm

$$\pi'(s) = \max\left[\frac{i[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)}, 1 - h(\pi_1, \pi_2)\right], \tag{14}$$

$$\pi'(s) = \min\left[1, \frac{i[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)} + 1 - h(\pi_1, \pi_2)\right], \tag{15}$$

$$\pi'(s) = i[\pi_1(s), \pi_2(s)] + 1 - h(\pi_1, \pi_2), \tag{16}$$

$$\pi'(s) = \max\left[\frac{\min(\pi_1, \pi_2)}{h}, \min[\max(\pi_1, \pi_2), 1 - h]\right]. \tag{17}$$

From these equations we conclude that these operators are CD.

In [7], the proposed measure of conflict is defined as $1 - h$ with

$$h = \sup_s i[\pi_1(s), \pi_2(s)]. \tag{18}$$

This measure, which assures that the resulting distribution π' is normalized, is well adapted for trapeze-shaped possibility distributions.

If we want to apply these operators with membership functions to fuzzy sets, for instance for multi-image classification problems, other measures of conflict may be used, like distances between fuzzy sets μ_i^1 and μ_i^2 representing the same class i in two different images 1 and 2. This allows to combine the information related to each class in a way which is adapted to the conflict between the sources concerning this class, as

$$\forall M \in \text{Image}, \mu_i(M) = F[\mu_i^1(M), \mu_i^2(M), \text{conflict}(\mu_i^1, \mu_i^2)]. \tag{19}$$

Such a form allows one to design CD operators well adapted for classification problems in data fusion, where the sources behave differently w.r.t. each class and differently from each others.

2) Dependence on Source Reliability: Let us now consider operators depending on the reliability of the sources. In possibility and fuzzy set theories, several operators can be built, depending on the reliability of the sources. Different situations can be considered [7].

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

TABLE I
COMPARISON OF DATA FUSION OPERATORS

Operator classification based upon behaviour w.r.t. data to be combined	Mathematical combination theory (framework)	Sec.	Specific mathematical operators	Behaviour of specific operator
CICB (Context Independent Constant Behaviour)	Bayesian	III-A.1	Product of probabilities	conjunctive
	Fuzzy sets, possibility	III-A.2	T-norm T-conorm mean	conjunctive disjunctive compromise
	Dempster-Shafer	III-A.3	orthogonal sum	conjunctive
CIVB (Context Independent Variable Behaviour)	Fuzzy sets, possibility	III-B.1	symmetric sums	all three behaviours possible, depending upon values to be combined
	MYCIN-like systems	III-B.2	certainty factors combination	disjunctive (if x and y are positive) conjunctive (if x and y are negative) compromise (else)
CD (Context Dependent)	Fuzzy sets, possibility	III-C.1	operators depending on conflict	conjunctive if sources are consonant disjunctive if sources are dissonant compromise if partial conflict
	Fuzzy sets, possibility	III-C.2	operators depending on source reliability	

- A subset of sources is reliable, but we do not know which one.
- Only an order is known on the reliabilities of the sources, but no precise values, such that priorities are defined on the sources.

As with conflict-dependent operators, these operators are CD and are adaptive. For the classification problem mentioned above, operators can be designed in the same way, taking into account possibly different reliabilities of the sources w.r.t. different classes (sources may report careless for some classes but not for other ones). For instance, when combining membership degrees, such an operator can take the form

$$\begin{aligned}\mu_i(M) = F[\mu_i^1(M), \mu_i^2(M), &\text{reliability}(\text{source } 1 \mid \text{class i}),\\ &\text{reliability}(\text{source } 2 \mid \text{class i}),\\ &\text{global-reliability}(\text{source } 1),\\ &\text{global-reliability}(\text{source } 2)].\end{aligned} \tag{20}$$

D. Summary

Table I summarizes the results presented in Sections III-A, III-B, and III-C.

IV. Further Properties

This section aims at giving criteria for refining the choice of an operator in each of the three classes described above. Since most of the properties of the operators are well known, we will mainly stress on the way they are related to data fusion and decision problem.

A. Ordering Operators

A first way for differentiating the operators consists in examining how they are sorted with respect to a partial ordering. Two operators F and F' are said ordered independently of the values of x and y if we have either $\forall(x, y) \in I^2, F(x, y) \leq F'(x, y)$, or $\forall(x, y) \in I^2, F(x, y) \geq F'(x, y)$. Ordered operators can be represented on the [0, 1] axis. Fig. 2 presents examples of operators of the three classes with their respective positions.

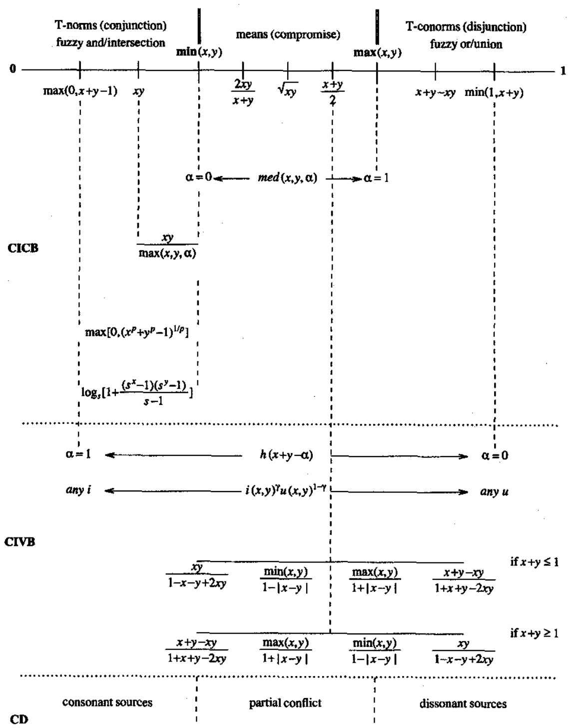


Fig. 2. Ordering operators.

Let us consider the examples of fuzzy CICB operators described in Section III-A. They are ordered in the following way:

- T-norms: $\forall(x,y) \in [0,1]^2, \max(0, x+y-1) \leq xy \leq \min(x,y)$;
- mean operators (in relation to "min" and "max"): $\forall(x,y) \in [0,1]^2, \min(x,y) \leq \frac{2xy}{x+y} \leq \sqrt{xy} \leq \frac{x+y}{2} \leq \sqrt{\frac{x^2+y^2}{2}} \leq \max(x,y)$;
- T-conorms: $\forall(x,y) \in [0,1]^2, \max(x,y) \leq x+y-xy \leq \min(1, x+y)$.

There exists a lot of parameterized families of operators which range continuously from one operator to another[7]. For instance, such families of T-norms (respectively, T-conorms) are summarized in [5], [6], [17], [19]. They are ordered w.r.t. the parameter they depend on. Some examples are as follows:

- $\frac{xy}{\max(x,y,\alpha)}$ [11] which ranges from product xy for $\alpha = 1$ to $\min(x,y)$ for $\alpha = 0$ (this family is decreasing w.r.t. the parameter α);
- $\max[0, (x^p + y^p - 1)]^{1/p}$ [6], which ranges from $\max(0, x+y-1)$ for $p = 1$ to the product for $p = 0$ [6];
- $1 - \min[1, (1-x)^p + (1-y)^p]^{1/p}$ [19], which ranges from $\max(0, x+y-1)$ for $p = 1$ to the "min" for $p = +\infty$;
- $\frac{xy}{\gamma + (1-\gamma)(x+y-xy)}$ (family of Hamacher);
- $\log_s[1 + \frac{(s^x-1)(s^y-1)}{s-1}]$ (family of Franck), which is a decreasing family w.r.t. the parameter s: in particular, it is equal to the "min" for $s = 0$, to the product for $s = 1$ and to $\max(0, x+y-1)$ for $s = +\infty$.

In a similar way for mean operators, $\mathrm{med}(x,y,\alpha)$ ranges from $\min(x,y)$ for $\alpha = 0$ to $\max(x,y)$ for $\alpha = 1$ (it is an increasing family of mean operators).

The parameter involved in such an operator allows one to define precisely where the operator is located with respect to others (see Fig. 2). For instance, the choice of a small value of α in the median operator makes it behave in a way closed to the behavior of "min". Choosing a greater value for α leads to an operator which is always greater than the previous one, whatever the values of the pieces of information to be combined. If α goes to 1, the operator is to behave like the "max".

For other parameterized operators, the class they belong to depends on the parameter, and they are also ordered when the parameter varies. Two examples are as follows:

- the function $h(x+y-\alpha)$ (with $h(t) = 0$ if $t \leq 0$, $h(t) = t$ if $0 \leq t \leq 1$, and $h(t) = 1$ if $t \geq 1$) ranges from $\max(0, x+y-1)$ for $\alpha = 1$ to $\min(1, x+y)$ for $\alpha = 0$; for these extreme values of α, the operator is CICB (conjunctive for $\alpha = 1$, disjunctive for $\alpha = 0$), else it is generally CIVB;
- i and u being any T-norm and T-conorm, the operator defined as $i(x,y)^\gamma u(x,y)^{1-\gamma}$ ranges from i for $\gamma = 1$ to u for $\gamma = 0$ and is in these both cases CICB, else it is CIVB.

Let us now consider the four examples of symmetrical sums given in Section III-B. They are ordered differently depending on the values to be combined: if $x + y \leq 1$,

$$\frac{xy}{1-x-y+2xy} \leq \frac{\min(x,y)}{1-|x-y|} \leq \frac{\max(x,y)}{1+|x-y|} \leq \frac{x+y-xy}{1+x+y-2xy},$$

and if $x + y \geq 1$,

$$\frac{xy}{1-x-y+2xy} \geq \frac{\min(x,y)}{1-|x-y|} \geq \frac{\max(x,y)}{1+|x-y|} \geq \frac{x+y-xy}{1+x+y-2xy}.$$

B. Interpretation of Properties in Terms of Data Fusion

Let us now try to interpret the various elements involved in numerical data fusion operators w.r.t. the processing of uncertain, imprecise, or ambiguous information.

At first, the elements of I are interpreted as measures derived from data given by one or several sensors, for instance a measure of membership to a class, a measure of evidence of presence, a measure of satisfaction of a criterion. Their mathematical form depends on the considered theoretical framework. The extreme values of I (say 0 and 1) play particular roles. The value 0 means that, for an event, the sensor provides a null measure, either because it considers the event as impossible, or because it has no information

[7]It should be noted that, depending on the value of the parameter, the operators may have different behaviors. For instance on Fig. 2, the median operator is a compromise for $0 < \alpha < 1$, a T-norm for $\alpha = 0$ and a T-conorm for $\alpha = 1$. In the same way, they may not always belong to the same class. For instance the operator $h(x,y,\alpha)$ (see formal definition in the text) is CIVB or CICB depending on α.

or a complete ignorance about the event. On the contrary, the value 1 means that the sensor considers the information, or the event, as sure and thus represents a total certainty. Values which are strictly between 0 and 1 represent degrees of uncertainty or partial knowledge on the information. They can also be interpreted as imprecision, or as quantity of information available about the event.

A conjunctive operator represents a consensus between pieces of information, or its common or redundant part. It reduces the less certain information and has at most confidence in the sensor which gives the smallest measure. It searches for a simultaneous satisfaction of criteria or objectives. On the contrary, a disjunctive operator increases the certainty we have about an information and has at least confidence in the sensor which gives the greatest or the most certain measure, or the most information. It expresses redundancy between criteria. A compromise operator provides a global measure, intermediate between the partial measures provided by each sensor [6]. Redundancy between pieces of information can be taken into account by compromise operators like fuzzy integrals [8]. By introducing weights on subsets of information, they are able to include interactions and dependencies between pieces of information, thus avoiding combination bias.

Commutativity and associativity properties express that the result of the combination is independent of the order in which the information is combined. Commutativity is commonly satisfied by information combination operators. Associativity is generally not satisfied by mean operators nor by symmetrical sums. For the other examples given in Section II, this property holds. These properties are even often imposed as axioms governing the construction of operators, as they are commonly recognized as minimal properties the operators should satisfy, although human reasoning not always combines information in a commutative and associative way. We may take as a counterexample of commutativity the task done by an aerial image photo-interpreter. Having two or more images at his disposal to analyse a scene, he generally works with one first, and, after a first basic interpretation with this image alone, upgrades his interpretation with complementary information extracted from the others.

Unit element may exist or not. If it exists (for instance for T-norms, T-conorms, Dempster rule, MYCIN), this means that, if a source provides such an answer, this answer combined with any other information will not modify a decision taken from this information alone. For T-norms, the unit element is 1. The combination of such an information with any other one well matches the idea of a consensus between a certainty about an event and another measure of this event. For T-conorms, the unit element is 0, which corresponds to a complete ignorance of a sensor, or the absence of information, and thus has no influence on a disjunction operator. In MYCIN, the unit element is 0, that is an information which says nothing about the event and thus does not influence the combination. In Dempster rule, the unit element is defined as the mass function

$$m_e(D) = 1, \forall A \neq D, m_e(A) = 0. \tag{21}$$

m_e has only one focal element which is the whole frame of discernment D, and thus represents a total absence of information.

Increasingness corresponds to a constraint generally imposed on the operators: if two sensors provide pieces of information or measures x' and y' greater than x and y, respectively, we expect from the combination of x' and y' a result that is also greater than the result obtained from x and y (representing more information, or more certainty).

Limit conditions, which appear in most theories, govern the behavior of the combination of measures in $\{0, 1\}$ (or more generally extreme values of I), and impose it to be compatible with the binary case. Thus, their interpretation is the same as for classical logic, where the reasoning deals only with values "true" and "false".

Continuity property, satisfied for most operators, assures the robustness of the information combination. If a sensor provides an information or measure x' slightly different from x, the combination of x' with any other value should not be very different from the one obtained with x. This property is not always imposed. It is for example possible to impose that some values completely determine the result and that small changes in these values drastically change the result and the derived decision. Let us take the example of catching a train or not, depending on the moment you leave home and on the speed you walk. If you start later and run faster, the response will be yes (you catch the train, more or less easily), until a precise moment. If you go later than this moment, you will not catch the train, whatever your speed.

Existence of a null element for an operator means that this value completely determines the result of the combination. It is enough that a sensor provides this value for the result of any combination being this value. For T-norms, the null element is 0, which is consistent with the idea that a consensus cannot provide any information from a set of measures among them one is 0. For T-conorms, the null element is 1: if a sensor provides a total certainty about an event, its combination by disjunction with any other information will also be a total certainty. For evidence theory, the null element is defined as the mass function

$$m_a(\{d_i\}) = 1, \forall A \neq \{d_i\}, m_a(A) = 0, \tag{22}$$

where d_i is an element of D. Then combining any mass function m with m_a provides

$$\begin{aligned}\forall m, \forall A, (m \oplus m_a)(A) &= \frac{\sum_{\substack{B,C \\ A=B\cap C}} m(B)m_a(C)}{1-\sum_{\substack{B,C \\ B\cap C=\emptyset}} m(B)m_a(C)} \\ &= \frac{\sum_{\substack{B \\ A=B\cap\{d_i\}}} m(B)}{1-\sum_{\substack{B \\ B\cap\{d_i\}=\emptyset}} m(B)}.\end{aligned} \tag{23}$$

The B's which contribute to the numerator are those which contain d_i if $A = \{d_i\}$, or those which do not contain d_i if $A = \emptyset$ (in this case, the combination rule provides 0 by definition), else no B contributes to the sum. Thus, we have

$$\forall A \neq \{d_i\}, (m \oplus m_a)(A) = 0 \tag{24}$$

$$(m \oplus m_a)(\{d_i\}) = \frac{\sum_{d_i \in B} m(B)}{1-\sum_{d_i \notin B} m(B)} = 1. \tag{25}$$

Therefore, m_a is null element for the Dempster rule of combination. It represents a total certainty about one event (singleton).

Idempotence means that measuring again an already known information will not change the already derived deduction. This property is not necessarily imposed for data fusion. For instance the Dempster rule of combination is not idempotent, in opposition to mean operators; the only idempotent T-norms and T-conorms are min and max. We may want on the contrary that the combination of two (uncertain) identical data reinforces or weakens the global confidence in the considered event. This is formalized as the Archimedian property. For T-norms, it expresses that the confidence decreases if we have twice the same uncertain information. This behavior is the same as for probabilistic fusion where, when multiplying probabilities, probability decreases. On the contrary for T-conorms, the Archimedian property expresses that the confidence in an information is reinforced if this information occurs twice. The kind of stability expressed by compromise operators and their idempotence is incompatible with the Archimedian property. Let us take an example of concurrent testimony combination, issued from two witnesses who report the same fact. If the two witnesses are in connivance and I know this fact, then I will combine their reports in an idempotent way, since it is not surprising that they tell the same, and thus the report of one of them will not change the confidence I have in the report of the other one. On the contrary, if they are independent I will reinforce the confidence in the fact they report if I trust them, and weaken this confidence if I do not trust them, and thus the Archimedian property will be used.

Nilpotence means that the accumulation of n pieces of information leads to the null element. For instance for T-conorms, a total certainty about an event is gained if we obtain a sufficient number of non null measures supporting that event, even uncertain. The operators $i(x,y) = \max(0, x+y-1)$ and $u(x,y) = \min(1, x+y)$ are examples of nilpotent T-norm and T-conorm. In evidence theory, m_a is clearly a nilpotent element. Let us consider a simple support mass function m_s, defined, for given A and $s \in [0,1]$, by ([9], [12])

$$m_s(A) = s, m_s(D) = 1-s, \forall B \in 2^D - \{A, D\}, m_s(B) = 0. \tag{26}$$

It can be recursively shown that

$$\oplus_n m_s(A) = 1-(1-s)^n, \oplus_n m_s(D) = (1-s)^n \tag{27}$$
$$\forall B \in 2^D - \{A, D\}, \oplus_n m_s(B) = 0. \tag{28}$$

For $s \in [0,1]$, we obtain $\lim_{n\to\infty} \oplus_n m_s = m_1$, which corresponds to a certainty on A. This means that the accumulation of m_s provides an information more and more certain on A, and this behavior is very close to nilpotence, although not mathematically equivalent. The combination of m_s with any mass function m provides a null result but for the B's such that $B \subset A$: the focal elements of $m \oplus m_s$ are those which are included in A. This means that all information given by m which does not contain A is eliminated.

As an example, let us consider successive reports from several cascaded witnesses (i.e., each witness reports what the previous one told him), none of them being completely reliable. In this case, the combination of their reports decreases when the number of witnesses increases, and thus, a nilpotent operator may be used.

The rules of excluded middle and noncontradiction, which are satisfied for some operators but not for all, have an interpretation in terms of reasoning, in particular in the domain of artificial intelligence and approximate reasoning. They concern the combination of contrary pieces of information and are respectively expressed as

$$F[x, c(x)] = 0, \quad \text{(noncontradiction)} \tag{29}$$
$$F[x, c(x)] = 1, \quad \text{(excluded-middle)} \tag{30}$$

where c denotes a complementation (i.e., a negation of the information). These rules are not necessarily imposed and may or may not be in conflict with other properties. For instance, the satisfaction of these principles and the idempotence property are mutually exclusive. For T-norms and T-conorms, it could be expected that the noncontradiction rule is satisfied by T-norms (as intersection operators) and the excluded middle rule by T-conorms (as union operators). This is however not always true: for instance, the "min" and "max" do not satisfy these rules, nor the product and the algebraic sum. On the contrary, nilpotent operators satisfy these two principles. In probability theory, the additivity relation $P(A) + P(A^C) = 1$ expresses the excluded middle rule. This is no more satisfied in Dempster-Shafer theory, where belief functions (resp. plausibility functions) are sub-additive (resp. supra-additive).

Examples where excluded middle is not desirable occur in problems where we want to introduce ignorance about an event and its complementary (this is by the way one of the key features of Dempster-Shafer theory), and where we thus need to relax the exhaustivity assumption made for instance in probability theory.

C. Fusion of More Than Two Images

In this section, we consider more than two pieces of information to be combined, denoted by $(x_1, \ldots, x_n) \in I^n$, with $n \geq 2$, and combined in the form $F(x_1, \ldots, x_n) \in I$. The following considerations describe how far the previous results obtained for $n = 2$ can be extended to $n > 2$.

1) Generalizing Definitions: Let us take first the Bayesian combination rule. It can be expressed as

$$p(E \mid x_1, \ldots, x_n) = \frac{p(x_n \mid E, x_1, \ldots, x_{n-1}) \cdots p(x_1 \mid E)p(E)}{p(x_1, \ldots, x_n)} \tag{31}$$

and this expression does not depend on the order in which the pieces of information are provided. It reduces to

$$\frac{p(E) \prod_{i=1}^n p(x_i \mid E)}{\prod_{i=1}^n p(x_i)} \tag{32}$$

in case of independence between sources.

In an analogous way, the orthogonal combination rule of evidence theory is directly generalized (see Section III-A), as well as the computation of conflict.

In a more general way, defining an operator acting on n variables from one acting on two variables is straightforward as long as the operator is associative (all operators considered in this paper being commutative). This is the case for Bayesian combination and Dempster-Shafer theory as seen before, but also for MYCIN combination rules, for T-norms, T-conorms, median operators and some symmetrical sums (σ_0 for instance) used in fuzzy set and possibility theories. The combination is simply performed by successively adding pieces of information

$$F(x_1, \ldots, x_n) = F[F[\cdots F[F(x_1, x_2), x_3] \cdots], x_n], \quad (33)$$

and the result does not depend on the order in which the x_i's are combined.

For nonassociative operators (most of mean operators, symmetrical sums, and CD operators) two ways may lead to the combination.

- The first way consists in applying the above formula for a given order on the x_i's, chosen in an adequate way w.r.t. the problem at hand (for instance, the ordering may represent priorities on the pieces of information to be combined).
- 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 arithmetical mean can be generalized as

$$\frac{x_1 + x_2 + \cdots + x_n}{n}, \quad (34)$$

instead of

$$(\cdots((x_1 + x_2)/2 + x_3)/2 + \cdots + x_n)/2 \quad (35)$$

obtained with the first approach. In the same way, the function g involved in the general form for symmetrical sums can be generalized to a function of n variables, leading to

$$\sigma(x_1, \ldots, x_n) = \frac{g(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n) + g(1 - x_1, \ldots, 1 - x_n)}. \quad (36)$$

It should be noted that this approach may provide results which do not involve all variables, like for $\sigma_{\min}$ and $\sigma_{\max}$. For CD operators, this second approach can be more complicated, especially for operators involving a measure of conflict between sources. Let us take the example of the normalized possibility measures given in Section III-C. All terms can be easily generalized, as they are based on T-norms and T-conorms, but the conflict term h. A formulation of a "global" conflict between n sources could be for instance $h(1, 2, \ldots, n) = \sup_s i[\pi_1(s), \pi_2(s), \ldots, \pi_n(s)]$, for which the following relationship holds w.r.t. conflicts computed by pairs

$$h(1, 2, \ldots, n) \leq \min[h(i, j), 1 \leq i \leq n, i < j \leq n]. \quad (37)$$

If the conflict is derived from a distance between two fuzzy sets or two possibility distributions, it is much more difficult to derive a global measure from the conflicts by pairs.

As far as source reliability is concerned, the generalization by the second approach is easier than for conflict since either reliability is available for each source separately, or the different sources reliability are ordered, even if precise values are unknown. Thus, they do necessitate computation involving several sources as for conflict.

2) Generalizing Properties and Behaviors: The concepts of conjunction, disjunction and compromise for an operator of n variables take the following forms

$$F(x_1, \ldots, x_n) \leq \min(x_1, \ldots, x_n), \quad \text{(conjunction)} \quad (38)$$

$$F(x_1, \ldots, x_n) \geq \max(x_1, \ldots, x_n), \quad \text{(disjunction)} \quad (39)$$

$$\begin{aligned} \min(x_1, \ldots, x_n) &\leq F(x_1, \ldots, x_n) \\ &\leq \max(x_1, \ldots, x_n). \quad \text{(compromise)}. \end{aligned} \quad (40)$$

Let us consider a conjunctive CICB operator of two variables. Its generalization to three variables satisfies [2]: $F[F(x_1, x_2), x_3] \leq \min(x_1, x_2, x_3)$, and similar relationships hold whatever the order of combination. Thus, we have: $F(x_1, x_2, x_3) \leq \min(x_1, x_2, x_3)$. This expression is easily extended to any number n of variables. Therefore, the generalization of a conjunctive CICB operator is a conjunctive CICB operator.

In a similar way, it can be proved that the generalization of a disjunctive (resp. compromise) CICB operator is a disjunctive (resp. compromise) CICB operator. This holds even for nonassociative operators: although the precise values of the combination depends on the order in which the variables are taken, the behavior of the operator does not. Also the ordering of operators, shown on Fig. 2, is still valid.

For CIVB operators, they also remain CIVB, but may follow more complicated rules than their equivalent with two variables. Let us take the example of associative CIVB operators, like associative symmetrical sums (but medians), for which the behavior has been described in Section III-B. For such an operator F, we have

$$x \leq 0.5 \text{ and } y \leq 0.5 \Rightarrow F(x, y) \leq \min(x, y) \; (\leq 0.5) \quad (41)$$

$$\Rightarrow \begin{cases} F[F(x, y), z) \leq \min(x, y, z) & \text{for } z \leq 0.5, \\ F(x, y) \leq F[F(x, y), z)] \leq z & \text{else.} \end{cases} \quad (42)$$

Thus, if all three values are less than 0.5, then F is a conjunction, and if two are less than 0.5, and one greater, then F is a compromise or a conjunction, depending on the precise values. In the same way, if all three values are greater than 0.5, then F is a disjunction, and if two are greater and one is less than 0.5, then F is a compromise or a disjunction, depending on the values. The precise rules differentiating the cases where F is a compromise depend on the operator. For instance, for σ_0, we have

$$\sigma_0(x, y, z) = \frac{xyz}{1 - x - y - z + xy + yz + zx}. \quad (43)$$

Assuming, without loss of generality, $x \leq y \leq z$, the rules of behavior are:

- conjunction iff $z \leq 1 - y$ (and thus $y \leq 0.5$): this means that two values have to be low (less than 0.5) and the third one is limited;

- disjunction iff $1-x \leq y \leq z$: this means that either two values have to be high and the third one low (if $x \leq 0.5$) or the three values have to be high;
- compromise in all remaining cases.

All these remarks can be extended to more than three variables.

Fig. 3 illustrates an example of behavior, where the combined value is compared to x, y and z.

3) Behavior of Nonassociative Operators: If the first approach for combining information with a nonassociative operator is chosen, the main question is in which order the pieces of information have to be combined. Several situations may occur.

- In some applications, the information is provided successively by each sensor and each piece of information has to be combined to the previous ones as soon as it is available (for instance in order to be able to take partial decisions with the information at hand); in this case, the order is fixed by the application.
- The order may be imposed by some priority between information, and some operators are designed for dealing with such cases (an example of processing queries in data bases with operators expressing "and possibly" for instance is described in [4]).
- In all other situations, we have generally to design criteria to find the adequate order, in particular for conflictual information: do we have to combine first the conflictual values or the consensual ones?

In the sequel, we will focus on the third kind of situations, and provide some examples illustrating the behavior of nonassociative operators w.r.t. the chosen order.

Let us take for F the arithmetical sum (a nonassociative CICB operator) and three variables such that $x \leq y \leq z$. The comparison between the results provided by the first approach with all possible orderings and the second approach leads to

$$\begin{aligned} x \leq F[F(y,z),x] \leq \frac{x+y+z}{3} \leq F[F(x,z),y] \\ \leq F[F(x,y),z] \leq z, \end{aligned} \tag{44}$$

for $y-x \geq z-y$, and

$$\begin{aligned} x \leq F[F(y,z),x] \leq F[F(x,z),y] \leq \frac{x+y+z}{3} \\ \leq F[F(x,y),z] \leq z, \end{aligned} \tag{45}$$

for $y-x \leq z-y$, with

$$F[F(x,z),y] - F[F(y,z),x] = \frac{y-x}{4}, \tag{46}$$

$$F[F(x,y),z] - F[F(x,z),y] = \frac{z-y}{4}. \tag{47}$$

This shows 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 will probably not be crucial. On the contrary, if the information is conflictual, the variations may be more significant. For instance, if x is very low and y and z are both high and close to each other, then combining the consensual values first will provide a significantly smaller result (R_1) than combining conflictual values first. Moreover, the value $\frac{x+y+z}{3}$ provided by the second approach is farther from the nonconflictual values than R_1.

TABLE II
EXAMPLE OF A MATRIX OF μ_j^k, $1 \leq j \leq C$, $1 \leq k \leq n$, FOR $C = 4$ AND $n = 3$

image class	I_1	I_2	I_3
1	0.5	0.3	0.7
2	0.4	0.3	0.2
3	0.7	0.7	0.5
4	0.1	0.8	0.6
max	$\mu_3^1 = 0.7$	$\mu_4^2 = 0.8$	$\mu_1^3 = 0.7$

TABLE III
FUSION BY DIFFERENT T-CONORMS AND COMPARISON OF DECISIONS

operator class	$\max(x,y)$	$x+y-xy$	$\min(x+y,1)$
1	0.7	0.895	1.0
2	0.4	0.664	0.9
3	0.7	0.955	1.0
4	0.8	0.928	1.0
max	$\mu_4 = 0.8$	$\mu_3 = 0.955$	$\mu_1 = \mu_3 = \mu_4 = 1.0$

A more complicated example is provided by $\sigma_{\min}$ (also a mean operator) (see Section III-B). We denote

$$\sigma_1 = \sigma_{\min}[\sigma_{\min}(x,y),z], \ \sigma_2 = \sigma_{\min}[\sigma_{\min}(z,y),x]. \tag{48}$$

As shown on Fig. 4 where $(\sigma_1 - \sigma_2)$ is drawn for three values of z, if x, y and z are of the same order of magnitude, there is no significant difference. The largest differences occur for a large conflict between x and z. In particular, if x is high and y and z are low, then σ_1 (obtained by combining conflictual values first) is less than σ_2 (where consensual values are combined first), and the difference on this example is quite high ($\approx$0.4, which can be very important w.r.t. decision). In the same way, if y and z are high and x is low, then $\sigma_1 \geq \sigma_2$, and the differences decrease when x increases.

Let us now consider a CIVB nonassociative operator, σ_+ for instance (see Section III-B). Fig. 5 illustrates the differences between $\sigma_1 = \sigma_+[\sigma_+(x,y),z]$ and $\sigma_2 = \sigma_+[\sigma_+(z,y),x]$. In this case, the differences depend not really on the conflict but more on relative values. On Fig. 5, it is clear that if $x \leq z$ then $\sigma_1 \geq \sigma_2$, the difference depending only slightly on y, and if $x \geq z$ then $\sigma_1 \leq \sigma_2$. This means that the result is smaller if the largest value is first combined to y.

D. Behavior w.r.t. Conflict and Decision

In this section, we discuss the properties of operators in terms of behavior w.r.t. conflict, and decisiveness (i.e., their ability to discriminate the situations they have to face).

This part will be illustrated by the example of multi-sensor image classification. The following notations will be used: $\mu_j^k(x)$ denotes the information (confidence or belief) on the membership of x to the class j ($j = 1 \cdots C$) provided by image k ($k = 1 \cdots n$) (the mathematical form of μ depends

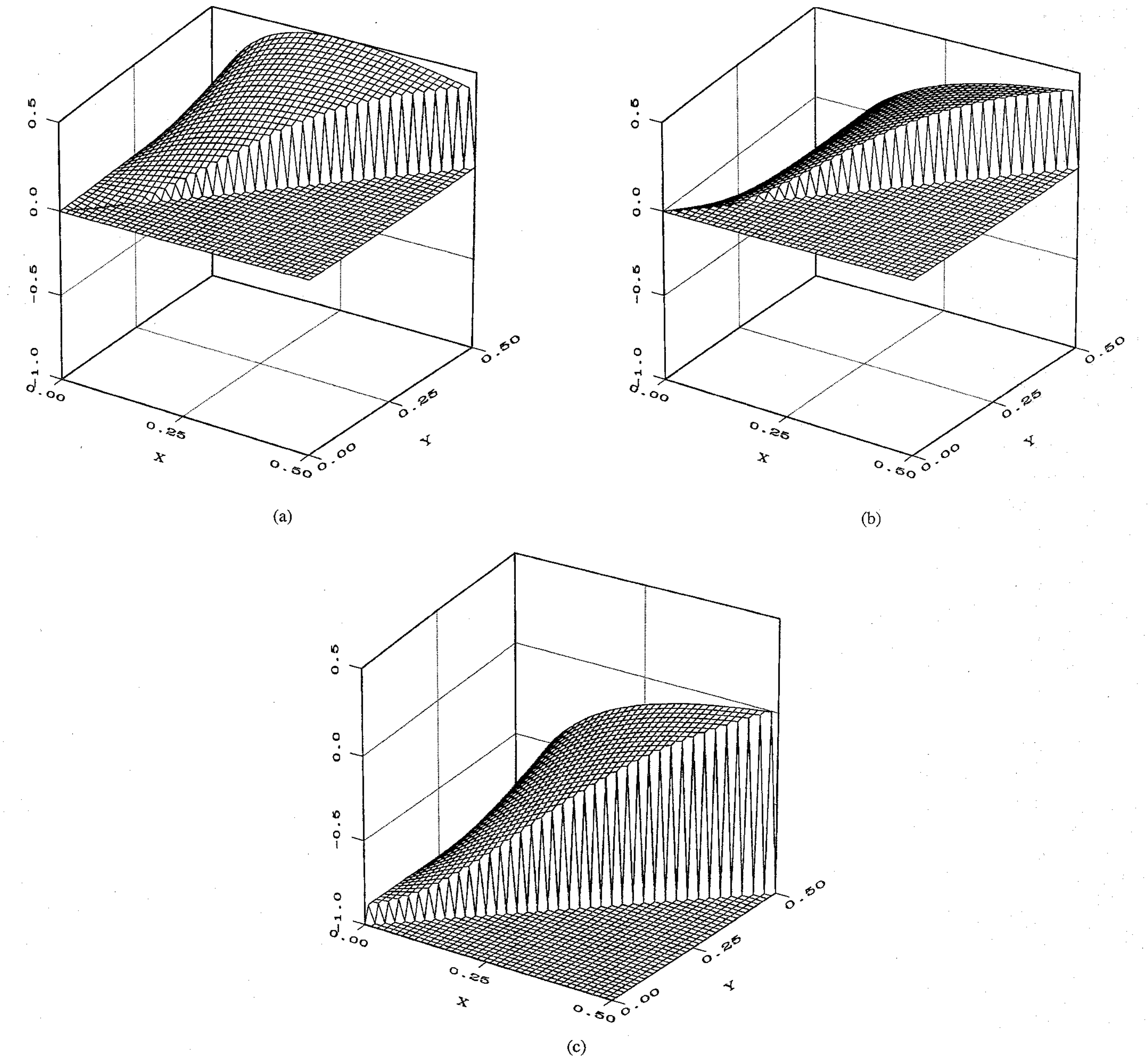


Fig. 3. (a)–(c) Behavior of an associative operator combining 3 values: differences between $\sigma_0(x, y, z)$ and x, y, and z, respectively, for $x < y < 0.5$ and $z = 0.9$; σ_0 behaves like a compromise (in this example $\max(x, y) < \sigma_0(x, y, z) < z$) but for very low values of x and y ($x \leq y \leq 1 - z = 0.1$).

on the chosen framework: probability, possibility, membership degrees to a fuzzy set, mass function, etc.); $\mu_j(x)$ denotes the combination of the $\mu_j^k(x)$'s by the operator F (i.e., a global information on class j resulting from the individual information provided by each image) $\mu_j(x) = F[\mu_j^k(x), k = 1 \cdots n]$.

Let us examine the following decision rules, which take a similar form as the classical rules used in fuzzy classification [1], but act on the result of information combination (i denotes the class chosen by the decision process)

$$\mu_i(x) = \max_{j=1}^{C} \mu_j(x), \tag{49}$$

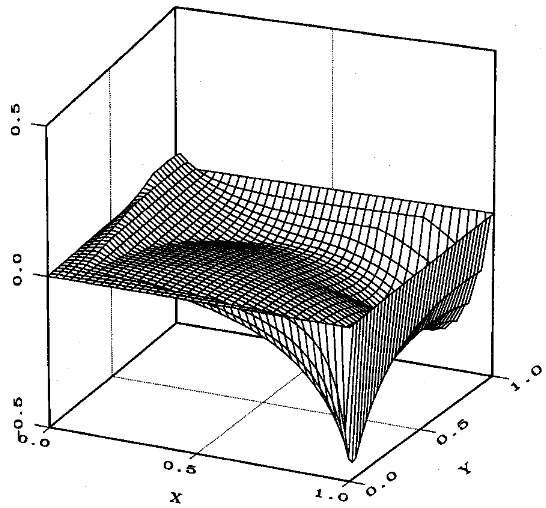
$$\mu_i(x) \geq d, \tag{50}$$

$$\begin{cases} \mu_i(x) - \max_{j \neq i} \mu_j(x) \geq \eta, \\ \text{or } \frac{\mu_i(x)}{\max_{j \neq i} \mu_j(x)} \geq A \ (A > 1). \end{cases} \tag{51}$$

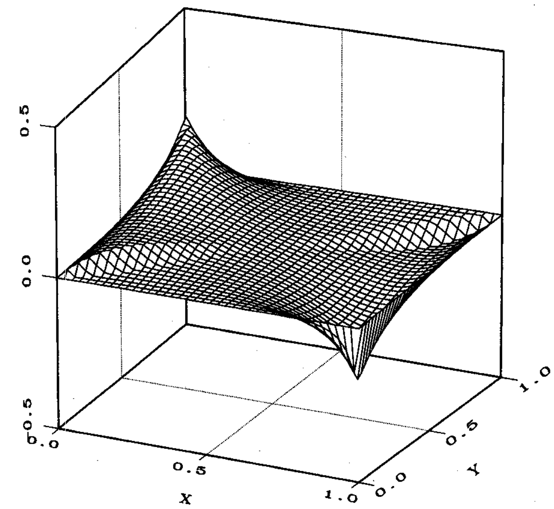
Condition (50) imposes that the confidence in class i is large enough in order to accept the decision "$x \in i$" (the parameter d represents a decision threshold). Condition (51) imposes that the decision in favor of a class is not ambiguous (i.e., decisive enough): the two largest values must be different enough.

Let us first take for F the "max", used in fuzzy set and possibility theories. According to rule (49), a point x will be assigned to class i iff

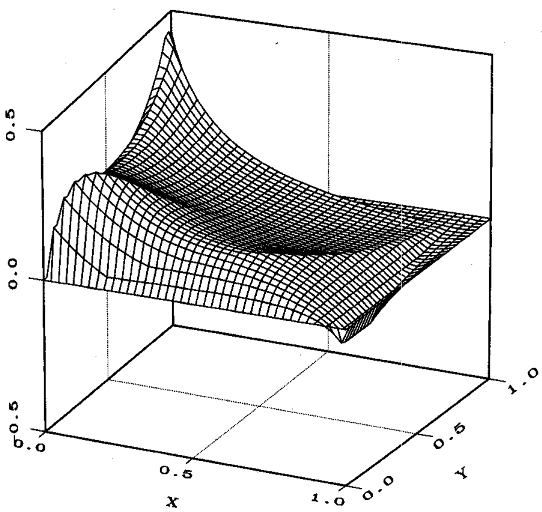
$$\begin{aligned} \mu_i(x) = \max_{j=1}^{C} \mu_j(x) &= \max_{j=1}^{C} \left[\max_{k=1}^{n} \mu_j^k(x) \right] \\ &= \max_{k=1}^{n} \left[\max_{j=1}^{C} \mu_j^k(x) \right]. \end{aligned} \tag{52}$$



(a)



(b)



(c)

Fig. 4. (a)–(c) Behavior of a nonassociative operator when combining pieces of information in different order: $\sigma_{\min}[\sigma_{\min}(x, y), z] - \sigma_{\min}[\sigma_{\min}(z, y), x]$ for z equal to 0.1, 0.5, and 0.9, respectively.

Thus, the decision process (49) after a fusion with the operator "max" is equivalent to

- a decision taken separately (locally) on each image according to rule (49) (for any k, $\mu_i^k(x) = \max_{j=1}^{C} \mu_j^k(x)$), the global decision being then taken by searching for the "est" local decision, i.e., the one for which $\mu_i^k(x)$ is maximum;
- a decision taken globally from all information, i.e., from the whole matrix of $\mu_j^k(x)$'s, by taking its maximum term.

Let us now consider for F any T-conorm. A point x will be assigned to class i iff

$$\begin{aligned}\mu_i(x) &= \max_{j=1}^{C} F(\mu_j^k(x), k = 1 \cdots n) \\ &\geq \max_{j=1}^{C} \left[\max_{k=1}^{n} \mu_j^k(x)\right] \geq \max_{k=1}^{n} \left[\max_{j=1}^{C} \mu_j^k(x)\right]. \quad (53)\end{aligned}$$

In this case, the decision is related to the maximum matrix coefficient only through an inequality. It is not necessarily taken in favor of the class for which this maximum is obtained. This is shown in Table III, which gives the results provided by the combination $F(\mu_j^k, k = 1 \cdots n)$ for different T-conorms, for the example of matrix of μ_j^k's presented in Table II for $C = 4$ and $n = 3$. The last row indicates the decisions taken locally on each image (with the corresponding membership value) (Table II) and on the result of the fusion (Table III).

The synthetic example in Table II is a very conflicting case (between the images for a given class and between classes in an image), where the individual decisions are different (class 3 for image 1, 4 for I_2, and 1 for I_3) and are high (values greater than 0.7). The global decision taken from these local decisions following the highest confidence value is in favor of class 4. The same result as the one obtained for the operator "max" is found, according to the above result. This operator does not take the important conflict on class 4

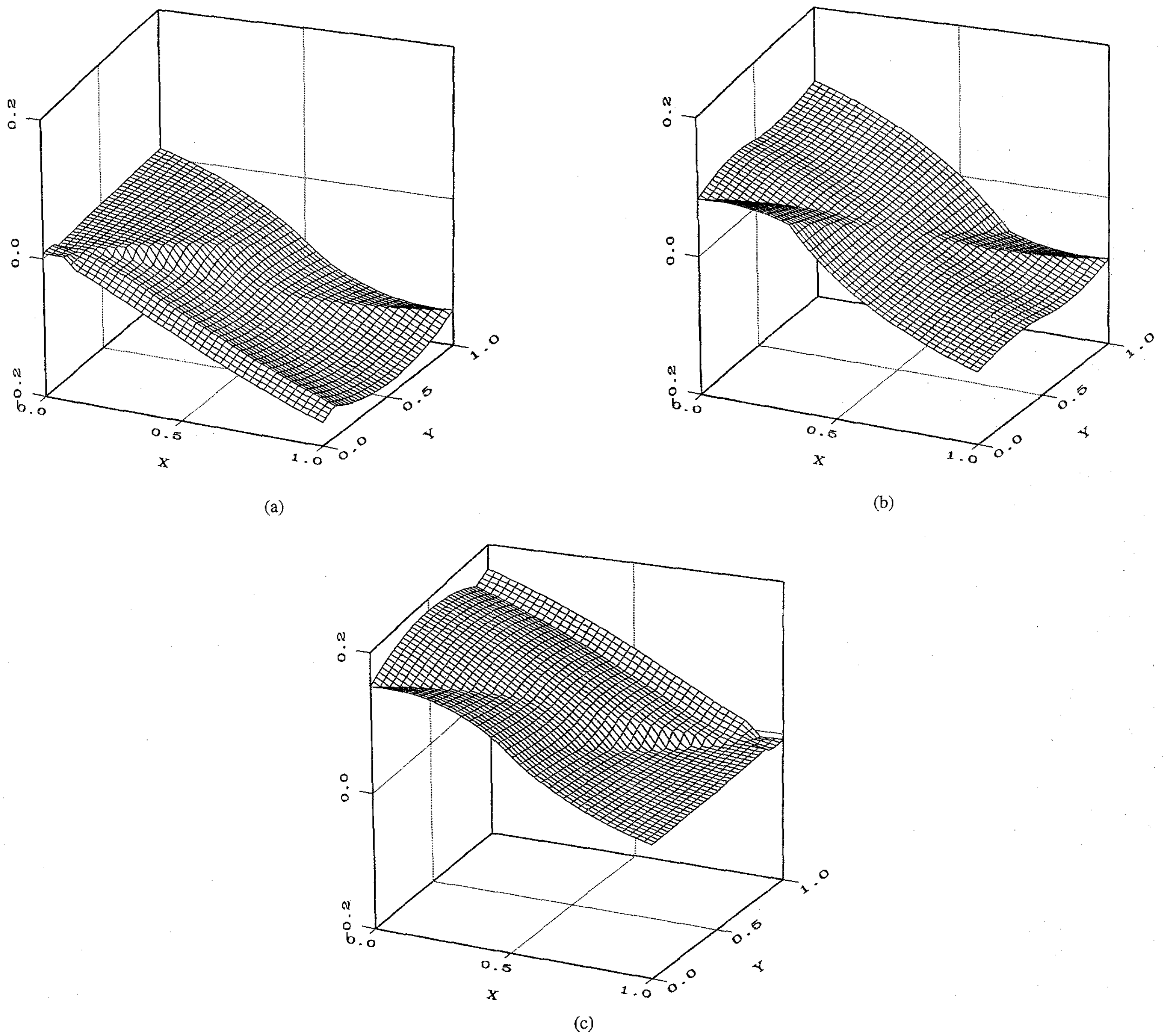


Fig. 5. (a)–(c) Behavior of a nonassociative operator when combining pieces of information in different order: $\sigma_+[\sigma_+(x,y),z] - \sigma_+[\sigma_+(z,y),x]$ for z equal to 0.1, 0.5, and 0.9, respectively.

into account, since the result is dictated only by the maximal values, without considering the possible deviations between the pieces of information given by the images on this class. On the contrary, the algebraic sum uses all values related to a class, including somehow the possible conflict: for instance, the result of the combination of images 2 and 3 for class 1 is less than for class 3 for which the conflict between the two images is smaller (the maximum value being the same). The decision for this operator is taken in favor of class 3, for which all images provide relatively high confidence with low conflict. As far as the bounded sum is concerned, the results show that this operator is not discriminant: it has a low decisiveness since it does not allow to decide between classes 1, 3, and 4. These phenomena are due to the fact that the "max" is the smallest T-conorm and all other reinforce the highest information: the decisiveness of the operators decreases when going to the right on Fig. 2 (a high T-conorm like $\min(1, x+y)$ provides the saturation value 1 as soon as $x + y \geq 1$, i.e., for a lot of values to be combined).

Let us now examine the behavior of T-conorms w.r.t. conditions (50) and (51). From the above derivations for the decision rule (49), it appears that the criterion $\mu_i(x) \geq d$ will be satisfied as soon as there exists one at least of the $\mu_j^k(x)$'s greater than d. For the T-conorm "max", this sufficient condition is also necessary. On the contrary for the other T-conorms, it is not necessary and will be satisfied more easily if the T-conorm is greater (on the right on Fig. 2), see for instance the results for the bounded sum in Table III. For the previous example, condition (50) is satisfied for the "max" iff $d \leq 0.8$, for the algebraic sum iff $d \leq 0.955$ and for the bounded sum iff $d \leq 1$. Therefore, in all cases, a large place is left when comparing these values with the traditional value $d = 0.5$.

TABLE IV
FUSION BY DIFFERENT T-NORMS AND COMPARISON OF DECISIONS

operator class	$\min(x, y)$	xy	$\max(0, x+y-1)$
1	0.3	0.105	0.0
2	0.2	0.024	0.0
3	0.5	0.245	0.0
4	0.1	0.048	0.0
max	$\mu_3 = 0.5$	$\mu_3 = 0.245$	$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0.0$

On the contrary, condition (51) is much more difficult to satisfy, in particular for large T-conorms. This is almost impossible for the bounded sum, unless we have only low confidence for all images and all classes (but possibly one); but in such a case, condition (50) will be difficult to satisfy. For the example in Table III, this criterion is satisfied for the "max" iff$\eta \leq 0.1$, and for algebraic sum iff $\eta \leq 0.027$!

Let us now revisit the previous example, under the light of T-norms as combination operators, as shown in Table IV. For any T-norm F, the decision rule (49) assigns x to class i iff

$$\begin{aligned}\mu_i(x) &= \max_{j=1}^{C} F\left[\mu_j^k(x), k = 1 \cdots n\right] \\ &\leq \max_{j=1}^{C}\left[\min_{k=1}^{n} \mu_j^k(x)\right] \\ &\leq \min_{k=1}^{n}\left[\max_{j=1}^{C} \mu_j^k(x)\right] \leq \max_{k=1}^{n}\left[\max_{j=1}^{C} \mu_j^k(x)\right]. \quad (54)\end{aligned}$$

This shows that for a T-norm, the obtained value is always smaller than the one obtained by a decision on each image individually and then a global decision by taking the best local decision. Also the class for which this decision is obtained may be different.

The decision taken as the best of the local decisions is again always in favor of class 4 (since the fusion operator has no influence). In Table IV, the operators "min" and product decide in favor of class 3 and thus privilege a less conflicting situation than the situation for class 4. The operator $\max(0, x+y-1)$ is not discriminant, since it is located far on the left on Fig. 2 (this phenomenon is analogous as the one already mentioned for the bounded sum).

Since T-norms are always smaller than the "min", condition (50) is much more difficult to satisfy if $d \geq 0.5$. In the previous example, it is satisfied for the "min" iff $d \leq 0.5$ and for the product iff $d \leq 0.245$, which is far from the usually chosen values (this shows that choosing a priori a decision threshold does not make sense as long as the fusion operator is not known). For the criterion (51), the "min" as well as the product are discriminant as long as η is not too high (typically, $\eta \leq 0.1$). The operator $\max(0, x+y-1)$ is not discriminant.

For probabilistic and Bayesian data fusion, a similar behavior is observed, as the involved operator is mainly the product.

For fusion performed with a mean operator, the resulting decision depends on the chosen operator but also on the order in which the pieces of informations are combined, as such an operator is generally not associative. If F is a mean operator,

TABLE V
FUSION BY DIFFERENT MEAN OPERATORS AND COMPARISON OF DECISIONS

operator class	$\frac{2xy}{x+y}$	$\sqrt{xy}$	$\frac{x+y}{2}$
1	0.488	0.521	0.550
2	0.253	0.263	0.275
3	0.583	0.592	0.600
4	0.274	0.412	0.525
max	$\mu_3 = 0.583$	$\mu_3 = 0.592$	$\mu_3 = 0.6$

operator class	$med(x, y, .5)$	$med(x, y, .7)$	$med(x, y, .2)$
1	0.5	0.7	0.3
2	0.4	0.4	0.2
3	0.5	0.7	0.5
4	0.5	0.7	0.2
max	$\mu_1 = \mu_3 = \mu_4 = 0.5$	$\mu_1 = \mu_3 = \mu_4 = 0.7$	$\mu_3 = 0.5$

we have

$$\min_{k=1}^{n} \mu_j^k(x) \leq F\left[\mu_j^k(x), k = 1 \cdots n\right] \leq \max_{k=1}^{n} \mu_j(x). \quad (55)$$

Therefore, the decision value according to (49) satisfies

$$\begin{aligned}&\max_{j=1}^{C}\left[\min_{k=1}^{n} \mu_j^k(x)\right] \leq \mu_i(x) \leq \max_{j=1}^{C}\left[\max_{k=1}^{n} \mu_j^k(x)\right] \\ &\Leftrightarrow \max_{j=1}^{C}\left[\min_{k=1}^{n} \mu_j^k(x)\right] \leq \mu_i(x) \leq \max_{k=1}^{n}\left[\max_{j=1}^{C} \mu_j^k(x)\right]. \quad (56)\end{aligned}$$

Here again, the decision will generally not be the same as the best of the local decisions, as illustrated in Table V. The behavior as compromise operators and the ordering between the operators is clearly apparent on this example. The decision taken by the three first operators is always in favor of class 3: these operators favor the class with the highest confidence values as far as the conflict is not too high (i.e., class 3 is chosen rather than class 4). For the median operators, it should be noted that the parameter α is obtained as a result as soon as there are values straddling α. Thus, low values of α lead to operators with a higher decisiveness and will select the class for which all pieces of information are greater than α (class 3 in our example for $\alpha = 0.2$). When α increases, the operator becomes less decisive (see Table V for $\alpha = 0.5$ and $\alpha = 0.7$). Typically, the decision threshold in condition (50) has to depend on the value of α. The median operators are not sensitive to the amplitude of possible conflict, as soon as the conflictual values are on both sides of α, which is generally the case for medium values of the parameter.

Table VI shows, again on the same example (Table II), the results obtained with several symmetrical sums. In all presented cases, the decision is taken in favor of class 3, i.e., the class which shows the highest values for a low conflict is chosen. Condition (50) is satisfied for the standard value $d = 0.5$. Considering condition (51), it can be observed that σ_0 has a high decisiveness ($\geq$0.3), and the other ones are much less decisive.

For the nonassociative operators of Tables V and VI, the results of Section IV-C apply. For the example of Table II,

TABLE VI
FUSION BY DIFFERENT SYMMETRICAL SUMS AND COMPARISON OF DECISIONS

operator class	σ_0	σ_+	$\sigma_{\min}$	$\sigma_{\max}$
1	0.500	0.544	0.556	0.545
2	0.067	0.360	0.231	0.313
3	0.845	0.547	0.625	0.583
4	0.400	0.524	0.455	0.531
max	$\mu_3 =$ 0.845	$\mu_3 =$ 0.547	$\mu_3 =$ 0.625	$\mu_3 =$ 0.583

TABLE VII
FUSION BY DEMPSTER-SHAFER COMBINATION RULE

class	Dempster-Shafer
1	0.249
2	0.057
3	0.580
4	0.114
max	$m(3) = 0.580$

TABLE VIII
FUSION BY THE MYCIN RULES

class	MYCIN
1	0.000
2	-0.808
3	0.640
4	0.000
max	$CF(3) = 0.640$

the behavior w.r.t. decisiveness depends only slightly on the order of combination.

Table VII presents the results obtained by Dempster-Shafer orthogonal rule of combination, for the example in Table II (after normalization of the confidence values). The conflict between 1 and 2 is 0.765, and between 3 and the combination of 1 and 2 is 0.749. Thus the global conflict between the 3 images, according to the formula given in section 2.1.3 is 0.941. The decision is taken in favor of class 3, with a good decisiveness. The global conflict is quite high, and the Dempster-Shafer rule, like several other operators described above, prefers the class with quite high confidence values and low local conflict (between the images for this class).

Table VIII presents the results obtained by the MYCIN rules (after a linear normalization of the confidence values into $[-1, 1]$). On this example, it appears that the rules have a high decisiveness: only the results for class 3 are positive, with quite high certainty factors. Thus the global set of images confirms only this class and disconfirms the other classes (or says nothing about them). Again, the class with the highest certainty factors for a low conflict is preferred.

A last remark about the examples presented above (Tables II–VII) concerns the interest of CD operators. Clearly, an information about reliability of the sources could drastically change the decision taken from these confidence values. For instance, if the sensor which provides image 2 has a very high reliability in comparison with the two other ones, it is very likely that we would prefer a decision in favor of class 4 (instead of class 3), even in case of strong conflict between the sensors about this class.

V. CONCLUSION: A GUIDE FOR CHOICE

A classification of data fusion operators has been proposed, according to the constance or nonconstance of their behavior and to the information they depend on (global information about the problem context or not). This classification provides a useful guide of choice of an operator among their great variety. We have shown that all commonly used operators in fusion theories like probability, Bayesian decision, fuzzy sets, possibility, Dempster-Shafer, or MYCIN fit in the proposed classification. In each class, the interpretation of the analytical and algebraic properties in terms of data fusion leads to a refinement of the choice. Simple examples were provided to show how a given problem may impose properties on the operators. At last, we discussed the properties of the operators according to their behavior in conflictual situations and in terms of decisiveness. Synthetical examples illustrate this aspect.

Therefore, these guides, which provide a better knowledge of the operators and of what can be expected from them, allow the user to address the problem of choosing an operator adapted to a given problem.

However, a difficulty remains concerning the choice of an adequate order when combining more than two sources with a nonassociative operator. Although some rules can be derived, as shown in the text, to describe the behavior of such operators when conflictual or consensual pieces of information are combined first, it remains difficult to derive a general strategy of processing.

ACKNOWLEDGMENT

The author would like to thank the reviewers for their useful comments, and for suggesting Table I and some additional references.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.
[2] I. Bloch, "Information combination operators for data fusion: A review with classification," Télécom Paris 94D013, Tech. Rep., Apr. 1994.
[3] I. Bloch and H. Maître, "Fusion de données en traitement des images," *Traitement Du Signal*, pp. 435–446, Dec. 1994.
[4] G. Bordogna and G. Pasi, "A fuzzy query language with a linguistic hierarchical aggregator," in *ACM Symp. Appl. Computing*, Phoenix, AZ, Mar. 1994, pp. 184–187.
[5] D. Dubois and H. Prade, "A review of fuzzy set aggregation connectives," *Inform. Sci.*, vol. 36, pp. 85–121, 1985.
[6] ——, *Théorie des Possibilités, Applications à la Représentation des Connaissances en Informatique*, Masson, Paris, 1988.
[7] ——, "Combination of information in the framework of possibility theory," in *Data Fusion in Robotics and Machine Intelligence*, M. Al Abidi *et al.* Eds. New York: Academic, 1992.
[8] M. Grabisch, "On the use of fuzzy integrals as a fuzzy connective," in *2nd IEEE Int. Conf. Fuzzy Systems*, San Francisco, CA, Mar. 28–Apr. 1, 1993, pp. 213–218.
[9] J. Guan and D. A. Bell, *Evidence Theory and its Applications*. Amsterdam: North-Holland, 1991.

[10] K. Menger, "Statistical metrics," in *Proc. Nat. Acad. Sci.*, USA, 1942, vol. 28, pp. 535–537.
[11] B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*. Amsterdam: North Holland, 1983.
[12] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[13] E. H. Shortliffe and B. G. Buchanan, "A model of inexact reasoning in medicine," *Math. Biosci.*, vol. 23, pp. 351–379, 1975.
[14] F. M. Song and P. Liang, "Inference with possibilistic evidence," *Uncertainty in Artificial Intelligence*, vol. 9, pp. 506–514, 1993.
[15] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. Dissertation, Tokyo Institute of Technology, 1974.
[16] P. Z. Wang and H. M. Zhang *et al.*, "Fuzzy set-operations represented by falling shadow theory," in *Proc. Int. Fuzzy Eng. Symp. '91*, Yokohama, Japan, 1991.
[17] R. R. Yager, "On a general class of fuzzy connectives," *Fuzzy Sets Syst.*, vol. 4, pp. 235–242, 1980.
[18] ——, "On ordered weighted averaging aggregation operators in multi criteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.
[19] ——, "Connectives and quantifiers in fuzzy sets," *Fuzzy Sets Syst.*, vol. 40, pp. 39–75, 1991.
[20] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.

Isabelle Bloch graduated from Ecole des Mines de Paris, France, in 1986, and received the Ph.D. degree from Télécom Paris (now Ecole Nationale Supèrieure des Télécommunications, Paris, France, in 1990.

Currently, Dr. Bloch is an Associate Professor, Dèpartement Images, Ecole Nationale Supèrieure des Télécommunications, Paris. Her research interests include 3D image and object processing, computational and discrete geometry, 3D and fuzzy mathematical morphology, decision theory, data fusion in image processing, fuzzy set theory, evidence theory, and medical imaging.