ELSEVIER

Pattern Recognition Letters 17 (1996) 905-919

Pattern Recognition Letters

Some aspects of Dempster–Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account

Isabelle Bloch *

École Nationale Supérieure des Télécommunications, Département Images, 46 rue Barrault, 75634 Paris Cedex 13, France

Received 15 September 1995; revised 6 February 1996

Abstract

This paper points out some key features of Dempster–Shafer evidence theory for data fusion in medical imaging. Examples are provided to show its ability to take into account a large variety of situations, which actually often occur and are not always well managed by classical approaches nor by previous applications of Dempster–Shafer theory in medical imaging. The modelization of both uncertainty and imprecision, the introduction of possible partial or global ignorance, the computation of conflict between images, the possible introduction of a priori information are all powerful aspects of this theory, which deserve to be more exploited in medical image processing. They may be of great influence on the final decision. They are illustrated on a simple example for classifying brain tissues in pathological dual echo MR images. In particular, partial volume effect can be properly managed by this approach.

Keywords: Data fusion; Dempster–Shafer evidence theory; Decision; Multi-modality imaging; Classification; Imprecision; Uncertainty; Conflict; Ignorance

1. Introduction

Recent technical advances have led to the multiplication of imaging systems, which are often used for observing a phenomenon from different points of view. They provide a large amount of information that must be interpreted as a whole in order to draw correct conclusions. This development has made data fusion in image processing an important step, now well recognized, in modern multi-source image analysis. In medical imaging in particular (Coatrieux et al., 1991), the physician may use images issued from different sources, each of them highlighting specific properties of tissues and pathologies. They may be images acquired with a single imaging technique using different acquisition parameters (for instance multi-echo MRI), or images obtained from several imaging techniques (for instance anatomical MRI imaging combined with functional PET imaging). The association of such images allows the medical expert to confirm and complete his diagnosis. However, the tools at his disposal for analysing these images consist of almost nothing but diagnosis workstations, on which one or two images can be displayed, and on which some elementary processings can be run. The fusion process itself is still mainly qualitative and mental (except for the preliminary step of registration, for which several software tools exist).

* E-mail: bloch@ima.enst.fr.

0167-8655/96/$12.00 © 1996 Elsevier Science B.V. All rights reserved
PII S0167-8655(96)00039-6

In most data fusion problems, the images to be combined are partly redundant, as they represent the same scene, and partly complementary, as they may highlight different characteristics. Typically, none of the images provide a completely decisive and reliable information. In addition, the information is often imprecise and uncertain, and these characteristics are inherent to the images, due to observed phenomenon, sensors, numerical reconstruction algorithms, resolution, etc. The aim of data fusion techniques is therefore to improve the decisions by increasing the amount of global information while decreasing its imprecision and uncertainty, by making use of redundancy and complementarity.

Up to now, the methods used for data fusion were generally derived from rule-based systems. They involved numerous pre-processing steps in order to extract high-level information, which was then used in the fusion step. Extreme cases are methods where a decision is taken on each image separately and the fusion step aims at combining these local decisions into a global decision. However, numerical methods also present some interesting advantages, since they allow to deal with the numerical information, along with its imperfect nature, until the last decision step. In particular, dealing with conflicting local decisions can be avoided. Such numerical methods are mainly used for pixel-based fusion, for instance for classification purposes (Bloch and Maître, 1994). Among these, probabilistic fusion and Bayesian inference are the most popular. But they suffer from serious drawbacks, which are still a matter for discussion. Alternatives to this approach are fuzzy sets theory, possibility theory, and Dempster–Shafer evidence theory. In this article, we will focus on the Dempster–Shafer approach (DS).

When considering medical applications, it seems that DS has been applied mainly for topics derived from artificial intelligence (DS actually issued from this domain, for dealing with reasoning under uncertainty), and thus concerns medical diagnosis, where propositional representations of evidences and knowledge are set in the DS framework (Smets, 1978; Gordon and Shortliffe, 1985; Lowrance, 1988; Baldwin, 1991; Baldwin, 1992). Evidences issued from the image signal (of "iconic" type rather than propositional) are seldom considered. To our knowledge, only a few papers report on image fusion by DS in medical imaging (Lee and Leahy, 1990; Chen et al., 1993) (both dealing with brain MRI), (Suh et al., 1990) (applied for MRI left ventricle). On the contrary, DS has been more widely used in satellite image processing (see e.g. (Garvey, 1986; Lee et al., 1987; Rasoulian et al., 1990; Van Cleynenbreugel et al., 1991; Cucka and Rosenfeld, 1992; Zahzah, 1992)). There are several reasons for this: on the one hand, data fusion problems in remote sensing are probably less recent, and on the other hand, comparing and judging results with respect to the ground truth is obviously easier in satellite image interpretation, since it does not necessitate any surgery on one's body to access the truth ...

However, DS offers a number of advantages which could be of great interest for medical image fusion and deserve to be detailed. In this article, we try to contribute to this task. The originality of the paper lies in underlining how some features of Dempster–Shafer decision theory may be further exploited in medical imaging.

In Section 2, we briefly outline the principles of DS reasoning, and present the main advantages of DS for medical imaging applications. In this section, a general point of view concerning the key features of DS is adopted. In Section 3, we choose a more particular point of view and focus on a particular application: brain tissue classification from dual-echo MR images of pathological brains. We have chosen these images since they constitute a good illustration of several problems which may be solved by DS. This application will be detailed under the light of the different aspects of DS.

2. Data fusion by Dempster–Shafer evidence theory: interest for medical imaging

2.1. Basic principles of Dempster–Shafer theory

DS allows for a representation of both imprecision and uncertainty through the definition of two functions: plausibility (*Pls*) and belief (*Bel*), both derived from a mass function m (or basic probability assignment) (Shafer, 1976; Guan and Bell, 1991). Mass functions are defined on the power set of the space of discernment D, i.e. a mass is attributed to each subset of D. In classification problems, D may for instance be the set of classes of interest, and a subset of D

represents a union of classes. This represents a major difference with probabilistic approaches which only assign probabilities to singletons (i.e. to subsets of D of cardinality 1). In the following, singletons will be called simple hypotheses, whereas subsets containing at least two elements of D are called compound hypotheses. A mass function m is thus a function from 2^D onto $[0,1]$, such that

$$m(\emptyset) = 0, \qquad \sum_{A \subset D} m(A) = 1. \tag{1}$$

A subset A with non-zero mass value is called a focal element.

The problem of assigning masses to hypotheses becomes more complicated if values have to be assigned to compound hypotheses. Section 3.2 is dedicated to this problem for medical image fusion.

Belief and plausibility functions are derived from the mass function, and are respectively defined by

$$Bel(\emptyset) = 0,$$
$$Bel(A) = \sum_{B \subset A} m(B) \quad \forall A \subset D,\ A \neq \emptyset, \tag{2}$$
$$Pls(\emptyset) = 0,$$
$$Pls(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subset D,\ A \neq \emptyset. \tag{3}$$

Clearly, we have the following properties:

$$Bel(D) = 1, \tag{4}$$
$$Pls(D) = 1, \tag{5}$$
$$Bel(A) \leqslant Pls(A) \quad \forall A \subset D, \tag{6}$$
$$Pls(A) = 1 - Bel(A^C) \quad \forall A \subset D. \tag{7}$$

It can be shown (Shafer, 1976; Guan and Bell, 1991) that the knowledge of any one of these three functions on 2^D is enough to derive the two others. DS provides an explicit measure of ignorance about an event A and its complementary A^C as the length of the interval $[Bel(A), Pls(A)]$ (called belief interval). It can also be interpreted as the imprecision on the "true probability" of A. The mass assigned to D can be interpreted as the global ignorance since this weight of evidence is not discernible among the hypotheses. In summary, as for probability theory, using numerical values in $[0,1]$ allows us to represent uncertainty, but, using the two functions *Bel* and *Pls*, DS is also able to represent imprecision.

If masses are assigned only to simple hypotheses ($m(A) = 0$ for $|A| > 1$), then the three functions m, *Bel* and *Pls* are equal and are a probability, called Bayesian mass function in (Shafer, 1976). Otherwise, there is no direct equivalence with probabilities.

Now, if we have evidence issued from several sources, which are modelized in the DS framework by means of the previously defined functions, these masses are combined by the orthogonal rule of Dempster (Shafer, 1976). For m_j being the mass function associated with source j ($j = 1, \ldots, n$), this rule is written, for all non-empty subset A of D:

$$(m_1 \oplus m_2 \oplus \cdots \oplus m_n)(A) = \frac{\sum_{B_1 \cap \cdots \cap B_n = A} m_1(B_1) m_2(B_2) \cdots m_n(B_n)}{1 - \sum_{B_1 \cap \cdots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \cdots m_n(B_n)},$$
$$(m_1 \oplus m_2 \oplus \cdots \oplus m_n)(\emptyset) = 0 \tag{8}$$

if this expression is defined, i.e. if

$$k = \sum_{B_1 \cap \cdots \cap B_n = \emptyset} m_1(B_1) m_2(B_2) \cdots m_n(B_n) < 1. \tag{9}$$

Similar equations can be derived for directly combining belief or plausibility functions. To some extent, k can be interpreted as a measure of conflict between the sources and is directly taken into account in the combination as a normalization factor. It represents the mass which would be assigned to the empty set if masses were not normalized. It is very important to take this value into account for evaluating the quality of the combination: when it is high (in case of strong conflict: $k \approx 1$), the combination may not make sense and may lead to questionable decisions (moreover, the combination rule is not continuous if k is very close to 1 (Dubois and Prade, 1988)).

When m, *Bel* and *Pls* are Bayesian mass functions (i.e. the only focal elements are singletons), issued from independent sources, Dempster's rule is consistent with the laws of probability and the combination of Bayesian mass functions results in a Bayesian mass function. Thus probability appears as a limit of DS, in the case where no ambiguity or imprecision exists and where only uncertainty has to be taken into account.

After the combination, the final decision is taken in favour of a simple hypothesis using one of several

rules (Appriou, 1991; Denœux, 1995): for instance, the maximum of plausibility (generally over simple hypotheses), the maximum of belief, or the maximum of belief without overlapping of belief intervals, i.e. in favour of $d \in D$ such that

$$Bel(d) \geqslant \max_{d' \in D, d' \neq d} Pls(d')$$

(this is a very strict condition, called the absolute decision rule).

2.2. Main advantages of Dempster–Shafer theory

As mentioned above, DS includes modelization of both imprecision and uncertainty and assigns masses to all elements of 2^D rather than to the elements of D only. These points constitute some of the main advantages of the DS approach. Indeed, they lead to a very flexible and rich modelling, which is able to fit a very large class of situations, occurring in particular in medical imaging. A few examples of situations where DS theory may be successfully used are:

- in ideal cases where all information relevant to the problem is known, i.e. cases where Bayesian fusion applies;
- when a source provides information concerning only a few of several classes: for instance brain PET images under certain condition allow for the detection of the brain surface but not of the head surface;
- when a source differentiates two classes and another may not (this example will be illustrated in Section 3): DS allows us to deal with hesitation or ambiguity between these two classes;
- in the case of partial volume effects (a particular case of the previous one): it can also be taken into account by assigning masses to the union of the two classes mixed in the considered area (an example will be given in Section 3.6);
- in cases where global source reliability has to be taken into account: this may be done by weakening all masses and reinforcing $m(D)$ (such that the normalization constraint is still satisfied);
- in cases where knowledge of source reliability is available only for some classes: it can be taken into account by modifying accordingly the masses assigned to these classes and by introducing ignorance;
- in cases where a priori information has to be introduced: even if it is not represented in a probabilistic manner, it can be taken into account if it induces a way to assign masses, in particular to compound hypotheses; for instance if we know that a source is not able to distinguish between two classes, then it is not worth trying to estimate masses for these two classes separately, but the estimation has to be made on the union of the classes and mass is then assigned to the corresponding compound hypothesis.

DS provides an efficient way for representing some extreme kinds of information. A total certainty about an element d of D is represented by the mass function m_C defined by

$$m_C(\{d\}) = 1, \quad \text{and thus}$$
$$m_C(A) = 0 \quad \forall A \subset D,\ A \neq \{d\}. \tag{10}$$

Analogously, total ignorance is represented by the mass function m_I:

$$m_I(D) = 1, \quad \text{and thus}$$
$$m_I(A) = 0 \quad \forall A \subset D,\ A \neq D. \tag{11}$$

These definitions are related to algebraic properties of the DS combination rule: m_C is a null element for $\oplus$, while m_I is the identity. This means that the combination of any mass function m by m_C results in m_C (if m is not completely conflicting, i.e. $k = 1$, with m_C), whereas the combination of m by m_I results in m. This perfectly fits the idea of total certainty, which cannot be changed by any other information, and of total ignorance, which cannot have any influence on any other information.

The set of mass functions, along with the $\oplus$ operator, has a strong algebraic structure, which is an additional advantage of DS. As has been pointed out, there exist an identity and null elements. Moreover, DS combination is commutative and associative. It is not idempotent (in general $m \oplus m \neq m$). Finally, it behaves in a conjunctive way, while conflict behaves in a disjunctive way (increases) when more sources are combined (Shafer, 1976; Bloch, 1996).

3. Application to brain tissue classification from multi-echo MRI data

In this section, we consider a particular application to medical imaging. It concerns the classification of

dual-echo MR images of the brain, of patients suffering from adrenoleukodystrophy (ALD) (Aubourg et al., 1992). Examples of both echoes from one slice are shown in Fig. 1. The two echoes acquired for each slice are registered.

3.1. Image description

Fig. 1 shows a good discrimination between brain, ventricles (V) and cerebro-spinal fluid (CSF) in the first image, which however is not able to distinguish between white matter (WM) and ALD, nor between grey matter (GM) and WM. On the contrary, the second image shows ALD very clearly (the white area), presents slight differences between WM and GM, but the ventricles have almost the same grey-levels than GM and their contours are imprecise. Therefore, these images constitute a typical example where data fusion is needed. They are also well suited for illustrating several aspects of DS. For the sake of clarity, we will deal with only 3 classes (C_1 = WM + GM, C_2 = V + CSF and C_3 = ALD), and illustrate using this simple case the influence of the various parameters involved in DS. Then, we show how the partial volume effect on the boundary between ALD and WM can be explicitly introduced as a mass on a compound hypothesis.

3.2. Mass function definition from problem modelling

The definition of mass functions remains a largely unsolved problem, which did not yet found a general answer. In image processing, they may be derived at three different levels. At the highest, most abstract level, information representation is used in a way similar to that in artificial intelligence and masses are assigned to propositions, often provided by experts (Neapolitan, 1992; Baldwin, 1991; Baldwin, 1992; Gordon and Shortliffe, 1985). Up to now, this kind of information is usually not derived from measures on the images. This is therefore beyond the scope of this paper. At an intermediate level, masses are computed from attributes, and may involve simple geometrical models (Chen et al., 1993; Van Cleynenbreugel et al., 1991; Cucka and Rosenfeld, 1992; Andress and Kak, 1988). This is well adapted to model-based pattern recognition but it is difficult to use for image fusion classification of complex structures (like the cortex) without a model. At the pixel level, mass assignment is inspired from statistical pattern recognition. The most widely used approach is as follows: masses on simple hypotheses are computed from probabilities or from the distance to a class centre (Garvey et al., 1981; Garvey, 1986; Lee et al., 1987; Strat,

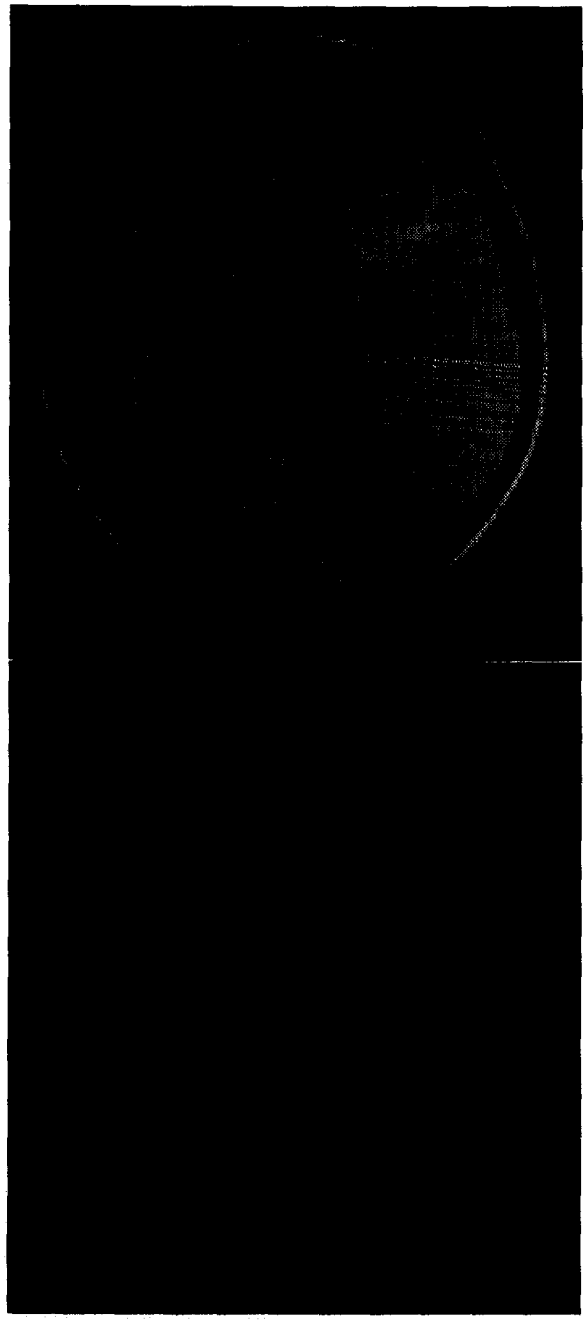

Fig. 1. Initial images: dual echo brain MRI (one slice).

1989; Rasoulian et al., 1990; Suh et al., 1990; Lee and Leahy, 1990; Lowrance et al., 1991; Appriou, 1991; Clarke and Wilson, 1991; Zahzah, 1992; Appriou, 1993; De Maertelaere et al., 1993; Ip and Ng, 1994). Then global ignorance ($m(D)$) is introduced as a weighting factor, often as a constant on all pixels (e.g. (Lee and Leahy, 1990; Zahzah, 1992)). In most cases no other compound hypothesis is considered, and this drastically limits the power of DS. One way to overcome this problem consists in using the decomposability property on the masses and the Barnett approach (only masses on singletons and on their complements are considered) (Barnett, 1981; Lee and Leahy, 1990; Chen et al., 1993; Gordon and Shortliffe, 1985). Although this approach may be interesting for pattern recognition purposes where each class has to be tested against all the others, it may however be questionable when a source provides information on a compound hypothesis which is neither a singleton nor the complement of a singleton. In (Lee et al., 1987), the use of mixed pixels in a probabilistic like approach is suggested. In (Suh et al., 1990), nested focal elements are used, by sorting probabilities. In our opinion, these approaches are quite promising but are also restrictive since they do not allow to take into account all occurring situations.

The way we assign masses in our example is based on a reasoning approach where knowledge about the information provided by each image is used to choose the focal elements. We argue that, although this approach may probably not be easily made automatic for any problem, it is more flexible and is able to take into account a larger variety of situations. In this way, based on the modelling previously introduced (Section 3.1), the focal elements of m_1 related to the first image are C_1, C_2, C_3, $C_1 \cup C_3$. Since C_1 and C_3 are not discriminated on this image, a possible choice for mass assignment consists in assigning the same mass to C_1, C_3 and $C_1 \cup C_3$. Null mass functions are assigned to the other compound hypotheses, since the corresponding classes cannot be confused. On the second image, it is difficult to separate brain and ventricles, and thus the focal elements are C_1, C_2, C_3 and $C_1 \cup C_2$, with the same masses on C_1, C_2 and $C_1 \cup C_2$. The introduction of global ignorance will be discussed in Section 3.5. An additional compound hypothesis will be introduced in Section 3.6 for the second image. The mass functions are simple functions of trapezoidal

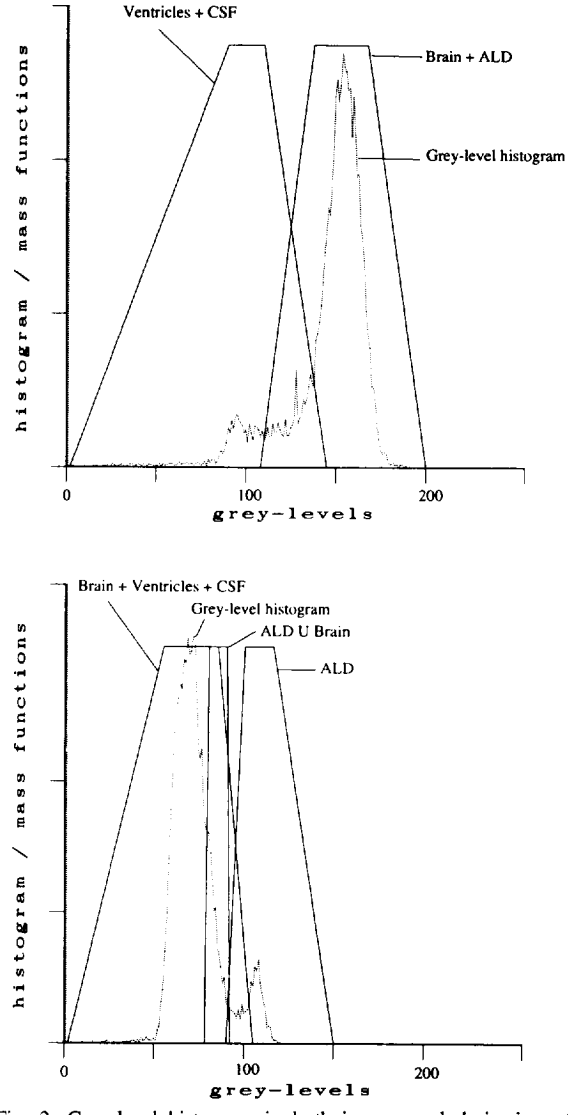

Fig. 2. Grey-level histogram in both images and derivation of mass functions (the scale is different for mass functions and for histograms). Mass functions are simple trapezoidal functions whose parameters are derived from histogram (the mass function on $C_1 \cup C_3$ (union of ALD and brain) will be used in Section 3.6). Afterwards, they are normalized.

shape, derived from grey-level histograms (Fig. 2), and such that the functions are normalized in order to verify $\sum_{A \subset D} m(A) = 1$. The parameter of the functions could be obtained in an automatic way, for instance from scale-space analysis of the histogram, as proposed in (Aurdal et al., 1995). This model is coarse but was found to be sufficient. Thus, in this applica-

tion, the classification is performed only from grey-level, and a pixel based fusion will be made. Finally, in the next two subsections, we will deal with:

$$\begin{cases} m_1(C_1) = m_1(C_3) = m_1(C_1 \cup C_3), \\ m_1(C_2) = 1 - 3m_1(C_1), \\ m_1(C_1 \cup C_2) = m_1(C_2 \cup C_3) = m_1(D) = 0. \end{cases} \quad (12)$$

$$\begin{cases} m_2(C_1) = m_2(C_2) = m_2(C_1 \cup C_2), \\ m_2(C_3) = 1 - 3m_2(C_1), \\ m_2(C_1 \cup C_3) = m_2(C_2 \cup C_3) = m_2(D) = 0. \end{cases} \quad (13)$$

3.3. Combination and decision

In this section, we discuss some aspects related to combination and decision. A first remark concerns the influence of the combination on the belief intervals. It is easy to prove that the combination of any mass function with a mass function not supporting compound hypotheses (i.e. a Bayesian mass function) results in a Bayesian mass function. Therefore the belief intervals reduce to 0. In the case where D contains only two elements, the belief intervals of the combination of any two mass functions are smaller than the original ones. In the application at hand (Eqs. (12) and (13)), the combination of m_1 and m_2 results in a Bayesian mass function. Therefore, the belief intervals reduce to 0. This effect may be of great interest if several images are combined. This means that each image makes the global information more precise in such cases. However the conflict increases, making the global result more questionable.

It should be noted that the measure of conflict is not an absolute measure of conflict between the images but depends on the problem at hand, on the way it is modelized in the DS framework, and on the way masses are assigned (on the focal elements in particular). In Fig. 3, the conflict between m_1 (given by Eq. (12)) and m_2 (given by Eq. 13)) is presented as a function of $m_1(C_2)$ and $m_2(C_3)$, along with the corresponding conflict image.

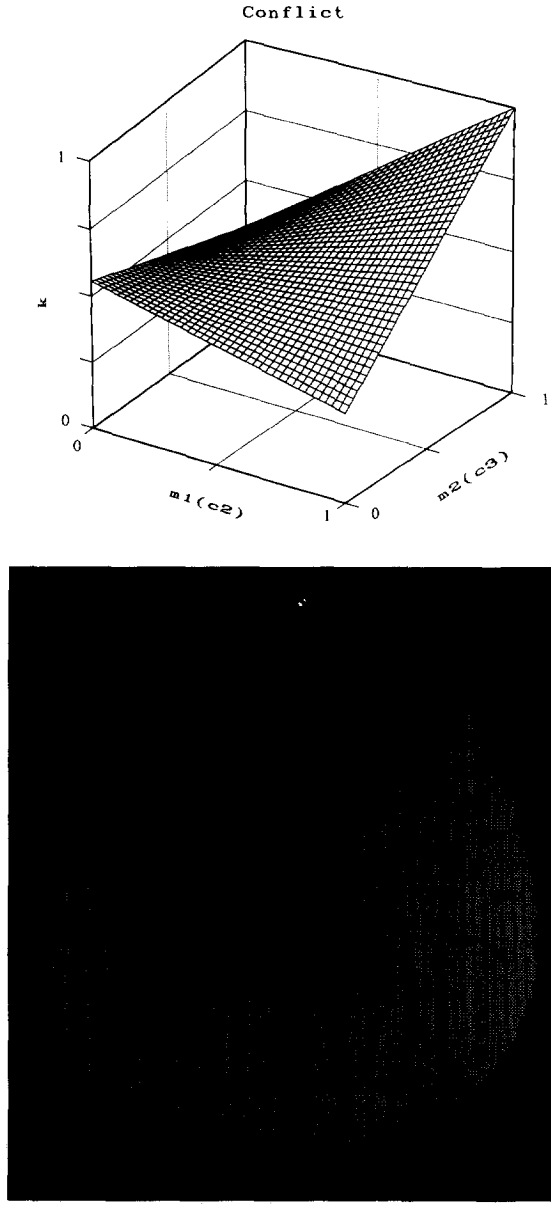

Fig. 3. Conflict between m_1 and m_2 (Eqs. (12) and (13)) as a function of $m_1(C_2)$ and $m_2(C_3)$, and conflict image for the considered application (the range of conflict values has been rescaled between 0 and 255, high grey-levels representing high conflict values).

Several decision rules have been used in the literature. The most frequently used is the maximum of belief among the simple hypotheses. The maximum plausibility rule is judged as being the best by some authors (Appriou, 1991; Appriou, 1993; De Maertelaere et al., 1993). The absolute rule has been used in (Suh et al., 1990). Other rules have been proposed like $\max(Bel + Pls)/2$ (Lee and Leahy, 1990) or $\max(Bel(A) - Bel(A^C))$ (Wesley, 1986) (these two rules are actually equivalent). It is remarkable that decision rules are seldom compared. Always taking a

decision in favour of a simple hypothesis actually reduces to making a crisp decision, and this does not always fit real situations in medical imaging, where pixels may belong to the union of classes and not to any one class exclusively. But since we have $Bel(A) \geqslant Bel(d)$ for any $d \in A$ and $Bel(D) = 1$, making the decision for compound hypotheses should be done with great care. It could for instance be decided in favour of a compound hypothesis if the evidences supporting simple hypotheses are not strong enough. In Fig. 4, decision has been taken following the maximum belief over all hypotheses excluding D. Note that in this simple case, the maximum of belief is equivalent to the maximum of plausibility, since $m_1 \oplus m_2$ is a Bayesian function. Therefore, using this decision rule, the decision is taken for a simple hypothesis at points where the other masses are null, and for a compound hypothesis elsewhere. Thus we obtain very interesting results, where the partial volume effect is detected as a compound hypothesis, whereas areas without ambiguity are well segmented. Fig. 5 shows the result obtained by taking a decision in favour of the simple hypothesis having the maximum belief.

The decision areas in Figs. 4 and 5 are obtained in the following way. Let us denote $m_1(C_2) = a$ and $m_2(C_3) = b$, where a and b depend on grey-levels. According to Eqs. (12) and (13), we have

$$\begin{aligned} m_1(C_1) = m_1(C_3) = m_1(C_1 \cup C_3) &= \frac{1-a}{3}, \\ m_2(C_1) = m_2(C_2) = m_2(C_1 \cup C_2) &= \frac{1-b}{3}. \end{aligned} \tag{14}$$

The combination by Dempster rule (Eq. (8)) provides the following results for the belief function (k being the conflict):

$$\begin{aligned} Bel(C_1) &= \frac{4(1-a)(1-b)}{9(1-k)}, \\ Bel(C_2) &= \frac{2a(1-b)}{3(1-k)}, \\ Bel(C_3) &= \frac{2(1-a)b}{3(1-k)}, \\ Bel(C_1 \cup C_2) &= \frac{2(1-b)(2+a)}{9(1-k)}, \\ Bel(C_1 \cup C_3) &= \frac{2(1-a)(2+b)}{9(1-k)}, \\ Bel(C_2 \cup C_3) &= \frac{2(a+b-2ab)}{3(1-k)}. \end{aligned} \tag{15}$$

If the decision is taken only in favour of a simple hypothesis, C_i will be chosen iff $Bel(C_i) > \max_{j \neq i} Bel(C_j)$. This is expressed as a function of a and b as

$$\begin{aligned} &\text{Decision } C_1 \Leftrightarrow a < 0.4 \text{ and } b < 0.4, \\ &\text{Decision } C_2 \Leftrightarrow a > 0.4 \text{ and } b < a, \\ &\text{Decision } C_3 \Leftrightarrow b > 0.4 \text{ and } b > a. \end{aligned} \tag{16}$$

This corresponds to the graph given in Fig. 5.

Analogously, if the decision is taken in favour of any hypothesis excluding D, we obtain

$$\begin{aligned} &\text{Decision } C_1 \cup C_2 \Leftrightarrow a > b \text{ and } b < 0.4, \\ &\text{Decision } C_1 \cup C_3 \Leftrightarrow a < b \text{ and } a < 0.4, \\ &\text{Decision } C_2 \cup C_3 \Leftrightarrow a > 0.4 \text{ and } b > 0.4, \\ &\text{Decision } C_1 \Leftrightarrow a = b = 0, \\ &\text{Decision } C_2 \Leftrightarrow a = 1, \\ &\text{Decision } C_3 \Leftrightarrow b = 1. \end{aligned} \tag{17}$$

This corresponds to the graph given in Fig. 4.

3.4. Evidence weightings

In this section we address the problem of weighting the masses with respect to each other, while keeping their sum equal to 1. If masses are assigned to simple hypotheses from probabilities (more precisely, from posterior conditional probabilities), they are generally deduced from frequencies (histogram for instance), leading to conditional probabilities. They are then naturally weighted by prior probabilities (and normalization factor). If global ignorance is introduced as a mass on D, the other masses are directly weakened by a factor $1 - m(D)$.

When masses are assigned to compound hypotheses, as in this application, the problem becomes more difficult. We have tested different weighting factors on $m_1(C_1) = m_1(C_3) = m_1(C_1 \cup C_3)$ and $m_2(C_1) = m_2(C_2) = m_2(C_1 \cup C_2)$, with respect to $m_1(C_2)$ and $m_2(C_3)$. The computation of k shows no systematic evolution of conflict. For instance, if $m_2(C_3)$ is low, the conflict decreases if the weight of evidence on $m_1(C_2)$ increases (in this case, the two sources are in better accordance in favour of C_2, i.e. the ventricle class). However, it was observed that attribut-

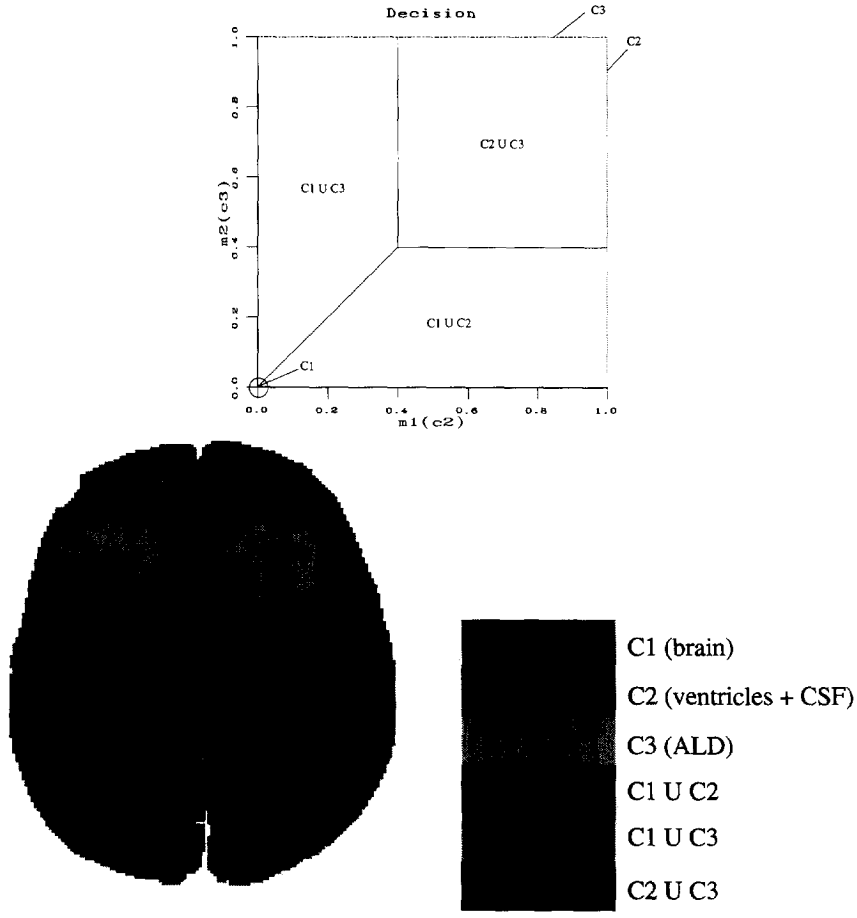

Fig. 4. Decision areas depending on $m_1(C_2)$ and $m_2(C_3)$ and decision image, by taking the maximum of belief over all hypotheses excluding D.

ing a mass on compound hypotheses (for instance on $m_1(C_1 \cup C_3)$, if we know that source 1 is not able to discriminate C_3 in the brain (C_1) areas) tends to decrease the conflict. This explains why there is little conflict in the pathological area in Fig. 3. In our opinion, this is a very important behaviour of DS, which highlights the interest of compound hypotheses.

The computation of decision areas (like in Figs. 4 and 5) shows that if the weight on $m_2(C_3)$ and $m_1(C_2)$ increases, the decision area in favour of C_1 decreases (when the decision is made based on simple hypotheses). This has also been observed in the decision images: Fig. 6 shows a zoom of the decision images over simple hypotheses, where ventricles and CSF are detected better if the weight on $m_2(C_3)$ and $m_1(C_2)$ is increased. Also, some differences can be observed on the small branches of ALD which are also detected better. It is remarkable that areas where differences occur are all classified in favour of compound hypothesis $C_1 \cup C_2$ (i.e. brain or ventricles + CSF) in Fig. 4 when decision is taken over all hypotheses excluding D (respectively $C_1 \cup C_3$ for the branches of ALD). With this decision rule, almost no differences are observed in the decision images when weights change. This supports the conclusion that compound hypotheses are important. This decision rule is more robust with respect to evidence weighting.

If the ambiguity is resolved by assigning masses only to the corresponding compound hypotheses (on $m_1(C_1 \cup C_3)$, with $m_1(C_1) = m_1(C_3) = 0$, and on $m_2(C_1 \cup C_2)$, with $m_2(C_1) = m_2(C_2) = 0$) only a few differences are observed (see Fig. 6). The decision

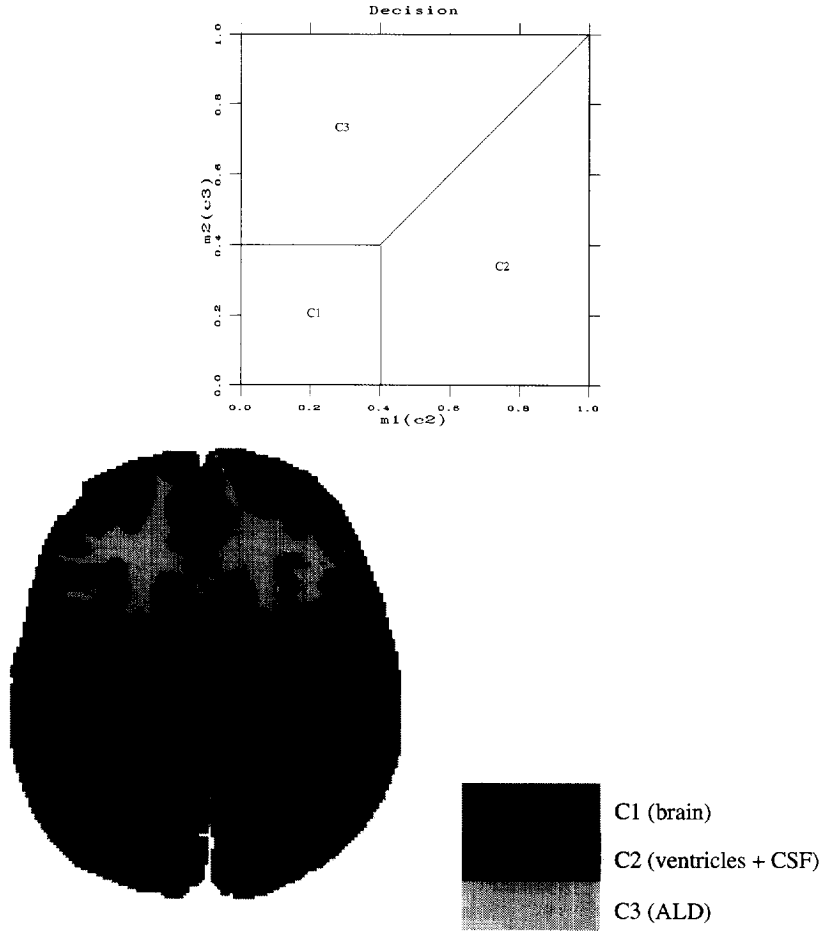

Fig. 5. Decision areas depending on $m_1(C_2)$ and $m_2(C_3)$ and decision image, by taking the maximum of belief over all simple hypotheses.

area in favour of C_1 is slightly increased and conflict is reduced.

3.5. Global ignorance

In this section we introduce a global ignorance, i.e. a mass on D, and test its influence. Thus the mass functions are now

$$\begin{cases} m_1'(C_1) = m_1'(C_3) = m_1'(C_1 \cup C_3), \\ m_1'(D) = 1 - m_1'(C_2) - 3m_1'(C_1), \\ m_1'(C_1 \cup C_2) = m_1'(C_2 \cup C_3) = 0. \end{cases} \tag{18}$$

$$\begin{cases} m_2'(C_1) = m_2'(C_2) = m_2'(C_1 \cup C_2), \\ m_2'(D) = 1 - m_2'(C_3) - 3m_2'(C_1), \\ m_2'(C_1 \cup C_3) = m_2'(C_2 \cup C_3) = 0. \end{cases} \tag{19}$$

This model allows for the introduction of reliability coefficients on each source, which serve as weighting factors for weakening masses on all hypotheses excluding D. Denoting these reliability coefficients by λ_1 and λ_2 in both images ($0 < \lambda_i \leqslant 1$) we thus obtain, for $i = 1, 2$,

$$\begin{aligned} &m_i'(A) = \lambda_i m_i(A) \quad \forall A \neq D, \\ &m_i'(D) = m_i(D) + (1 - \lambda_i)(1 - m_i(D)) \end{aligned} \tag{20}$$

(with $m_i'(D) = 1 - \lambda_i$ if $m_i(D) = 0$).

The computation of conflict provides $k' = \lambda_1 \lambda_2 k$, and therefore the conflict decreases. On the contrary, the ignorances on a hypothesis and its complement increase, since we have

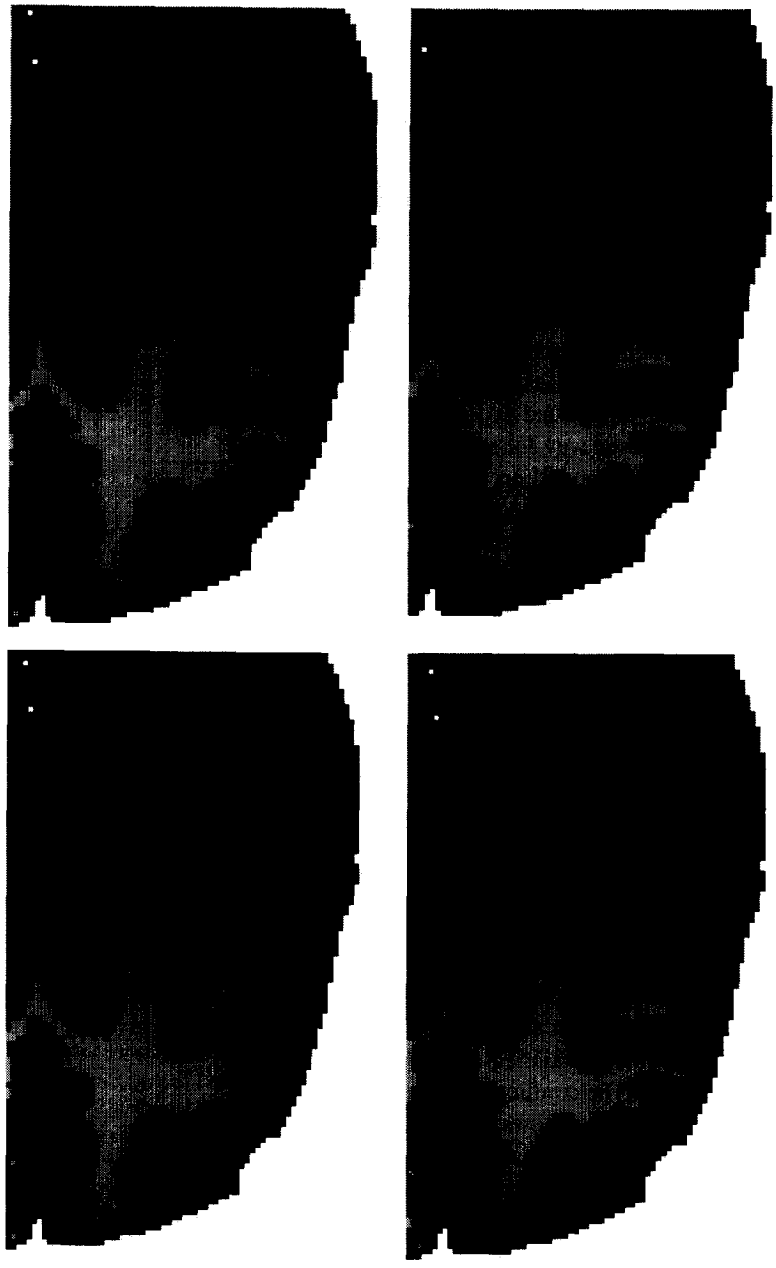

Fig. 6. Decision by maximum of belief over all simple hypotheses, for increasing weight on $m_2(C_3)$ and $m_1(C_2)$ (3 first images), and with mass only on the compound hypothesis (last image) corresponding to ambiguity between classes on each image ($m_2(C_3)$ and $m_1(C_2)$ are kept unchanged).

$$\forall A \subset D,$$
$$Pls_i'(A) - Bel_i'(A) = [Pls_i(A) - Bel_i(A)] + (1 - \lambda_i)[1 - (Pls_i(A) - Bel_i(A))] \geqslant [Pls_i(A) - Bel_i(A)].$$

Since the result of the combination $m_1 \oplus m_2$ is no longer a Bayesian function, the decision based on the maximum of belief is no longer equivalent to the decision based on the maximum of plausibility. Indeed, some small differences have been observed in the de-

cision images.

We observed very small differences in the decision made based on the simple hypotheses. However, decisions over all hypotheses excluding D are always made in favour of a compound hypothesis, as expected, if $\lambda_i \neq 1$. In our application, we do not have strong evidence for an image being more or less reliable than the other one (in a global sense). Therefore assigning a mass to $m(D)$ is not very useful, and this is confirmed by the results. On the contrary, there is a strong evidence of partial ignorance, depending on the image, leading to ambiguities between classes, and this has been introduced by masses on compound hypotheses. Again, this supports the proposed way for assigning masses, which relies on problem modelling, rather than applying a classical probabilistic approach and global reliability factors.

3.6. *Partial volume effect between white matter and ALD*

The images used in our experiments correspond to thick slices. Therefore, an important partial volume effect can be observed, in particular between white matter and ALD (see Fig. 1). We introduce this knowledge as a mass value on $C_1 \cup C_3$ in the second image, which thus explicitly takes into account this partial volume effect. Again, a trapezoidal function has been used, whose parameters are determined based on the histogram, as seen in Fig. 2 (an automatic determination of partial volume effect and associated mass function could be obtained for instance like in (Géraud et al., 1995), using a morphological and statistical analysis of interfaces between tissues). The conflict is thus reduced. The computation of $m_1 \oplus m_2$ shows that it is no longer a Bayesian belief function. The decision areas are changed, as seen in Fig. 7. The computation of these decision areas is similar as the one explained for Figs. 4 and 5. Here, the separation curves between regions have no longer simple linear expressions.

The decision images in Fig. 8 show the decision results obtained respectively on all hypotheses excluding D and on all simple hypotheses, first with $m_2(C_1 \cup C_3) = 0$ and then with increasing value of $m_2(C_1 \cup C_3)$. This figure shows that the decision on all hypotheses includes all partial volume between ALD and WM in $C_1 \cup C_3$, and does not change if the weight affected to $m_2(C_1 \cup C_3)$ increases, proving the robustness with respect to evidence weighting. On the contrary, more partial volume is included in ALD on the decision images over all simple hypotheses. This modelling allows us to mimic the way the physician would make this decision, depending on his objective. On the left most image, where the partial volume effect is not taken into account, the area classified as ALD presents no ambiguity. On the right most image, on the contrary all partial volumes are included in the ALD (this actually corresponds to the segmentation performed by the physicians), and the area classified as brain includes no ambiguous region.

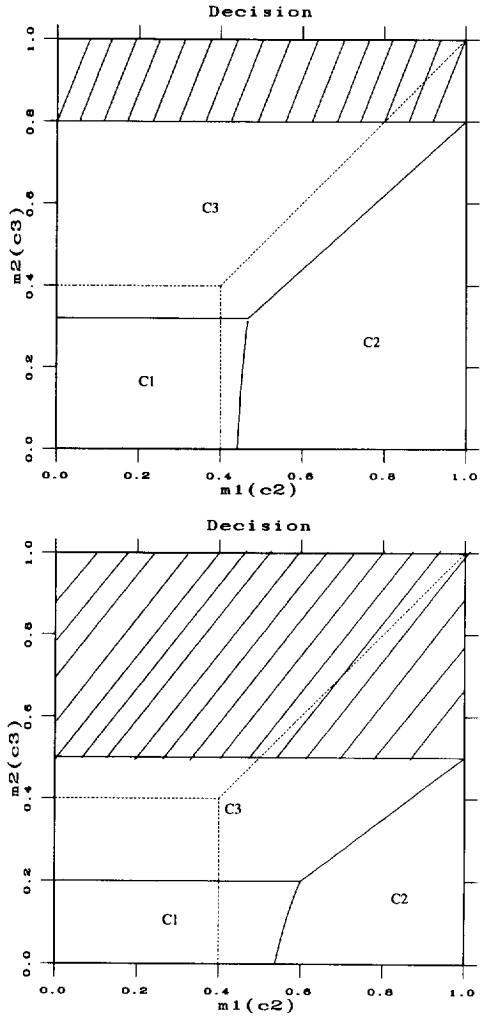

Fig. 7. Decision areas depending on $m_1(C_2)$ and $m_2(C_3)$ with introduction of an increasing mass on $C_1 \cup C_3$ in the second image ($m_2(C_3)$ has to be less than $1 - m_2(C_1 \cup C_3)$). The dashed lines represent the previous limits (see Fig. 5).

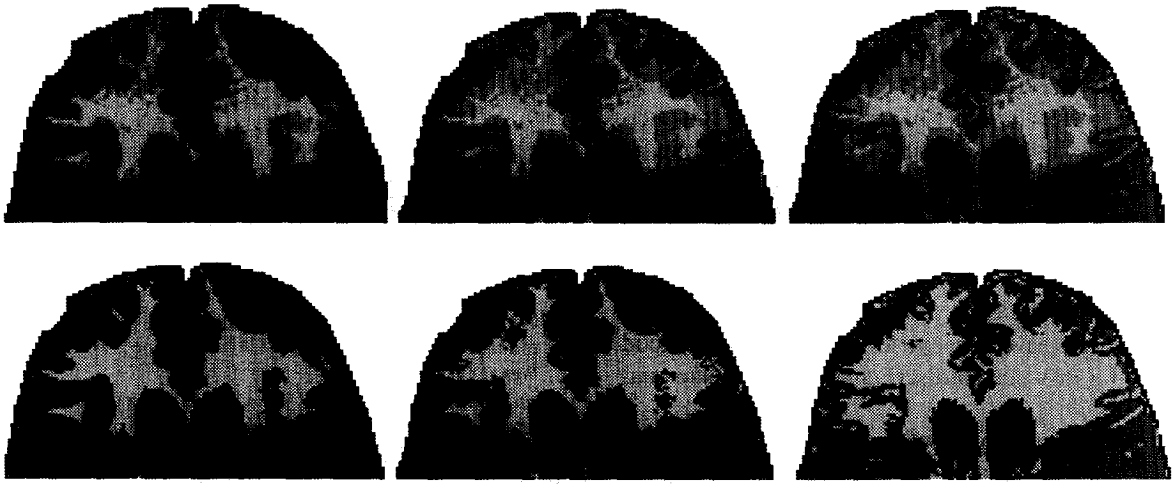

Fig. 8. Decision by maximum of belief over all hypotheses excluding D (first row) and over all simple hypotheses (second row), without mass on $C_1 \cup C_3$ (left) and with introduction of a mass on $C_1 \cup C_3$ (representing the partial volume effect between brain and ALD) in the second image, with increasing weight (middle and right).

4. Conclusion

The large diffusion of Bayesian techniques is due to its sound mathematical bases and to the knowledge issued from the wide experience in modelization or learning. It can be used in complex inference networks and the conditions for correct use are well known and may be checked by statistical tests. But this approach also suffers from severe limits. They have led to a heated debate which will not be discussed here. Let us just mention some arguments. The additivity property, imposing that the probability of an event and of its contrary must add to 1, may be too constraining in medical imaging (Smets, 1978). The use of Bayesian formalism necessitates the knowledge of many pieces of a priori information about the problem at hand, and we are often faced with the limitation of statistical tests relating symptoms or observations to pathologies and with the difficulty to have an estimation of prior probabilities. Moreover, knowledge which cannot be easily modelled by probabilities is often difficult to introduce in the method. Another limit of probabilistic approaches concerns both absence of information and ignorance, which cannot be properly managed in this framework (Shafer, 1976). Furthermore, imprecise information is not easily modelled by probabilities, as opposed to uncertain information. Even subjective probabilities, which are nothing but a formal way, chosen by human beings, for representing events, are not able to overcome all these drawbacks.

We outlined in this paper some features of Dempster–Shafer evidence theory which can be very useful for medical image fusion for classification, segmentation or recognition purposes, and which constitute advantages over classical probabilistic and Bayesian approaches. They include the high flexibility of the modelling offered by DS, taking into account both imprecision and uncertainty, global ignorance or source reliability, ability of sources to provide reliable information or not about each class of interest, and prior information not necessarily expressed as probabilities. These aspects have been illustrated on an example in images of pathological brains acquired with dual echo MRI, where none of the images provides a complete and reliable information, thus making the fusion process necessary. In this example, the advantages of DS approach have been demonstrated.

First of all, an adapted modelization is possible in the DS framework, in particular we proposed to assign masses to compound hypotheses representing the unability of an image to discriminate between two classes on one hand, and partial volume effect on the other hand. Such a modelling already constitutes an improvement over other applications of DS. The reasoning process proposed in this paper for choosing the set of focal elements and for assigning masses is very powerful as long as the number of classes remains low. However, in cases with large numbers of

classes, it would become too tedious, and unsupervised methods are necessary, for instance like the one proposed in (Mascle et al., 1995) for SAR imaging. Since a reduced number of classes is used, computation is very fast (on a Sparc Center 2000). Again, in cases with large numbers of classes, the computation time increases exponentially in the worst case (if all compound hypotheses are taken into account).

Secondly, even a rough definition of mass functions proved to be sufficient and robust. Experiences with more classes have been performed and provide good results, for instance for discriminating grey matter and white matter.

In addition, reliability of an image concerning each class can be introduced by weighting accordingly the corresponding masses, and its influence on the combination and on the conflict has been tested.

Finally, the decisions have been taken according to two rules: a classical one, where a decision is always taken in favour of a simple hypothesis, and a second, original one, where it can also be decided in favour of a compound hypothesis. This second rule has several advantages: it is robust with respect to evidence weighting, as seen when testing different weights on mass functions and on partial volume, and it fits reality by highlighting in particular regions with partial volume effect, and is adapted to the physician's way of reasoning.

References

Andress, K.M. and A.C. Kak (1988). Evidence accumulation and flow control in a hierarchical spatial reasoning system. *AI Magazine*, 75–94.

Appriou, A. (1991). Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 27–40.

Appriou, A. (1993). Formulation et traitement de l'incertain en analyse multi-senseurs. *Quatorzième Colloque GRETSI*, Juan les Pins, 1993, 951–954.

Aubourg, P., C. Adamsbaum, M.C. Lavallard-Rousseau, A. Lemaitre, F. Boureau, M. Mayer and G. Kalifa (1992). Brain MRI and electrophysiologic abnormalities in preclinical and clinical adrenomyeloneuropathy. *Neurology* 42, 85–91.

Aurdal, L., X. Descombes, H. Maître, I. Bloch and C. Adamsbaum (1995). Analysis of adrenoleukodystrophy from dual echo MR images: Automatic segmentation and quantification. *Computer Assisted Radiology CAR'95*, Berlin, June 1995, 35–40.

Baldwin, J.F. (1991). A new approach to inference under uncertainty for knowlegde based systems. In: R. Kruse and P. Siegel, Eds., *Symbolic and Quantitative Approaches to Uncertainty*, Marseille, 1991. Springer, Berlin, 107–114.

Baldwin, J.F. (1992). Inference for information systems containing probabilistic and fuzzy uncertainties. In: L. Zadeh and J. Kacprzyk, Eds., *Fuzzy Logic and the Management of Uncertainty*. Wiley, New York, 353–375.

Barnett, J.A. (1981). Computational methods for a mathematical theory of evidence. *Proc. 7th IJCAI*, Vancouver, 1981, 868–875.

Bloch, I. (1996). Information combination operators for data fusion: A comparative review with classification. *IEEE Trans. Syst. Man Cybernet.* 26, 52–67.

Bloch, I. and H. Maître (1994). Fusion de données en traitement d'images: Modèles d'information et décisions. *Traitement du Signal* 11, 435–446.

Chen, S.Y., W.C. Lin and C.T. Chen (1993). Evidential reasoning based on Dempster–Shafer theory and its application to medical image analysis. *SPIE* 2032, 35–46.

Clarke, M. and N. Wilson (1991). Efficient algorithms for belief functions based on the relationship between belief and probability. In: R. Kruse and P. Siegel, Eds., *Symbolic and Quantitative Approaches to Uncertainty*, Marseille, 1991. Springer, Berlin, 48–52.

Coatrieux, J.L., C. Roux and R. Collorec (1991). Fusion d'informations en imagerie médicale tridimensionnelle. *Bull. de Liaison de la Recherche en Informatique et Automatique* 132, 12–16.

Cucka, P. and A. Rosenfeld (1992). Evidence-based pattern matching relaxation. Technical Report CAR-TR-623, Center of Automation Research, University of Maryland, May 1992.

De Maertelaere, P., P. Ravazzola, P. Ghesquière and A. Beltrando (1993). Architectures et méthodes de fusion pour la classification multi-sources. *Actes du Quatorzième Colloque GRETSI*, Juan-les-Pins, 1993, 983–986.

Denœux, T. (1995). A *k*-nearest neighbor classification rule based on Dempster–Shafer theory. *IEEE Trans. Syst. Man Cybernet.* 25.

Dubois, D. and H. Prade (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.* 4, 244–264.

Garvey, T.D. (1986). Evidential reasoning for land-use classification. In: *Analytical Methods in Remote Sensing for Geographic Information Systems*, Internat. Assoc. Pattern Recognition, Technical Committee 7 Workshop, Paris, October 1986.

Garvey, T.D., J.D. Lowrance and M.A. Fishler (1981). An inference technique for integrating knowledge from disparate sources. *Internat. Joint Conf. on Artificial Intelligence*, 1981, 319–325.

Géraud, T., L. Aurdal, H. Maître, I. Bloch and C. Adamsbaum (1995). Estimation of partial volume effect using spatial context. Application to morphometry in cerebral imaging. *IEEE Medical Imaging Conf.*, San Francisco, CA, October 1995.

Gordon, J. and E.H. Shortliffe (1985). A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence* 26, 323–357.

Guan, J. and D.A. Bell (1991). *Evidence Theory and its Applications*. North-Holland, Amsterdam.

Ip, H.H.S. and J.M.C. Ng (1994). Human face recognition using Dempster–Shafer theory. *ICIP*, Vol. II, Austin, TX, 1994, 292–295.

Lee, R.H. and R. Leahy (1990). Multi-spectral classification of MR images using sensor fusion approaches. *SPIE Medical Imaging IV: Image Processing* 1233, 149–157.

Lee, T., J.A. Richards and P.H. Swain (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Trans. Geoscience Remote Sensing* 25, 283–293.

Lowrance, J.D. (1988). Automatic multisource data analysis. *Internat. Lithosphere Project Research Conf. on Advanced Data Intergration in Mineral and Energy Resource Studies*, Sotogrande, Spain, December 1988.

Lowrance, J.D., T.M. Strat, L.P. Wesleyy, T.D. Garvey, E.H. Ruspini and D.E. Wilkins (1991). The theory, implementation and practice of evidential reasoning. SRI project 5701 final report, SRI, Palo Alto, CA, June 1991.

Mascle, S., I. Bloch and D. Vidal-Madjar (1995). Unsupervised multisource remote sensing classification using Dempster–Shafer evidence theory. *SPIE/EUROPTO Synthetic Aperture Radar and Passive Microwave Sensing* 2584, Paris, September 1995, 200–211.

Neapolitan, R.E. (1992). A survey of uncertain and approximate inference. In: L. Zadeh and J. Kaprzyk, Eds., *Fuzzy Logic for the Management of Uncertainty*. Wiley, New York, 55–82.

Rasoulian, H., W.E. Thompson, L.F. Kazda and R. Parra-Loera (1990). Application of the mathematical theory of evidence to the image cueing and image segmentation problem. *SPIE Signal and Image Processing Systems Performance Evaluation* 1310, 199–206.

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

Smets, P. (1978). Medical diagnosis: Fuzzy sets and degree of belief. *Colloque Internat. sur la Théorie et les Applications des Sous-Ensembles Flous*, Marseille, September 1978.

Strat, T.M. (1989). Decision analysis using belief functions. Technical Note 472, SRI, September 1989.

Suh, D.Y., R.M. Mersereau, R.L. Eisner and R.I. Pettigrew (1990). Automatic boundary detection on cardiac magnetic resonance image sequences for four dimensional visualization of the left ventricle. *First Conf. on Visualization in Biomedical Computing*, Atlanta, GA, 1990, 149–156.

Van Cleynenbreugel, J., S.A. Osinga, F. Fierens, P. Suetens and A. Oosterlinck (1991). Road extraction from multi-temporal satellite images by an evidential reasoning approach. *Pattern Recognition Lett.* 12, 371–380.

Wesley, L.P. (1986). Evidential knowledge-based computer vision. *Opt. Engrg.* 25, 363–379.

Zahzah, E. (1992). Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites. Thèse de Doctorat, Université Paul Sabatier, Toulouse, 1992.