Document Type Recognition Using Evidence Theory

Thierry Géraud[1], Geoffroy Fouquier[1], Quôc Peyrot[1], Nicolas Lucas[2], and Franck Signorile[2]

[1] EPITA Research and Development Laboratory, 14-16 rue Voltaire, F-94276 Le Kremlin-Bicêtre cedex, France,,
thierry.geraud@lrde.epita.fr,
WWW home page: http://www.lrde.epita.fr/
[2] SWT, 21 rue des Genêts, F-94310 Orly, France

Abstract. This paper presents a method for document type recognition when a database of document types is given. To that aim, we define each document type by a set of characteristics whose nature can vary from one to another. For instance, a characteristic can be "having a flower-shaped logo on top-left" as well as "having about 12pt fonts". Since some characteristics are usually featured by several different document types, this recognition problem is not trivial. This paper shows that Boolean logic is not relevant and that fuzzy approaches lead to many false ambiguities. We then describe how to use the mathematical theory of evidence to get successful recognition results.

1 Introduction

Recognizing the type of a document when a database of document types is given is an important issue. For instance, consider an automatic process that inputs document images and that outputs indexing results. If your first step is document type identification, you can use prior information to perform indexing and thus you can expect getting results.

In this paper, we assume that each type is defined by a set of characteristics derived from its description in natural language. For instance, a characteristic can be "having a flower-shaped logo on top-left", "having about 12pt fonts", or "containing a bar code". Let us take a simple example summarized by the table below.

	type 1 (t_1)	type 2 (t_2)	type 3 (t_3)
flower logo (W)	yes	no	no
12pt fonts (F_{12})	yes	no	yes
bar code (B)	no	yes	yes

In this example, type 1 features the flower logo and 12pt fonts but does not feature any bar code. Another way to read this table is to consider one characteristic, say 12pt fonts, to argue that only types 1 and 3 feature that characteristic.

Given an input document image, all characteristics can be checked; for instance, if we get { flower = no; 12pt = no; bar code = yes }, we can state that this particular document is from type 2.

Unfortunately, relying on Boolean logic to take decisions usually does not provide good results. There are three main reasons for that.

- First, when an error occurs while checking a characteristic, this single error leads to either a false decision, or to the impossibility to take any decision at all. For instance, if we get { flower = no, 12pt = no, bar code = no }, no decision can be taken although we can guess that the bar code presence would not have been properly detected and that the document type might be number 2. Another illustration about the error-prone aspect of Boolean logic arises when considering types 2 and 3: an error while checking the 12pt fonts characteristic always leads to a false decision.
- The second reason why Boolean logic is not well suited to our problem is due to the imprecise nature of characteristics. For instance, we never expect to get exactly "12.0pt" when we measure fonts from a document; actually, we want to estimate if fonts are "about 12pt", that is, we want to handle imprecision. Another example comes from the "flower-shaped logo on top-left" characteristic. Since such a vague notion as "top-left" is involved, we cannot reasonably rely on Boolean estimation.
- Last, document images are corrupted by noise and other distortions so there is no guarantee that we can be sure of our estimation (presence or not) of a given feature. For instance, some documents, before being scanned, are manually processed and a stick can be glued upon them. So, to handle the fact that a sticker might hide our flower logo, our property should rather be expressed as "having a flower-shaped logo, but which may not be visible". Boolean logic just cannot handle such uncertainty.

Conversely, fuzzy set theory and evidence theory are well suited to represent both the imprecision and the uncertainty that we have to deal with. The latter theory, also called "mathematical theory of evidence" or Dempster-Shafer's theory, is presented by Shafer (1976) and Guan and Bell (1991). This theory is not as well-spread as fuzzy set theory. However, it has already been applied to several recognition problems that are listed by Sentz and Ferson (2002). In the field of document processing, as far as we know, the only application of this theory has been proposed by Lalmas (1997) about structured document analysis. The application that we present here does not rely on the notion of structure.

In this paper, making schemes using either theory, we present how to take reliable decision about document type from characteristic evaluation. As the paper focuses on the decision process, we do not explain how to select the most relevant features to distinguish between different document types and we do not explain how to estimate the "more or less" presence of a given characteristic in a document. So we state that :

- a document type is described by a set of characteristics;

– evaluating if a document more or less features a characteristic gives a "score" *between* 0 and 1.

The value 0 means that the document does not feature the characteristic at all, whereas the value 1 means that the document totally features the characteristic. An intermediate value represents the "more or less" aspect of the estimation. Back to our example, a feature estimation from a document can be: { flower = 0.9; 12pt = 0.8; bar code = 0.1 }.

This paper is organized as follows. Section 2 describes an approach based on fuzzy set theory. Then, section 3 is dedicated to solutions relying on the mathematical theory of evidence. Last we conclude in section 4.

2 Fuzzy Approach

2.1 Fuzzy Set Theory

In fuzzy set theory, we first need a set. Since our problem is document type recognition, let us denote by D the set of documents. The subsets of D are considered to be fuzzy. Let us denote S_i a subset such as $\forall i,\ S_i \subset D$ and $\cup_i S_i = D$. Given a document $d \in D$, we then have fuzzy membership values: $\forall i,\ \mu_{S_i}(d) \in [0,1]$ and $\sum_i \mu_{S_i}(d) = 1$, where $\mu_{S_i}(d)$ denotes the degree of d belonging to S_i.

If we have n different types of documents, we can define their corresponding sets: $t_1, \ldots, t_n$, and we have $\forall i = 0..n,\ t_i \subset D$. Given a document d, we are interested in calculating the values $\mu_{t_0}(d), \ldots, \mu_{t_n}(d)$ in order to take a decision. By contrast with the Boolean logic approach, we handle nothing but fuzzy values until the final decision step so that we keep as much information as possible during our computation.

Remembering that we deal with characteristics, we can define, for each characteristic, a set with its proper fuzzy subsets:

$$\begin{aligned} W &= W_{yes} \cup W_{no} \\ F &= F_{<12} \cup F_{12} \cup F_{>12} \\ B &= B_{yes} \cup B_{no} \end{aligned}$$

where W_{yes} is the subset "having a flower-shaped logo on top-left", F_{12} is the subset "having about 12pt fonts", B_{no} is the subset "having *no* bar code", and so on. For a given document d, each characteristic is evaluated into fuzzy membership values.

A first way to link D with characteristics is to consider that $D = W \times F \times B$ and then to derive the definitions of types. For instance, according to the table given in section 1: $t_1 = W_{yes} \times F_{12} \times B_{no}$. This modeling stresses that characteristics are independent and that each document is valuated into a vector of three components. For instance, we have $d = (\ 0.9_{/W_{yes}} + 0.1_{/W_{no}},\ \ 0.2_{/F_{<12}} + 0.8_{/F_{12}} + 0.0_{/F_{>12}},\ \ 0.1_{/B_{yes}} + 0.9_{/B_{no}}\ )$.

A second way of modeling our problem is to consider that $D = W = F = B$, which means that several fuzzy partitions are defined over D. Continuing with the same example, we have $t_1 = W_{yes} \cap F_{12} \cap B_{no}$, and $d = 0.9_{/W_{yes}} + 0.1_{/W_{no}} = 0.2_{/F_{<12}} + 0.8_{/F_{12}} + 0.0_{/F_{>12}} = 0.1_{/B_{yes}} + 0.9_{/B_{no}}$.

With both models (Cartesian product and partitions), fuzzy set theory leads to the same conclusion:

$$\mu_{t_1}(d) = \min(\ \mu_{W_{yes}}(d),\ \mu_{F_{12}}(d),\ \mu_{B_{no}}(d)\ ).$$

Last, following our numerical example, we obtain $d = 0.8_{/t_1} + 0.1_{/t_2} + 0.1_{/t_3}$.

More generally, if we have k independent characteristics, let us note c^j a characteristic (where $j = 1..k$) and c^j_i the subset of c^j corresponding to t_i. For instance, in the previous example, $c^1 = W, c^2 = F$, $c^3 = B$, $c^1_1 = W_{yes}, c^2_1 = F_{12}$, and $c^3_1 = B_{no}$. The resulting membership values w.r.t. document types are:

$$\forall d \in D,\ \forall i = 0..n,\ \ \mu_{t_i}(d)\ =\ \min_{j=1..k}\ \mu_{c^j_i}(d). \tag{1}$$

A very simple decision rule is to assign a document to the type which gives the greatest membership value:

$$\forall d \in D,\ \ \omega(d) = \arg\ \max_{i=1..n}\ \mu_{t_i}(d). \tag{2}$$

where ω denotes the decision function. By extension, we define the following function:

$$\forall d \in D,\ \ \omega_2(d) = \arg\ \max_{i \in [1,n] - \omega(d)}\ \mu_{t_i}(d), \tag{3}$$

which gives the second "best" decision.

Some other decision rules are often used, taking into account the fact that we sometimes prefer *not* to take any decision instead of taking an erroneous one. This situation happens when $\mu_{t_{\omega(d)}}$ is too low, that is, when we are not sure to have recognized a type of our database. It also happens when the difference $\mu_{t_{\omega(d)}} - \mu_{t_{\omega_2(d)}}$ is too low, that is, when a too strong ambiguity in taking a decision is noticed.

Actually the fuzzy *set* approach is a particular case of fuzzy *fusion* presented hereafter.

2.2 Fuzzy Fusion Schemes

Our problem can be formulated in a different way. Each characteristic gives a clue to decide that a document belongs to a particular type; an evaluation of a document featuring a characteristic is a piece of information. Finally we can state that a characteristic is nothing but a source of information and that our problem consists in fusing the different sources in order to take a decision. An allegory to understand the difference between this new approach and the one of previous section is the following. If we were in a probabilistic context, taking a decision could be either a direct result of our problem modeling (such as in the previous section) or an *estimation* problem from input data (such as in this present section).

When we reread equation (1) from this point of view, it happens that it is a fusion formula which follows the general fusion pattern:

$$\forall d \in D,\ \forall i = 0..n,\ \ \mu_{t_i}(d) \ = \ \oplus_{j=1..k}\, \mu_{c_i^j}(d), \tag{4}$$

where $\oplus$ symbolizes any fuzzy fusion operator. We can then choose an operator amongst the wide set of fuzzy fusion operators listed by Bloch (1996). Our choice depends on the behavior expected from the fusion.

This operator can be conjunctive, which translates the following idea: "deciding to assign d to t_i means that we *simultaneously* strongly recognize all features c_i^j in document d". Conjunctive operators are T-norms and verify $\oplus \leq min$. As one can notice, min falls in this category and the fuzzy model of section 2.1 is a particular case of conjunctive fuzzy fusion. Conjunctive operators are *severe* since all characteristics should be well recognized to get an unambiguous document type identification. However, we assume that sometimes some characteristics cannot be retrieved, for instance due to stickers hiding parts of the documents. Thus, such severe operators would not be tolerant enough *vis-à-vis* false estimations of characteristic presence and many results would be ambiguous. Put differently, these operators are not well suited to handle strong uncertainty.

Though, the operator behavior can be less severe than a conjunction; it can be a compromise, which corresponds to another idea: "deciding to assign d to t_i means that we *globally* properly recognize all features c_i^j in document d". Compromise operators are means and verify min $< \oplus <$ max. Simple compromise operators are the arithmetical mean (denoted by $+mean$ later on) and the geometrical mean. Their behavior is definitively appropriate to take into account uncertainty in our document type recognition problem: these operators are tolerant to estimation errors.

2.3 Limitations of Fuzzy Approach

The fuzzy framework presented in the previous section unfortunately does *not* take into account that some characteristics can be shared by different types, and conversely that some other characteristics are really specific for their respective types. Actually, we are not able to introduce such information *explicitly* into this framework.

For instance, if we compare both fuzzy fusions respectively dedicated to t_2 and t_3:

$$\begin{aligned} \mu_{t_2} &= \oplus(\ \mu_{W_{no}}, 1 - \mu_{F_{12}}, \mu_{B_{yes}}\ ) \\ \mu_{t_3} &= \oplus(\ \mu_{W_{no}}, \quad \mu_{F_{12}}, \quad \mu_{B_{yes}}\ ) \end{aligned}$$

we see that information about shared and specific characteristics are *present* implicitly in formulas, but there is no way to emphasize on the fact that the second characteristic ($c^2 = F$) is crucial to distinguish between type 2 and type 3. If we have:

$$\mu_{W_{no}} = 0.9; \quad \mu_{F_{12}} = 0.8; \quad \mu_{B_{yes}} = 0.7 \qquad \textit{case 1} \tag{5}$$

then, with $\oplus$ being an arithmetical mean, we obtain $\mu_{t_2} = 0.6$ and $\mu_{t_3} = 0.8$ and we can state that the decision is ambiguous. Though, choosing t_3 seems obvious since the *only* difference between both types comes from $\mu_{c_3^2} = 0.8$, whereas $\mu_{c_2^2} = 0.2$.

In that particular case, the fact that the fusion operator is a compromise does not help avoiding ambiguity. However, another example with $\oplus$ being a disjunctive operator (min) can easily be settled:

$$\mu_{W_{no}} = 0.8; \quad \mu_{F_{12}} = 0.7; \quad \mu_{B_{yes}} = 0.5 \qquad \textit{case 2} \tag{6}$$

gives $\mu_{t_2} = 0.3$ and $\mu_{t_3} = 0.5$. This is another case of ambiguity, whereas the decision should be dictated only by $\mu_{c_3^2} = 0.7$ being much greater than $\mu_{c_2^2} = 0.3$.

The fuzzy framework thus leads to poor recognition results. Ambiguities occur even when simple logic rules can tell that there cannot be any ambiguity. This drawback comes from the fact that we handle each document type separately (Cf. equation (1)), that we consider and valuate each piece of information independently. This approach is not relevant since different types can have some characteristics in common.

A simple text-book example is the following. Considering the set of people *Greg*, *Jack*, and *Tom*, somebody says: “I can’t remember who’s the biggest fool but I’m positive that it’s either *Greg* or *Tom*”. With fuzzy set theory, we can model this assertion by $0.5_{/Greg} + 0.5_{/Tom} + 0_{/Jack}$. However, this is unsatisfactory because having a membership degree of 0.5 for *Greg* means that he’s just “half” a fool and this is definitely *not* what we meant. Rather, a proper translation of the assertion is $1_{/(Greg \text{ or } Tom)} + 0_{/Jack}$ but it is then out of the scope of fuzzy set theory.

3 Evidence Theory Approach

Evidence theory, also called Dempster-Shafer theory, has been built to handle situations such as the “*Greg*, *Jack*, and *Tom*’s case” in the previous section.

3.1 Basics of Mathematical Theory of Evidence

The hypothesis set Θ, also called “frame of discernment”, represents a set of mutually exclusive and exhaustive propositions; in our case, $\Theta = \{\, t_1, \ldots, t_n \,\}$. By extension of the set theory, inclusion, intersection, and union of a couple of hypotheses are defined as follows:

$$\begin{cases} A \subseteq B & \Leftrightarrow \text{ if } A \text{ is true, then } B \text{ is true} \\ (A \cap B) \text{ is true} & \Leftrightarrow A \text{ is true and } B \text{ is true} \\ (A \cup B) \text{ is true} & \Leftrightarrow A \text{ is true or } B \text{ is true .} \end{cases} \tag{7}$$

Evidence on a subset A of Θ is valued with a mass $m(A)$. m is said to be a mass function.

Nota bene: To stick to common notation of evidence theory, we omit the fact that these valuations depend on the document being considered. Precisely, we should have written $m(A)(d)$ instead of just $m(A)$. In the following, every expression formed as $\mathit{something}(A)$ with $A \subset \Theta$ actually means that a document d is given and thus should be understood as $\mathit{something}(A)(d)$. This notation simplification is due to the fact that valuations in evidence theory usually focus on *one* observation. In our case, this observation is a document and the problem of assigning a type occurs for *each* document d.

Subsets of Θ with non null masses are called focal elements and compose the kernel of the mass function. We have the following properties:

$$\begin{cases} \forall A \subset \Theta, \;\; m(A) \in [0,1] \\ \quad \sum_{A \subset \Theta} m(A) = 1 \\ \qquad\quad m(\emptyset) = 0. \end{cases} \tag{8}$$

The belief function, $\mathit{bel} : A \subset \Theta \rightarrow \mathit{bel}(A)$, represents the amount of evidence which implies A:

$$\mathit{bel}(A) = \sum_{B \subset A} m(B).$$

The plausibility function, $\mathit{pls} : A \subset \Theta \rightarrow \mathit{pls}(A)$, represents the amount of evidence that does not refute A:

$$\mathit{pls}(A) = 1 - \mathit{bel}(\overline{A}) = \sum_{B,\, B \cap A \neq \emptyset} m(B)$$

where $\overline{A}$ is the complementary hypothesis of A, that is, $\Theta - A$. We have $\forall A \subset \Theta$, $0 \leq \mathit{bel}(A) \leq \mathit{pls}(A) \leq 1$. Another way to interpret the meaning of plausibility $\mathit{pls}(A)$ is to think that it is the maximum uncertainty value of A. The interval $[\, \mathit{bel}(A), \mathit{pls}(A)\,]$ represents therefore the uncertainty about A; it is called the belief interval and allows us to define ignorance:

$$\mathit{ign}(A) = \mathit{pls}(A) - \mathit{bel}(A).$$

Last, the doubt about A is the amount of evidence that does refute A:

$$\mathit{dou}(A) = \mathit{bel}(\overline{A}).$$

When several information sources give evidence about the same set of hypotheses Θ, their respective mass functions are defined, say $m_1, \ldots, m_s$. These masses can be combined to fuse information and to get a *single* mass that owns the knowledge of the whole set of sources. To that aim, Dempster has proposed a combination rule, also called orthogonal sum, and denoted by "$\oplus$" by Shafer (1976). First, a measure of conflict between sources is calculated:

$$K = \sum_{\cap_{i=1}^{s} B_i = \emptyset} \left(\prod_{i=1}^{s} m_i(B_i) \right).$$

This represents the mass that would be assigned to the empty set after combination. We have $0 \leq K \leq 1$. If $K = 1$ the sources are totally contradictory and their combination thus has no sense at all. The lower K is, the more their combination makes sense. Last, the mass combination is defined as follows:

$$\text{if } K \neq 1, \quad m_1 \oplus \ldots \oplus m_s(A) \; = \; \frac{1}{1-K} \sum_{\cap_{i=1}^{s} B_i = A} \left(\prod_{i=1}^{s} m_i(B_i) \right).$$

To make these formulas clear, the combination of two masses leads to:

$$m_1 \oplus m_2(A) \; = \; \frac{\displaystyle\sum_{B_1 \subset A \text{ and } B_2 \subset A \text{ such as } B_1 \cap B_2 = A} m_1(B_1)\, m_2(B_2)}{1 \; - \displaystyle\sum_{B_1 \subset A \text{ and } B_2 \subset A \text{ such as } B_1 \cap B_2 = \emptyset} m_1(B_1)\, m_2(B_2)}.$$

Dempsters's combination rule has very strong properties. First, the result $m_1 \oplus \ldots \oplus m_s$ is a mass function, that is, it verifies the properties given by equation (8). Second, this combination rule is commutative and associative. Other algebraic properties are given by Guan and Bell (1991).

Extra information about evidence theory and recent advances concerning this theory are available thanks to Yager et al. (1994) and Lee and Zhu (95).

3.2 Evidence and Characteristics

A mass function is bound to a particular source of information, a characteristic in our case. Continuing with our first example (Cf. the table of section 1), we then have one mass function per characteristic. Until now, we have read the table column per column and we have ended up with formulas such as equation (1) in section 2.1 and equation (4) in section 2.2. The point of view enlightened by the evidence theory is now different since we are first interested by the behavior of each source / characteristic with respect to our hypotheses / document types, that is, a reading of the table rows.

The first row tells that the "flower logo" only appears on documents from type 1. A corresponding mass, m_W, should then be defined upon the singleton subset $\{t_1\} \subset \Theta$. This mass value can be derived from the equivalent fuzzy membership degree introduced in section 2.1:

$$m_W(\{t_1\}) = \mu_{W_{yes}}.$$

When the flower-shaped logo is not present in a document, this can be explained *either* by this document not being from type 1 *or*, as noticed in section 1, by this logo actually being there but hidden by a sticker. This uncertainty leads to the following funny assertion: "when it is not t_1, it is either t_1, t_2, or t_3"! This statement is modeled by:

$$m_W(\Theta) = 1 - \mu_{W_{yes}}.$$

An interpretation based on set theory relies on the definitions given by equation (7). Since $\Theta = \{t_1\} \cup \{t_2\} \cup \{t_3\}$, having a non-null mass for Θ means that either $\{t_1\}$, $\{t_2\}$, or $\{t_3\}$ can be true.

We proceed identically for $m_{F_{12}}$ and m_B:

$$\begin{aligned} m_{F_{12}}(\{t_1, t_3\}) &= \mu_{F_{12}} \\ m_{F_{12}}(\Theta) &= 1 - \mu_{F_{12}} \\ m_B(\{t_2, t_3\}) &= \mu_{B_{yes}} \\ m_B(\Theta) &= 1 - \mu_{B_{yes}}. \end{aligned}$$

Then we fuse the three sources:

$$m_u = m_W \oplus m_{F_{12}} \oplus m_B,$$

and last, for every singletons, $\{t_1\}$, $\{t_2\}$, and $\{t_3\}$, we compute the belief and plausibility values from the mass function m_u. For instance, we obtain the value $bel_a(\{t_1\})$ that gives the final amount of evidence with implies type 1. We then just have to decide to assign a type to the document or to state that its type is unknown. The four most popular decision rules are the following:

- maximum of belief;
- maximum of plausibility;
- maximum of belief without overlapping of belief intervals (also called absolute decision rule);
- maximum of $(bel + pls)/2$, which is a compromise, conversely to the previous rules which are conjunctive.

Let us mention another evidential modeling of our problem. If we do *not* want to take into account a global uncertainty, we just have to never valuate masses for Θ. This approach is then more conventional but also less robust to handle difficult cases such as the presence of stickers. Without global uncertainty, we have assertions such as "when it is not t_1, it is either t_2 or t_3". Therefore this alternate modeling is as follows:

$$\begin{aligned} m_W(\{t_1\}) &= \mu_{W_{yes}} \\ m_W(\{t_2, t_3\}) &= 1 - \mu_{W_{yes}} \end{aligned}$$

$$\begin{aligned} m_{F_{12}}(\{t_1, t_3\}) &= \mu_{F_{12}} \\ m_{F_{12}}(\{t_2\}) &= 1 - \mu_{F_{12}} \\ m_B(\{t_2, t_3\}) &= \mu_{B_{yes}} \\ m_B(\{t_1\}) &= 1 - \mu_{B_{yes}} \end{aligned}$$

and finally:

$$m_{\not u} = m_W \oplus m_{F_{12}} \oplus m_B. \tag{9}$$

3.3 Comparative Results

Table 1. Results with mass function m_u.

case 1

	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$
belief	0.03	0.00	0.54
plausibility	0.32	0.19	0.97
$(bel + pls)/2$	0.18	0.10	0.75
fuzzy $+mean$	0.40	0.60	0.80

case 2

	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$
belief	0.11	0.00	0.31
plausibility	0.56	0.27	0.89
$(bel + pls)/2$	0.33	0.13	0.60
fuzzy $+mean$	0.47	0.53	0.67

Results are depicted by table 1 and correspond to the cases respectively given by equations (5) and (6) in section 2.3. As one can notice, $bel_a(\{t_3\})$, equal to 0.54, is much greater than other belief values and the same goes for $pls_a(\{t_3\})$, equal to 0.97, as compared with other plausibility values. Finally, with both numerical examples, modeling our problem with evidence theory leads to no ambiguity for both cases; both decisions are "type t_3".

With fuzzy fusion (sections 2.2 and 2.3), both examples led to "false" ambiguities. The couple of two last lines in table 1 give the values obtained for each singleton when proceeding to compromises. Line "$(bel + pls)/2$" is an evidential compromise, whereas line "fuzzy $+mean$" is a fuzzy compromise. Results depict that the evidential approach is far better than the fuzzy one. For instance, we can compare the difference between the best value and the second best value, obtained with the evidential compromise:

$$\delta_{evidence}(d) \;=\; \frac{bel + pls}{2}(\{t_{\omega(d)}\})(d) \;-\; \frac{bel + pls}{2}(\{t_{\omega_2(d)}\})(d)$$

and the equivalent difference obtained with the fuzzy compromise:

$$\delta_{fuzzy}(d) \;=\; +mean(\{t_{\omega(d)}\})(d) \;-\; +mean(\{t_{\omega_2(d)}\})(d)$$

with ω and ω_2 as defined by equations (2) and (3). These differences estimate the degree of unambiguity in taking final decisions. With d_1 and d_2 denoting respectively the document of case 1 and the document of case 2, it comes:

$$\begin{aligned}
\delta_{evidence}(d_1) &= 0.75 - 0.18 = 0.57\\
\delta_{fuzzy}(d_1) &= 0.80 - 0.60 = 0.20\\
\delta_{evidence}(d_2) &= 0.60 - 0.33 = 0.27\\
\delta_{fuzzy}(d_2) &= 0.67 - 0.53 = 0.14.
\end{aligned}$$

We finally observe that disambiguation is better —degrees of unambiguity are greater— with the evidential fusion scheme than with the fuzzy one.

Table 2. Extra Results.

case 1						
	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$	$\{t_1,t_3\}$	$\{t_2,t_3\}$	$\{t_1,t_2,t_3\}$
m_u	0.03	0.00	0.54	0.23	0.14	0.06
$m_{\not u}$	0.04	0.19	0.77	0.00	0.00	0.00
μ	0.22	0.33	0.44	undef	undef	undef

case 2						
	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$	$\{t_1,t_3\}$	$\{t_2,t_3\}$	$\{t_1,t_2,t_3\}$
m_u	0.11	0.00	0.31	0.31	0.13	0.13
$m_{\not u}$	0.15	0.26	0.60	0.00	0.00	0.00
μ	0.28	0.32	0.40	undef	undef	undef

Table 2 compares the results obtained for:

- m_u, our vanilla mass function;
- $m_{\not u}$, the mass function obtained when considering that there is no global uncertainty, see equation (9);
- μ, the normalized fuzzy membership function derived from the results of the fuzzy "$+mean$" fusion.

As one can notice, m_u is valuated for subsets that are *not* singletons, which is not the case of $m_{\not u}$ since this scheme does not handle uncertainty. Last, the normalized fuzzy membership function provides poor results.

4 Conclusion

In this paper, we have shown that the mathematical theory of evidence is highly relevant to perform document type recognition when document types are described by imprecise characteristics and when some characteristics are featured by several document types. We have applied this theory onto an effective document image database —several thousands documents and about one hundred different document types. Recognition results are about perfect even for documents presenting heavy defects. This is due to the fact that imprecision and uncertainty are properly handled by the evidential information fusion.

Implementation Issues. We provide a general C++ library, `eVidenZ`, dedicated to experiments with the mathematical theory of evidence. We also provide a generic image processing library, `olena`; information about it are given by Darbon et al. (2002). Both libraries are free software under the GNU Public Licence (GPL) and can be downloaded from our web site:

`http://www.lrde.epita.fr`

Acknowledgements The authors would like to thank David Lesage and Nicolas Burrus for having developed the `eVidenZ` library, the `olena` past and present developing teams for their great work, and Didier Verna for his valuable comments on this paper.

Bibliography

I. Bloch. Information combination operators for data fusion: A comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1):52–67, 1996.

J. Darbon, T. Géraud, and A. Duret-Lutz. Generic implementation of morphological image operators. In *Mathematical Morphology, Proceedings of the 6th International Symposium VI (ISMM)*, pages 175–184, Sydney, Australia, April 2002. Sciro Publishing.

J. Guan and D. Bell. *Evidence Theory and its Applications.* North-Holland, New-York, 1991.

M. Lalmas. Dempster-shafer's theory of evidence applied to structured documents: Modelling uncertainty. In *Proceedings of the 20th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 110–118, Philadelphia, PA, USA, July 1997. ACM.

E. S. Lee and Q. Zhu. *Fuzzy and Evidence Reasoning.* Studies in Fuzziness. Physica-Verlag, 95.

K. Sentz and S. Ferson. Combination of evidence in dempster-shafer theory. Technical Report SAND 2002-0835, Thomas J. Watson School of Engineering and Applied Science, Binghamton University, New-York, USA, April 2002. URL http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf.

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

R. R. Yager, M. Fedrizzi, and J. Kacprzyk, editors. *Advances in the Dempster-Shafer Theory of Evidence.* Wiley Professional Computing. Wiley, 1994.