Target Identification Based on the Transferable Belief Model Interpretation of Dempster-Shafer Model. Pars II: Applications.

François Delmotte
LAMIH, Université de Valenciennes
Le Mont Houy, 59313 Valenciennes Cedex 9, FRANCE
francois.delmotte@univ-valenciennes.fr
and
Philippe Smets
IRIDIA - Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6, 1050 Bruxelles, Belgium
psmets@ulb.ac.be
http://iridia.ulb.ac.be/˜psmets

Abstract

This paper explains how multisensor data fusion and target identification can be performed within the transferable belief model, a model for the representation of quantified uncertainty based on belief functions. In the first part, we present the underlying theory. Here, we present some illustrative examples to clarify how we consider the transferable belief model should be used. Some simulations are presented in order to compare the speed of these algorithms based on belief functions with the more classical ones based on probability functions. The results show that the transferable belief model approach is in fact never less efficient than the probability approach, contrary to previously published results. The results presented here can be extended directly to many problems of data fusion and diagnosis.

Keywords: Belief functions, transferable belief model, General Bayesian Theorem, pignistic probabilities, target identification.

I. Introduction

In the first part of this paper, we have introduced the transferable belief model a model for the representation of quantified beliefs based on belief functions. We have explained what are the essential tools used in the TBM for problems of target identification even though they apply directly to data fusion and diagnosis problems. The tools are the General Bayesian Theorem and the pignistic transformation. Our model mimics the probabilistic approach except that every probability function is replaced by a belief function. The latter being much more general than the former, we can handle degrees of uncertainty hard to represent in probability theory. In particular we can represent the state of total ignorance, what provides a solution to the problem of choosing the adequate prior in the diagnosis process. With the TBM, a prior representing total ignorance is available and can be used directly. Of course if justified priors are available, they are included in our model that degrades nicely into the classical probability approach when all the ingredients needed for such an analysis are available.

In the second part, section II, we show through an example that the General Bayesian Theorem and the probability solutions can be diametrically opposed indicating thus that the choice of the model is not just an academic exercise, but an issue which consequences can be very serious. In section III, we reproduce in part the study presented in [1], [2] and show that the TBM approach is computationally never less efficient that its probability analogous. In the last three sections, we show how to use the TBM for multisensor target identification problems. We consider three cases of sequentially observed sensors which results refine the previous ones. In Section IV, the likelihoods are known on non overlapping subsets of the initial frame. In Section V, the likelihoods are known for every element of a partition of the initial frame, but the granularity of the partition is finer at each level. In Section VI, the likelihoods are generated by each sensor on partially overlapping subsets of the overall frame. We conclude in Section VII.

For easier cross-referencing, equation numbers proceed throughout the two parts. Section references starting with I- refer to part I of this paper.

II. Example 2. An Embarrassing Example Comparing the TBM with a Probability Approach

We present now an example where the probability and the TBM approaches strongly disagree. This example is useful is showing that the choice between the two models can be essential in practice, and not just an intellectual game. The example cannot be used to 'prove' that one of the two models is right or wrong, as neither common sense nor rationality requirements can definitively help us in deciding which of the two diverging conclusions we derive is the 'correct' one. Some hints about this choice are discussed at the end of this section.

A. The Problem.

Suppose a set of sensors $S_1, S_2 \dots S_K$ which purpose is to identify friends from foes. For each sensor, we know what is the probability that it is in working condition or broken. Let $p_i = P(S_i \; in \; Working \; Condition)$.

Let X_i be the domain of the data sensor S_i can collect, with $X_i = \{x_i, y_i\}$. When in working condition, the sensor is a perfect detector, i.e., x_i is equivalent to $Friend$ and y_i to Foe. When broken, we totally ignore how the sensor would react when observing a friend or a foe.

For simplicity sake, suppose we have 30 sensors, and all sensors are either of high quality (HQS) or of low quality (LQS), with sensors $S_1 \dots S_{11} \in HQS$ whereas $S_{12} \dots S_{30} \in LQS$. For each $S_i \in HQS$, $p_i = .99$ and for each $S_i \in LQS$, $p_i = .90$.

The collected data is the vector $data = (x_1 \dots x_{10}, y_{11}, x_{12}, y_{13} \dots y_{30})$, i.e., x_i for $i \in$

$\{1, \ldots 10, 12\}$ and y_i for $i \in \{11, 13 \ldots 30\}$. So 10 out of the 11 high quality sensors and one among the 19 low quality sensors support the target is a friend, the others support it is a foe. What should we conclude after fusing these data? It is hard to decide as common sense can hardly help. We show now that the TBM concludes with a probability of .91 that the target is a friend, whereas the probability analysis concludes with a probability .92 that it is a foe. Values were of course chosen in order to get this enormous discrepancy. Nevertheless such a discrepancy is quite embarrassing. It shows that selecting the model deserves serious attention as conclusions can strongly depend on this choice.

Target	Frd			Foe		
S_i Status	Wrk	Brk		Wrk	Brk	
$P(Status)$	p_i	$1 - p_i$		p_i	$1 - p_i$	
X_i	$p(d_i)$	$p(d_i)$	$p(d_i \vert Frd)$	$p(d_i)$	$p(d_i)$	$p(d_i \vert Foe)$
x_i	1	.5	$p_i + .5(1 - p_i)$	0	.5	$.5(1 - p_i)$
y_i	0	.5	$.5(1 - p_i)$	1	.5	$p_i + .5(1 - p_i)$

TABLE I

EXAMPLE 2. PROBABILITY APPROACH. FOR SENSOR S_i, VALUES OF $p(d_i)$ WHERE THE TARGET T CAN BE Frd (FOR FRIEND) OR Foe AND S_i STATUS CAN BE Wrk (FOR WORKING) AND Brk (FOR BROKEN). p_i IS THE PROBABILITY SENSOR S_i STATUS IS Wrk. THE COLUMNS $p(d_i|Frd)$ AND $p(d_i|Foe)$ PRESENT THE PROBABILITY ON X_i GIVEN THE TARGET IS Frd OR Foe, THUS THE LIKELIHOODS GIVEN TO THE TARGETS WHEN THE OBSERVED DATA IS x_i OR y_i, RESPECTIVELY.

B. Bayesian Analysis.

In order to proceed with a Bayesian analysis, we need first to assess $\alpha = P(x_i|S_i = Broken, Friend) = P(x_i|S_i = Broken, Foe)$, where we accept that the behavior of the broken sensor does not depend on the target's nature. A strict Bayesian claims that a probability can be assigned to any event, and thus that α can and must be assessed. The most natural assessment here (and the one most users will apply in practice) is $\alpha = .5$.

The Bayesian analysis proceeds then as follow. The relevant data are presented in Table I. We must compute $P(Friend|data)$ where $data$ is the $data = (d_1 \ldots d_{30})$ vector . We have:

$$\begin{aligned} P(Friend|data) &\propto P(Friend)P(data|Friend) \\ &= P(Friend) \prod_{i=1...30} P(d_i|Friend) \end{aligned}$$

assuming the conditional independence of the data given the nature of the target.

We have then:

$$\begin{aligned} P(d_i|Friend) &= P(d_i|Friend, S_i = Working)P(S_i = Working|Friend) \\ &\quad + P(d_i|Friend, S_i = Broken)P(S_i = Broken|Friend) \\ &= P(d_i|Friend, S_i = Working)P(S_i = Working) \\ &\quad + P(d_i|Friend, S_i = Broken)P(S_i = Broken) \\ &= P(d_i|Friend, S_i = Working)p_i \\ &\quad + P(d_i|Friend, S_i = Broken)(1 - p_i) \end{aligned}$$

assuming the fact the sensor is in working condition or not is independent of the nature of the target. These values are displayed in the columns $p(d_i|Frd)$ and $p(d_i|Foe)$ of Table I. Their numerical values are in the present example:

$$\begin{aligned} P(d_i|Friend) &= 1 \times .99 + .5 \times .01 = .995 && \text{if } S_i \in HQS \text{ and } d_i = x_i \\ &= .5 \times .01 = .005 && \text{if } S_i \in HQS \text{ and } d_i = y_i \\ &= 1 \times .90 + .5 \times .10 = .95 && \text{if } S_i \in LQS \text{ and } d_i = x_i \\ &= .5 \times .10 = .05 && \text{if } S_i \in LQS \text{ and } d_i = y_i \end{aligned}$$

Given the observed data, we have:

$$P(Friend|data) \propto P(Friend) \times .995^{10} \times .005^{1} \times .95^{1} \times .05^{18} = 1.72339E - 26$$

Identically with foe, we get:

$$P(Foe|data) \propto P(Foe) \times .005^{10} \times .995^{1} \times .05^{1} \times .95^{18} = 1.92983E - 25$$

Assuming equi prior probability of Friend and Foe ($P(Friend) = P(Foe) = .5$), we get:

$$P(Friend|data) = \frac{1.72339E-26}{1.72339E-26 + 1.92983E-25} = 0.08$$
$$P(Foe|data) = \frac{1.92983E-25}{1.72339E-26 + 1.92983E-25} = 0.92$$

Hence the probability analysis using $P(x_i|S_i = Broken, Friend) = P(x_i|S_i = Broken, Foe) = .5$ and an equi *a priori* probability on the nature of the target leads to the conclusion that the target is a foe.

Strict Bayesian might argue that he $\alpha = .5$ was not correct and that another value for α must be used. This is not a serious problem here, as once the α is determined, it is always possible to find a set of data so that the Bayesian and the TBM conclusions will diverge as strongly as here.

One might argue that this strict Bayesian analysis is not an adequate probability analysis, and that we should perform a sensitivity analysis, i.e., we must consider all possible values for α. The result is totally uninformative as we get $P(Friend|data)$ varying from 0 (when $\alpha = 1$) to 1 (when $\alpha = 0$). So the sensitivity analysis leaves the user totally at lost: all the data support is that the probability that the target is a $Friend$ is anywhere in $[0, 1]$, a truism of course.

C. TBM Analysis.

The TBM analysis leads to a conclusion opposite to the one reached by the strict Bayesian (and the practical - some would say naive - user). It proceeds as follows. We must build the plausibility over $\{Friend, Foe\}$ given each type of observation and each sensor quality (see Table II). In the 'working' case, the sensors are perfect, hence the mass 1 on $Friend$ with x_i and on Foe with y_i. When the sensor is 'broken', we are in a state of total ignorance about what might be the target, hence a mass 1 is given to $\{x_i, y_i\}$ for both possible data. Table II presents the corresponding plausibility functions on X_i (columns pl^{X_i}), and the values of $pl^{X_i}[Frd]$ and $pl^{X_i}[Foe]$ taking into consideration the p_i values (they are the weighted average of the former).

In order to combine the data, we compute $pl^H[data] = \textcircled{\cap}_{1\ldots30} pl^H[d_i]$. This is easily

Target	Frd			Foe		
S_i Status	Wrk	Brk		Wrk	Brk	
$P(Status)$	p_i	$1-p_i$		p_i	$1-p_i$	
X_i	pl^{X_i}	pl^{X_i}	$pl^{X_i}[Frd]$	pl^{X_i}	pl^{X_i}	$pl^{X_i}[Foe]$
x_i	1	1	1	0	1	$1-p_i$
y_i	0	1	$1-p_i$	1	1	1
x_i, y_i	1	1	1	1	1	1

TABLE II

EXAMPLE 2. TBM APPROACH. FOR SENSOR S_i, VALUES OF $p_{X_i}[T, S_i\ Status]$ WHERE THE TARGET T CAN BE Frd (FOR FRIEND) OR Foe AND S_i STATUS CAN BE Wrk (FOR WORKING) AND Brk (FOR BROKEN). p_i IS THE PROBABILITY SENSOR S_i STATUS IS Wrk. THE COLUMNS $pl^{X_i}[Frd]$ AND $pl^{X_i}[Foe]$ PRESENT THE PLAUSIBILITY ON X_i GIVEN THE TARGET IS Frd OR Foe, THUS THE LIKELIHOODS GIVEN TO THE TARGETS WHEN THE OBSERVED DATA IS x_i OR y_i, RESPECTIVELY.

achieved using the commonality function $q^H[d_i]$ as:

$$q^H[data](T) = \prod_{i=1...30} q^H[d_i](T),\ \forall T \subseteq H.$$

Table IV presents the details of this computation and the resulting pignistic probabilities $BetP$. The TBM approach concludes that the target is a Friend with $BetP(Friend) = .91$.

Notice that if one had replaced the vector (1 1 1) of pl^{X_i} by (.5 .5 1) in the Brk columns of Table II, the results of the TBM analysis become the same is those of the probability approach. So the source of the difference between the results of the two approaches comes from the fact we represent total ignorance by equal probabilities in the probability approach and by a vacuous belief function in the TBM approach. Which representation is adequate is a matter of personal opinions.

D. Choosing between the two models.

As shown before, the two models, the Bayesian and the TBM, give totally opposite conclusions, which is quite disquieting. Its origin has been found in the representation of the uncertainty about how the sensor would react when broken: with probabilities, we use

$pl^H[x_i]$	$HQS : p_i = .99$		$LQS : p_i = .90$	
	x_i	y_i	x_i	y_i
$Friend$	1.00	0.01	1.00	0.10
Foe	0.01	1.00	0.10	1.00
H	1.00	0.010	1.00	1.00

TABLE III

EXAMPLE 2. VALUES OF THE PLAUSIBILITY FUNCTIONS $pl^H[d_i]$ OVER $H = \{Friend, Foe\}$ GIVEN THE DATA $= d_i$ ACCORDING TO THE SENSOR QUALITY. THESE DATA ARE COMPUTED FROM THOSE IN TABLE II BY APPLYING THE GBT.

$q^H[x_i]$	HQS		LQS		$q^H[data]$	$BetP$
	x_i	y_i	x_i	y_i		
$\emptyset$	1.00	1.00	1.00	1.00	1.00	
$Friend$	1.00	0.01	1.00	0.10	1.0E-20	.909
Foe	0.01	1.00	0.10	1.00	1.0E-21	.091
H	0.01	0.01	0.10	0.10	1.0E-41	
$frequencies$	10	1	1	18		

TABLE IV

EXAMPLE 2. INDIVIDUAL VALUES OF THE COMMONALITY FUNCTION $q^H[x_i]$, AND $q^H[data]$, THE RESULT OF THEIR COMBINATION BY THE CONJUNCTIVE COMBINATION RULE. THE COLUMN $BetP$ PRESENTS THE PIGNISTIC PROBABILITIES OVER H.

the uniform distribution, and with the TBM, we use the vacuous belief.

We feel that we cannot leave each model stand against the other without giving readers some help to choose among them. A possible answer comes from the study of the informativity of the sources.

In the probabilistic setting, the entropy of Shannon has been defined in order to assess and compare the informativity of distributions. The TBM can represent a broader range of uncertainty, and the entropy of Shannon cannot be applied directly to bba's. Some researches [3], [4], [5], [6], [7], [8] have been undertaken to try to measure the informativity of a belief function.

According to these measures of uncertainty, the uniform distribution used in probability theory is not the most uninformative among all the possible representations of absolute uncertainty; it includes already ‘something’. On the contrary, the vacuous belief is always the least informative.

Although still arguable, these views of uncertainty measures could help us to decide what model is appropriate in presence of such a large uncertainty.

III. Computational Efficiency: Comparison with the Bayesian Model

We present an example inspired by [2] where the authors try to show that the use of belief functions is computationally less efficient than the use of probability functions. In fact, we show, with the same example, that the converse holds once belief functions are used according to the theory developed here.

Using their approach, we will compare the number of identical observations required by the probability model and the TBM so that the best hypothesis reaches a level of 0.99 probability.

A. Example 3. One Sensor, Theoretical Analysis

The first comparison involves one sensor. Let $H = \{h_1, h_2, \dots h_n\}$ denote a set of n hypotheses, for instance the type of aircraft under observation. Hypothesis could be F15, B 737, Sukkoi 21, Rafale, F18... One of these hypotheses, denoted h_0, corresponds to the actual one.

In order to determine the value of h_0, a sensor, like an Electronic Support Measure (ESM), makes an observation. Let X denote the set of possible values this observation can take. Suppose the sensor measurement is $x \in X$. Table V presents the values of the plausibility function given to the fact that the observation is x for each possible value of H. These values are the likelihoods given to each $h_i \in H$ once the observation is x. For simplicity sake, we use identical likelihoods for all hypotheses except the first one. Table VI presents some of the values of the plausibility and belief functions induced by x on H.

Suppose You make k independent observations. The result is the vector $d = (d_1, d_2 \dots d_k)$.

data	$l(h_1)$	$l(h_2)$	$\ldots$	$l(h_n)$
x	a	b	$\ldots$	b

TABLE V

EXAMPLE 3. LIKELIHOODS ON H PRODUCED BY THE OBSERVATION x.

H	$pl^H[x]$	$bel^H[x]$
h_1	a	$a(1-b)^{(n-1)}$
h_2	b	$(1-a)b(1-b)^{(n-2)}$
h_1, h_2	$a+b-ab$	$(a+b-ab)(1-b)^{(n-2)}$
h_2, h_3	$2b-b^2$	$(2b-b^2)(1-b)^{(n-2)}$
h_1, h_2, h_3	$a+2b-2ab-b^2+ab^2$	$(a+2b-2ab-b^2+ab^2)(1-b)^{(n-3)}$

TABLE VI

EXAMPLE 3. VALUES OF *pl* AND *bel* FOR SOME REPRESENTATIVE SUBSETS OF H, COMPUTED FROM THE LIKELIHOODS OF TABLE V.

The domain of d is $X \times X \ldots \times X = X^k$. Suppose each $d_j = x$ for $j = 1, \ldots k$. So d can be written as x^k.

Table VII presents $l(h_i|d)$ with a and b being two positive reals in $(0, 1)$ and $a > b$.

data	$l(h_1\|d)$	$l(h_2\|d)$	$\ldots$	$l(h_n\|d)$
d	a^k	b^k	$\ldots$	b^k

TABLE VII

EXAMPLE 3. LIKELIHOODS ON H PRODUCED BY THE OBSERVATION $d = (d_1, d_2 \ldots d_k)$ WHERE $d_j = x$ FOR $j = 1, \ldots k$.

Based on the likelihoods of Table VII, we assess the beliefs about the value of h_0.

A.1 Probability Approach

The initial knowledge state is total ignorance, so we assume an *a priori* probability $P_0^H(h_i) = 1/n,\ \forall h_i \in H$.

Suppose we have observed k times the data x. With the Bayes's rule we compute the probability at step k. Let x^k represent the collected data, i.e., k times x, and X^k its domain. We have,

$$P^H[x^k](h_1) = \frac{P^{X^k}[h_1](x^k)}{P^{X^k}[h_1](x^k) + (n-1)P^{X^k}[h_2](x^k)} \tag{30}$$

$$= \frac{a^k}{a^k + (n-1)b^k} \tag{31}$$

and for $i = 2 \ldots n$,

$$P^H[x^k](h_i) = \frac{P^{X^k}[h_2](x^k)}{P^{X^k}[h_1](x^k) + (n-1)P^{X^k}[h_2](x^k)} \tag{32}$$

$$= \frac{b^k}{a^k + (n-1)b^k}. \tag{33}$$

The number of steps k_P required by the Bayesian approach is the value of k such that:

$$\frac{a^k}{a^k + (n-1)b^k} \geq 0.99.$$

k_P is the smallest integer k with:

$$k \geq \frac{log(99(n-1)}{log(a/b)}.$$

A.2 TBM Approach

Suppose we represent the uncertainty with belief functions. The GBT needs $pl^X[hi](x), i = 1 \ldots n$. As we only need these plausibilities on x, we can as well consider the space X as made of two singletons, x, and $\overline{x}$.

There are three ways the masses $m^X[h_1]$ can be allocated on X.

case 1	case 2	case 3
$m^X[h_1](x) = a$	$m^X[h_1](x) = a - c$	$m^X[h_1](x) = 0$
$m^X[h_1](\overline{x}) = 1 - a$	$m^X[h_1](\overline{x}) = 1 - a - c$	$m^X[h_1](\overline{x}) = 1 - a$
$m^X[h_1](X) = 0$	$m^X[h_1](X) = c$	$m^X[h_1](X) = a$

In the three cases $pl^X[h_1](x) = a$, so it does not matter for the GBT which of the three forms is used. The same holds for $m^X[h_i], i = 2 \ldots n$. The first case is the closest from the probability approach, and corresponds probably to the most common type of data one can expect for application. It just means that the conditional beliefs over the data are probabilistic in nature. The major difference between the GBT and the probabilistic approach comes from the way the prior knowledge is represented, a vacuous *a priori* being usable within the TBM whereas probability theory does not allow such a flexibility

To achieve the comparison, we need $BetP(h_1)$ after observing k data. Thanks to the properties presented in Section I-III-D.2, all we need are the likelihoods given x^k. They are a^k and b^k. The bba on H is (see relation (22)):

$$\begin{aligned} m^H[x^k](h_1) &= a^k(1-b^k)^{n-1} \\ m^H[x^k](h_1,h_2,\dots h_i) &= a^k b^{k(i-1)}(1-b^k)^{n-i} \\ m^H[x^k](h_2,\dots h_i) &= (1-a^k)b^{k(i-1)}(1-b^k)^{n-i} \end{aligned}$$

or in general, $\forall h \subseteq H$

$$m^H[x^k](h) = a^{k|h_1\cap h|}(1-a^k)^{1-|h_1\cap h|}b^{k|\overline{h_1}\cap h|}(1-b^k)^{(n-|\overline{h_1}\cap h|-1)}$$

A.3 Computing $BetP$

We derive the equation for $BetP^H[x^k]$ using the symbols as defined above.

Theorem III.1:

$$BetP^H[x^k](h_1) = \frac{a^k}{nb^k}\,\frac{1-(1-b^k)^n}{1-(1-a^k)(1-b^k)^{n-1}} \tag{34}$$

Proof. Let $a^k = y, b^k = z$. By relation (21), we have:

$$\begin{aligned} BetP^H[x^k](h_1) &= \sum_{h_1\in h\subseteq H}\frac{m^H[x^k](h)}{|h|(1-m^H[x^k](\emptyset)} \\ &= \frac{1}{(1-m^H[x^k](\emptyset))}\sum_{h\subseteq\overline{h_1}}\frac{y}{1+|h|}z^{|h|}(1-z)^{n-1-|h|} \\ &= \frac{1}{(1-m^H[x^k](\emptyset))}\sum_{i=0}^{n-1}\frac{y}{1+i}\binom{n-1}{i}z^i(1-z)^{n-1-i} \\ &= \frac{y}{(1-m^H[x^k](\emptyset))}\frac{1}{nz}(1-(1-z)^n) \\ m^H[x^k](\emptyset)) &= (1-y)(1-z)^{n-1} \\ BetP^H[x^k](h_1) &= \frac{y}{nz}\,\frac{1-(1-z)^n}{1-(1-y)(1-z)^{n-1}} \end{aligned}$$

□

A.4 Convergence Speed

We prove that the ratio $BetP^H[x^k](h_1)/P^H[x^k](h_1)$ is always larger than 1, thus that the TBM converges faster than the probability model, contrary to what Buede and Girardi [2] concludes.

Theorem III.2: $BetP^H[x^k](h_1) \geq P^H[x^k](h_1)$.

Proof. Let $a^k = y, b^k = z$. By relations (34) and (30), we have:

$$\frac{BetP^H[x^k](h_1)}{P^H[x^k](h_1)} = \frac{y}{nz} \frac{1-(1-z)^n}{1-(1-y)(1-z)^{n-1}} \frac{y+(n-1)z}{y}$$

To prove the ratio is larger than 1, we must show that

$$(1-(1-z)^n)(y+(n-1)z) \geq nz(1-(1-y)(1-z)^{n-1})$$

or equivalently that:

$$(1-(1-z)^n)(y+(n-1)z) - nz(1-(1-y)(1-z)^{n-1}) \geq 0$$

which, after arithmetic manipulations, becomes equal to:

$$(y-z)(1-(1-z)^n - nz(1-z)^{n-1}) \geq 0.$$

As $y \geq z$ by hypothesis, we must only show that:

$$1 \geq (1-z)^n + nz(1-z)^{n-1}.$$

This is equivalent to showing that:

$$\begin{aligned} 1 &\geq (1-z)^{n-1}(1+(n-1)z) \\ 1-(1-z)^{n-1} &\geq (n-1)z(1-z)^{n-1} \\ \frac{1-(1-z)^{n-1}}{1-(1-z)} &\geq (n-1)(1-z)^{n-1} \text{ as } z \in [0,1) \\ \sum_{i=0}^{n-1}(1-z)^i &\geq (n-1)(1-z)^{n-1} \\ \sum_{i=0}^{n-1}(1-z)^{-(n-i-1)} &\geq n-1 \end{aligned}$$

As the n terms $(1-z)^{-(n-i-1)}$ in the summation are larger or equal to 1 as $z \in [0,1)$, their sum is larger than n, thus proving the inequality. □

The ratio $BetP^H[d](h_1)/P^H[d](h_1)$ is thus always larger than 1 when $a > b$, for all $k = 1, 2 \ldots$. Therefore the number of steps k needed so that $BetP^H[d](h_1) \geq .99$ is never larger than the number of steps needed so that $P^H[d](h_1) \geq .99$.

This result contradicts the conclusions of Buede and Girardi [2]. The possible origin of the discrepancy is discussed in Section III-D.

In fact the ratio is very close to 1 and the number of step are essentially the same in both cases. In any case, this efficiency criterion cannot be used against the TBM approach. Instead it could have been used against the probabilistic approach, were it not for the fact that the difference is too small to produce a convincing argument against the probabilistic approach.

In Table VIII, we compare the number of steps needed to reach a .99 threshold for the pignistic probability (left term) or the posterior probability (right term). The TBM is sometime a little more efficient, but the differences do not seem to be of practical usefulness.

B. Example 4. Two Sensors Problem, Theoretical Analysis

Suppose we use two sensors S_1 and S_2 that observe the data x_1 and x_2, respectively. Let the likelihoods they generate in such cases be:

- $l_1(h_i) = a$ for $i = 1 \ldots j$, and b for $i = j+1 \ldots n$
- $l_2(h_i) = b$ for $i = 1 \ldots j-1$, and a for $i = j \ldots n$

where $a > b$. So alone, the sensors cannot discriminate the hypothesis h_j when it holds, whereas together, they do it nicely.

The same data x_1, x_2 has been collected k times.

$P^X[h_2](x)$	$P^X[h_1](x)$								
	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
0.10	4	4	4	4	3/4	3	3	3	3
0.12	5	4	4	4	4	4	3/4	3/4	3
0.14	5	5	5	4	4	4	4	4	3/4
0.16	6	5	5	5	4/5	4	4	4	4
0.18	6	6	5	5	5	5	4/5	4	4
0.20	7	6	6	6	5	5	5	4/5	4
0.22	8	7	6	6	6	5	5	5	4/5
0.24	9	8	7	6/7	6	6	5	5	5
0.26	10	8	8	7	6/7	6	6	5/6	5
0.28	11	9	8	8	7	6/7	6	6	5/6
0.30	12	10	9	8	8	7	6/7	6	6
0.32	14	12	10	9	8	7/8	7	6/7	6
0.34	16	13	11	10	9	8	7	7	6/7
0.36	19	15	12	11	9	9	8	7	7
0.38	22	17	14	12	10	9	8/9	8	7
0.40	27	19	15	13	11	10	9	8	8
0.42	35	23	17	14	12	11	10	9	8
0.44	47	27	20	16	13	12	10/11	9/10	9
0.46	72	34	23	18	15	13	11	10	9
0.48	147	44	27	20	16	14	12	11	10
0.50	-	63	33	23	18	15	13	12	10/11

TABLE VIII

EXAMPLE 3. WITH $n = 6$, EACH PAIR OF NUMBERS r/s GIVES THE VALUES OF K NEEDED TO GET $BetP^H[d](h_1) \geq .99$ FOR r AND $P^H[d](h_1) \geq .99$ FOR s ACCORDING TO THE VALUES OF $a = P^X[h_1](x)$ (COLUMNS) AND $b = P^X[h_2](x)$ (ROWS). WHEN $r = s$, ONLY ONE NUMBER IS INDICATED. WHEN $r \neq s$, THE TWO NUMBERS ARE INDICATED.

Assuming the observations collected with the two sensors are independent, we can thus compute the likelihoods for the joint data. They are:

$$
\begin{aligned}
l_{12}(h_i) &= a^k b^k && \text{for } i = 1 \ldots j-1 \\
&= a^{2k} && \text{for } i = j \\
&= a^k b^k && \text{for } i = j+1 \ldots n
\end{aligned}
$$

We are thus back to the previous example, and the same proof shows that the pignistic probabilities computed in the TBM converge faster than the posterior probabilities computed in the probabilistic approach.

C. Example 5. Numerical Comparison on a Practical Example

This part revisits the simulations used by Buede and Girardi [2]. We use the same data described in their paper, and we compare the speed of convergence to the winning hypothesis by two decision systems: the classical Bayesian approach and TBM approach.

Contrary to the results obtained by Buede and Girardi, the TBM-based algorithm will never be slower than the Bayesian algorithm.

They simulate a problem of decision concerning an aircraft engagement. The aircraft fled by the user detects another aircraft, and the question is to classify it. There are ten possible aircrafts. Table IX presents the various hypothesis, their class and their nature.

Identity	Type	Class	Nature
1	F15	Fighter	Friend
2	F16	Fighter	Friend
3	ATF	Fighter	Friend
4	B2	Bomber	Friend
5	Mig27	Fighter	Foe
6	Mig25	Fighter	Foe
7	Mig29	Fighter	Foe
8	Mig31	Fighter	Foe
9	Tu26	Bomber	Foe
10	Boeing	Commercial	Neutral

TABLE IX

The various hypotheses of the problem.

The user possesses a multisensor system to detect and recognize possible aircrafts: it consists of an Electronic Support Measure (ESM), an Identification Friend or Foe (IFF) and a Radar Sensors. We assume here that these three sensors have already been trained on the possible aircrafts. Tables X, XI, XII present the confusion matrices, i.e., the conditional probabilities about the sensor observation for each possible aircraft.

The ESM sensor has been trained to discriminate between the 10 types. The IFF sensor can only discriminate between the two classes: Friend of Foe, and the Radar sensor can only distinguish between the three natures: Fighter, Bomber, or Commercial.

Claimed	Actual aircraft									
to be	F15	F16	ATF	B2	Mig27	Mig25	Mig29	Mig31	Tu26	Boeing
F15	0.526	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
F16	0.053	0.526	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
ATF	0.053	0.053	0.526	0.053	0.053	0.053	0.053	0.053	0.053	0.053
B2	0.053	0.053	0.053	0.526	0.053	0.053	0.053	0.053	0.053	0.053
Mig27	0.053	0.053	0.053	0.053	0.526	0.053	0.053	0.053	0.053	0.053
Mig25	0.053	0.053	0.053	0.053	0.053	0.526	0.053	0.053	0.053	0.053
Mig29	0.053	0.053	0.053	0.053	0.053	0.053	0.526	0.053	0.053	0.053
Mig31	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.526	0.053	0.053
Tu26	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.526	0.053
Boeing	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.526

TABLE X

Confusion matrix for the ESM sensor.

Claimed	Actual aircraft									
to be	F15	F16	ATF	B2	Mig27	Mig25	Mig29	Mig31	Tu26	Boeing
Friend	0.909	0.909	0.909	0.909	0.091	0.091	0.091	0.091	0.091	0.091
Not Friend	0.091	0.091	0.091	0.091	0.909	0.909	0.909	0.909	0.909	0.909

TABLE XI

Confusion matrix for the IFF sensor.

Claimed	Actual aircraft									
to be	F15	F16	ATF	B2	Mig27	Mig25	Mig29	Mig31	Tu26	Boeing
Fighter	0.833	0.833	0.833	0.083	0.833	0.833	0.833	0.833	0.083	0.083
Bomber	0.083	0.083	0.083	0.833	0.083	0.083	0.083	0.083	0.833	0.083
Commer.	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.833

TABLE XII

Confusion matrix for the Radar sensor.

For the experiment, we varied three parameters. There was a possible misassociation: when detecting an aircraft, if a misassociation occurs, another aircraft is detected. The probability of a misassociation is 0.0 or 0.2 or 0.4. It was also possible that a sensor could not take a measure, sending in this case an empty information. An empty information was modeled by a uniform distribution in the probability framework, and an vacuous belief function in the TBM framework. The probability that a sensor could not send a message ranges from 0 to 0.4 by 0.1 steps. The no-report problem could apply on each sensor separately, each pair of sensors and the three sensors simultaneously.

We rerun all the simulations performed by Buede and found out that the number of identical observations needed to reach a .95 *a posteriori* probability was always the same in both approaches. We also discovered that the difference between the pignistic probability and the Bayesian probability observed when reaching the .95 threshold was positive in 70% with the pignistic probability, being larger than the Bayesian probability and within 10e-10 in 30% (what looks more like a rounding error). The difference was never negative. The largest difference encountered was 0.004. The TBM conclusions were always a little bolder than the one reached by the Bayesian approach, even though the difference seems really useless for any practical purpose.

These results indicate only that the claims made against the computational complexity of the belief based model were inexact.

D. Origin of the difference

There are three essentials differences between our GBT solution and the solution published in [1], [2], denoted hereafter the X solution.

- The interpretation of the confusion matrix data. In the X solution, the bba on H satisfies: $m^H[x](h_x) = l(x|h_x)$ and $m^H[x](H) = 1 - l(x|h_x)$ where $h_x \in H$ is the most likely hypothesis under x, i.e., $l(x|h_x) > l(x|h_k)\ \forall h_k \in H, h_k \neq h_x$. Its origin is not explained. In the GBT solution, the computation of $m^H[x]$ is derived from the GBT, and the GBT uses the same confusion matrix as the probabilists.
- The relation used in the X solution to compute $m^H[x^n]$ after collecting n data (see ([1], eq. (7.28) page 247), and [2], eq. (1) page 570, where u_j must be read as u). The equation is only correct if $m^H[x]$ is a simple support function, what happens for some of the cases analyzed by the X solution. But in that simple case, $m^H[x^n](h_x)$ could be computed much more efficiently by $m^H[x^n](H) = (1 - l(h_x|x))^n$ and $m^H[x^n](h_x) = 1 - (1 - l(h_x|x))^n$. To show the inadequacy of the published equation in the more general case where only singletons of the H domain and H itself receive positive bbm's (what seems to be what the X solution is supposed to handle), consider the case with $H = \{a, b\}$, $m^H[x](a) = m^H[x](b) = 1/4$ and $m^H[x](a, b) = 2/4$. Table XIII presents the value of the normalized bba after collecting n times the same data x, with $n = 2$ and 3. The published equation

gives for $n = 3$:

$$m^H[x^3](h_x) = \frac{1/4\ [\ (1/4 + 2/4)2 + 2/4 + (2/4)2]}{2 * 1/4\ [\ (1/4 + 2/4)2 + 2/4 + (2/4)2] + (2/4)3} = \frac{21}{50}$$

whereas the solution obtained by Dempster's rule of combination is 19/46.

• The X solution focuses on comparing the *a posteriori* probabilities with the belief $bel^H[x_n](h_x)$ computed in the X model instead of the pignistic probabilities. Basing decision on *bel* is usually not advised. In the GBT, we use the pignistic probabilities for the comparisons, so comparing comparable objects.

These discrepancies lead to the published conclusions that we have shown to be incorrect.

H	$m^H[x]$	$m^H[x^2]$	$m^H[x^3]$
$\{a\}$	1/4	5/14	19/46
$\{b\}$	1/4	5/14	19/46
$\{a, b\}$	2/4	4/14	8/46

TABLE XIII

COMPUTATION OF THE NORMALIZED BBA'S $m^H[x^n] = \oplus_{i=1,\ldots,n} m^H[x]$ USING DEMPSTER'S RULE OF COMBINATION.

IV. EXAMPLE 6. NESTED SENSORS WITHOUT OVERLAP

We analysis three cases of sequentially observed data collected by various sensors and which results refine the previous ones. The cases vary according to which likelihoods are known:

• the likelihoods are generated on non overlapping subsets of the initial frame, (section IV),

• the likelihoods are generated on every element of a partition of the initial frame, but the granularity of the partition is finer at each level (section V),

• the likelihoods are generated by each sensor on partially overlapping subsets of the overall frame (section VI).

These three cases are practical cases we encountered, but rephrased and simplified for the presentation. Their solutions illustrate how the GBT can be used in practice.

We want to identify between three types of vehicles: Buses ($h_1 = B$), Cars ($h_2 = C$) and Trucks ($h_3 = T$). There are two models of Buses, denoted h_{11}, h_{12}, two models of Cars, denoted h_{21}, h_{22} and two models of Trucks, denoted h_{31}, h_{32} (see Figure 1).

Suppose we have 4 sensors. The first, denoted S_0, measures the variable X_0 that can distinguish between Buses, Cars and Trucks. So its frame of discernment is $H_0 = \{h_1, h_2, h_3\}$. The second, denoted S_1, measures X_1 and can distinguish between the two models of Buses. The third, denoted S_2, measures X_2 and can distinguish between the two models of Cars. The fourth, denoted S_3, measures X_3 and can distinguish between the two models of Trucks. Their frames of discernment of S_1, S_2, S_3 are H_1, H_2, H_3, respectively, where $H_i = \{h_{i1}, h_{i2}\}$. The granularity of each frame of discernment is important. So the frame of discernment H_0 has three singletons. Similarly the frame of discernment H_1 has two singletons, and the same holds for H_2 and H_3. The overall frame of discernment H has in fact six elements, the $h_{ij},\ i = 1, 2, 3,\ j = 1, 2$. H_0 is a coarsening of H, whereas H_1, H_2 and H_3 are disjoint subsets of H.

The sensors S_0, S_1, S_2, S_3 produce the bba's $m^{H_i}, i = 0, 1, 2, 3$, on their respective frames.

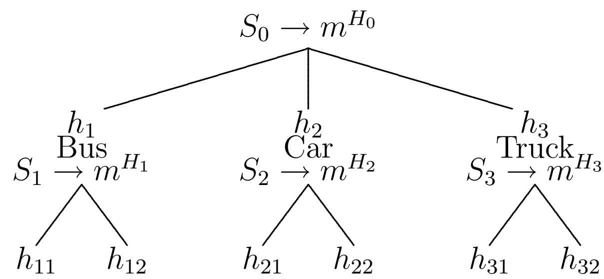

Fig. 1. Example 6. The tree of hypotheses and the domain of the four sensors with the beliefs they produce.

As defined H_1, H_2 and H_3 do not share a common refinement (they are not compatible frames, see [9]), so conjunctive combination rules cannot be applied directly. In order to get a bba on H, we build the ballooning extension (see Section I-II-F) of the m^{H_i}'s on H. For instance $m^{H_1}(h_{11})$ will be extended on H so that it will be allocated to the $h_{11} \cup H_2 \cup H_3$. By construction, these extensions share the same frame of discernment and the combination rules can then be applied to them.

When we state that S_1 can distinguish between the types of buses, we mean also that we have no idea whatsoever about how sensor S_1 would react if it was facing a car or a truck. This might occur if S_1 had never been used to measure the X_2 values on cars and X_3 values on trucks. Identical properties hold, up to permutation, for S_2 and S_3.

Suppose the four measurements are x_0, x_1, x_2, x_3. The four sets of likelihoods are presented in Tables XIV and XVI.

	Bus h_1	Car h_2	Truck h_3
$l(h_i\|x_0)$	0.60	0.40	0.10
$Bel^{H_0}[x_0]$	0.41	0.18	0.03
$Pl^{H_0}[x_0]$	0.77	0.51	0.13

TABLE XIV

Example 6. Likelihoods on H_0 produced by the observations x_0, , and the normalized beliefs and plausibilities they induce on the singletons of H_0.

H_0	$\emptyset$	h_1	h_2	h_1, h_2	h_3	h_1, h_3	h_2, h_3	h_1, h_2, h_3
$m^{H_0}[x_0]$	0.22	0.32	0.14	0.22	0.02	0.04	0.02	0.02

TABLE XV

Example 6. The basic belief assignment on H_0 given x_0, computed from the likelihoods of Table XIV.

We first compute the bba $m^{H_0}[x_0]$ on H_0 given x_0 from the likelihoods produced by S_0 (i.e., the line $l(h_i|x_0)$ of Table XIV). The resulting bba is presented in Table XV. Its computation is done by applying equation (22). So

$$m^{H_0}[x_0](h_1) = .60 \times (1 - .40) \times (1 - .1) = .32$$

For each sensor $S_i,\ i = 1, 2, 3$, we compute $m^{H_i \Uparrow H}[x_i]$, the ballooning extension of $m^{H_i}[x_i]$ on H.

	Bus $i=1$		Car $i=2$		Truck $i=3$	
	h_{11}	h_{12}	h_{21}	h_{22}	h_{31}	h_{32}
$l(h_{ij}\vert x_i)$	0.45	0.55	0.90	0.10	0.40	0.70
$Bel^{H_i}[x_i]$	0.27	0.40	0.89	0.01	0.15	0.51
$Pl^{H_i}[x_i]$	0.60	0.73	0.99	0.11	0.49	0.85

TABLE XVI

EXAMPLE 6. LIKELIHOODS ON H_1, H_2, H_3 PRODUCED BY THE OBSERVATIONS x_1, x_2, x_3, , AND THE NORMALIZED BELIEFS AND PLAUSIBILITIES THEY INDUCE ON THE SINGLETONS.

We conjunctively combine these three bba and (the vacuous extension of) $m^{H_0}[x_0]$:

$$m^H[x_0, x_1, x_2, x_3] = (\textcircled{\cap}_{i=1,2,3}\ m^{H_i \Uparrow H}[x_i])\ \textcircled{\cap}\ m^{H_0}[x_0]^{\uparrow H}.$$

This last bba is the final bba on H built from all collected data.

For practical applications, the computation can be strongly speed up as, in practice, we hardly need all bbm, but only *bel* and *pl* (and maybe *BetP*) on the elements of H. Table XVII presents these end results for these elements, i.e., the normalized *Bel* and *Pl* functions.

The analysis of the data show that S_0 supports that the target is a Bus (h_1), but after collecting all data, it appears that the h_{21} is probably the best supported target type (thus not a Bus). The data were purposely selected in order to enhance this embarrassing result. In practice, the data leave us essentially embarrassed in choosing between h_{11}, h_{12} and h_{21}, the other alternatives having been eliminated.

These data also enhance the danger of premature decision making. Suppose we apply an iterated procedure by first observing S_0's data, and decide them to collect only S_1's data as far as the first step leads us to consider the target was a Bus. It would save the cost of collecting S_2 and S_3's data, but the end result would have been erroneous, as the h_{21} hypothesis would have been rejected, whereas it seems nevertheless the best hypothesis in the present case. This illustrates the dilemma between cost reduction obtained by taking intermediate decisions versus larger expenses resulting from delayed decisions with 'better' results.

In practical applications where cost reduction is an issue, a pre-posterior sensitivity analysis has to be realized at each step in order to decide if collecting further data would affect the results and are worth the effort. The method is essentially mimicking the strategy followed by the Bayesians. We don't explore this methodology further here, but it can 'easily' be performed within the TBM.

Comments. It might be worth mentioning that the same results are obtained if we define the domain of sensors $S_i, i = 1, 2, 3$, as H, and describe the belief over X_i given the $h_{ij} \notin H_i$ as a vacuous belief function over X_i. In that case the likelihoods $l(h_{ij}|x_k)$ are 1 if $i \neq k$. We then apply the GBT on these three sets of likelihoods and proceed with three belief functions defined on the same frame H. The computation load would be much heavier in that case as we would work on unnecessarily large frames. Nevertheless, this latter approach enhances the fact that we are really ignorant about what might be X_i when sensor S_k face object in H_i, $k \neq i$. The vacuous belief function provides exactly the right representation to describe such a state of ignorance.

	h_{11}	h_{12}	h_{21}	h_{22}	h_{31}	h_{32}
$Bel^H[x_0, x_1, x_2, x_3]$	0.104	0.156	0.240	0.003	0.006	0.000
$Pl^H[x_0, x_1, x_2, x_3]$	0.397	0.485	0.530	0.059	0.059	0.103
$BetP^H[x_0, x_1, x_2, x_3]$	0.231	0.300	0.364	0.028	0.025	0.052

TABLE XVII

EXAMPLE 6. NORMALIZED BELIEFS AND PLAUSIBILITIES AND PIGNISTIC PROBABILITYS INDUCED ON THE SINGLETONS BY x_0, x_1, x_2, x_3.

V. Example 7. Sensors in a Hierarchical Tree

As in Section IV, we want to identify vehicles. They can be categorized in three types: Buses (h_1), Cars (h_2) and Trucks (h_3). Each type can be subdivided according to the auto-maker: Buses can be VanHool (h_{11}) or Mercedes (h_{12}), Cars can be VW (h_{21}), Audi (h_{22}) or Ford (h_{23}), Trucks can only be GMC (h_{31}). For each auto-maker, there are two models of vehicles (h_{ij1}, h_{ij2}), like 'Beetle' and Passat for VW, A4 and A6 for Audi ... Figure 2 presents the tree describing the relation between the hypotheses.

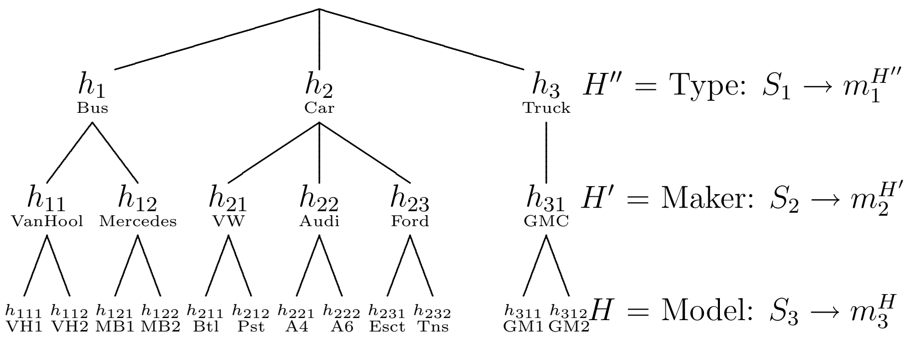

Fig. 2. Example 7. The tree of hypotheses and the domain of the three sensors with the beliefs they produce.

There are three sensors, denoted S_1, S_2, S_3 which measure the values of the three variables X_1, X_2, X_3, respectively. S_1 is able to distinguish between the three types, S_2 between the 6 auto-makers, and S_3 between the 12 models. It means that

• the frame of discernment $H = \{h_{ijk} : i = 1 \dots 3, j = 1 \dots n_i, k = 1, 2\}$ where h_{ijk} is one of the 12 models.

• there is a coarsening $H' = \{h_{ij} : i = 1, 2, 3, j = 1 \dots n_i\}$ of H where the elements of H' are the 6 auto-makers.

• there is a coarsening $H'' = \{h_i : i = 1, 2, 3\}$ of H' where the elements of H'' are the 3 types.

• S_1 reports likelihoods on the types (H''), S_2 on the maker (H') and S_3 on the model (H). It results from the fact that, for S_1, we only know the probability over X_1 given h_1, h_2 and h_3. We have no information about what would be these probabilities given more refined hypothesis, like for instance h_{11}. So S_1 reports only likelihoods on H''. Similarly S_2 reports likelihoods on H', and S_3 on H.

• we denoted by $m_1^{H'}, m_2^{H'}$ and m_3^H the bba induced by the three sensors S_1, S_2, S_3, respectively.

The practical aim is to compute the normalized beliefs, plausibilities and pignistic probabilities for the singletons of H (the models) considering all available likelihoods. The computation could be done in a straightforward manner by vacuously extending $m_2^{H'}$ and $m_1^{H''}$ on H, and conjunctively combining the three bba's so defined on H. But this is not computationally efficient as it means we would have to work on the space 2^H. A more efficient algorithm can be described that needs only to produce the full bba on H'. As one can expect that the cardinality of the antepenultimate level in a tree is really smaller that the one of the last level, the computational benefit can be serious.

In the present example, suppose the three sensors have observed x_1, x_2, x_3, respectively. Let l denote the corresponding likelihoods. So we have three sets of likelihoods:

- $l_1(h_i)$ for $h_i \in H''$
- $l_2(h_{ij})$ for $h_{ij} \in H'$
- $l_3(h_{ijk})$ for $h_{ijk} \in H$

Let m^H_{123} denotes the bba obtained on H after conjunctively combining the bba's produced by the three sensors, and $m^{H'}_{12}$ the one on H' produced by combining sensors S_1 and S_2.

$$m^{H'}_{12} = m^{H'}_2 \textcircled{\cap} m^{H''\uparrow H'}_1$$
$$m^{H}_{123} = m^{H}_3 \textcircled{\cap} m^{H'\uparrow H}_{12}$$

In appendix, we detail successively the computation of pl^H_{123}, bel^H_{123} on the singletons of H and $m^H_{123}(\emptyset)$, the latter being needed for normalizing the end results.

A. Computation for the Example

Tables XVIII and XIX present some details of the computation for the case presented in Figure 2. The object is an VH1 (h_{111}). We use likelihoods of .1 for the wrong hypothesis, and .4 for the correct ones (see Table XVIII). We do not detail the subsets of h_2 and h_3 as the values are all the same as those given in the columns h_2 and h_3. The terms $bel^{H''}_1$ and $pl^{H''}_1$ are computed from l_1 by equations (23) and (24). The bba $m^{H''}_1$ corresponding is detailed in Table XIX in the column $m^{H''\uparrow H'}_1$. In that last table, we also presented the masses $m^{H'}_2$ computed from l_2, and the masses obtained by conjunctively combining $m^{H''}_1$ and $m^{H'}_2$ into $m^{H'}_{12}$ (the small masses are not presented). Table XVIII presents then the beliefs and plausibilities on the singletons of H' and H. Finally, the normalized results on the singletons are displayed at the line Bel^H_{123} and Pl^H_{123}. The hypothesis VH1 (h_{111}) is strongly supported, VH2 (H_{112}) get a small support, and all other hypotheses can be neglected. We don't present the pignistic probabilities as they are always between Bel and Pl, and thus they would hardly bring any useful detail in the present context.

In conclusion, together the three sensors point strongly toward the fact the vehicle is a VH1. The purpose of this numerical example is only to illustrate the computation procedure and as such the data do not deserve a more detailed analysis.

H''	h_1				h_2	h_3	$m(\emptyset)$
l_1	.4				.1	.1	
$bel_1^{H''}$	.324				.054	.054	.486
$pl_1^{H''}$	.4				.1	.1	
H'	h_{11}		h_{12}		h_{2j}	h_{3j}	
l_2	.4		.1		.1	.1	
$bel_{12}^{H'}$	.139		.023		.006	.008	.786
$pl_{12}^{H'}$	.16		.04		.01	.01	
H	h_{111}	h_{112}	h_{121}	h_{122}	h_{2jk}	h_{3jk}	
l_3	.4	.1	.1	.1	.1	.1	
bel_{123}^{H}	.056	.009	.003	.003	.001	.001	.913
pl_{123}^{H}	.064	.016	.004	.004	.001	.001	
Bel_{123}^{H}	.65	.11	.03	.03	.01	.01	
Pl_{123}^{H}	.73	.18	.05	.05	.01	.01	

TABLE XVIII

EXAMPLE 7. VALUE OF bel AND pl FOR THE SINGLETONS OF H'', H' AND H, AND THEIR NORMALIZED VALUES Bel AND Pl ON H. RIGHTMOST COLUMN GIVES THE MASSES GIVEN TO $\emptyset$ USED FOR NORMALIZATION.

VI. EXAMPLE 8. SENSORS ON PARTIALLY OVERLAPPING FRAMES

We present here a case of non compatible frames and a method to extend the conjunctive combination rule in order to handle non compatible frames with some partial overlap. The solution is a 'careful' solution. This topic has been studied in [10], [11], [12] who discuss the present careful solution but also present other bolder solutions (see also [13]).

h_{11}	h_{12}	h_{21}	h_{22}	h_{23}	h_{31}	$m_1^{H''\uparrow H'}$	$m_2^{H'}$	$m_{12}^{H'}$
0	0	0	0	0	0	.486	.364	.786
1	0	0	0	0	0		.236	.139
0	1	0	0	0	0		.039	.023
1	1	0	0	0	0	.324	.026	.015
0	0	1	0	0	0		.039	.006
0	0	0	1	0	0		.039	.006
0	0	0	0	1	0		.039	.006
0	0	0	0	0	1	.054	.039	.006
0	0	1	1	1	0	.054		
1	1	1	1	1	0	.036		
1	1	0	0	0	1	.036		
0	0	1	1	1	1	.006		
1	1	1	1	1	1	.004		
						1.0	.782	.987

TABLE XIX

EXAMPLE 7. SUBSETS OF H' WITH THE BBA'S $m_1^{H''}$ EXTENDED ON H', $m_2^{H'}$ AND THEIR CONJUNCTIVE COMBINATION. THE SUBSETS WITH $m_{12}^{H'}$ SMALL THAN .001 ARE OMITTED. THE 6 LEFTMOST COLUMNS REPRESENT SUBSETS OF H' WITH THE 1 INDICATING WHICH ELEMENTS BELONG TO THE SUBSET. AT BOTTOM LINE, SUM OF THE MASSES DISPLAYED.

Suppose a sensor S_1 that has been trained to recognize h_1 objects and h_2 objects and a second sensor S_2 that has been trained to recognize h_2 objects and h_3 objects (like $h_1 =$ airplanes, $h_2 =$ helicopters and $h_3 =$ rockets). Sensor S_1 never saw an h_3 object, and we know nothing on how S_1 would react if it was observing an h_3 object. Beliefs provided by S_1 are always on the frame of discernment $\{h_1, h_2\}$. The same holds for S_2 with h_1 and h_3 permuted.

A new object X is presented to the two sensors. Both sensors S_1 and S_2 express their beliefs as $m_1^{H'}$ and $m_2^{H''}$, the first on the frame $H' = \{h_1, h_2\}$, the second on the frame

$H'' = \{h_2, h_3\}$. How to combine these two bba's into a bba m_{12}^H on a common frame $H = \{h_1, h_2, h_3\}$?

The careful solution consists in ballooning extending each bba on the frame H and then conjunctively combining them.

In this example the first sensor supports that X is h_1, whereas the second claims that X is h_2. If X had been h_2, how comes the first sensor did not say so? So the second sensor is probably facing an h_1 and just states h_2 because it does not know what an h_1 is. So we feel that the common sense solution is $X = h_1$, what is confirmed by $BetP_{12}^H$, the pignistic probability computed from m_{12}^H, as its largest value .655 is given to h_1.

Just to enhance the simplicity of the belief function solution, we examine what might be an Bayesian solution of this problem. Suppose the two sensors S_1 and S_2. Sensor S_1 generates a probability function on H, denoted P_1^H, but we only know $P_1 = P_1^H[H']$, the value of P_1^H after conditioning it on H'. The same holds for sensor S_2 with $P_2 = P_2^H[H'']$. We must aggregate P_1 and P_2 in order to derive a probability function on H. The major issue is in the reconstruction of P_1^H and P_2^H from P_1 and P_2. It means how to 'de-condition' a probability function on H when all you know is the result of its conditioning on some strict subset of H. There are infinitely many solutions. Introducing the maximum entropy principle leads to a solution strongly linked to the insufficient reason principle and suffers of all the well known weaknesses of this principle. How probabilists would solve that problem is surely not obvious. To get a working solution, they will probably introduce artificial extra assumptions in order to be able to use the probability model.

VII. Conclusions

The transferable belief model (TBM) has been proposed as a mathematical model to represent quantified beliefs. It covers the same domain as the one covered by Bayesian probabilities, but instead of using an additive measure, it quantifies beliefs by a belief function. The TBM is more general than the Bayesian model as the latter is a special case of the first. The TBM provides a more flexible way to represent uncertainty, and can nicely represent any form of partial up to total ignorance.

The TBM fits perfectly for the problem of multisensor data fusion thanks to the various tools that are part of the model: the combination rules that permit the conjunctive or

H	$m_1^{H'}$	$m_2^{H''}$	$m_1^{H'\Uparrow H}$	$m_2^{H''\Uparrow H}$	m_{12}^H	pl_{12}^H	$BetP_{12}^H$
h_1	.6				.42	.90	.655
h_2	.1	.7			.07	.32	.190
h_3		.2			.02	.30	.155
$\{h_1, h_2\}$	.3			.7	.21	.98	
$\{h_1, h_3\}$			.6	.2	.24	.93	
$\{h_2, h_3\}$		.1	.1		.01	.58	
$\{h_1, h_2, h_3\}$			.3	.1	.03	1	

TABLE XX

Example 8. Basic belief assignment $m_1^{H'}$ and $m_2^{H''}$ on two partially overlapping frames, with their ballooning extensions on the common frame H and their conjunctive combination m_{12}^H on H with its related plausibility and pignistic probability functions.

disjunctive combination of belief functions induced by various distinct pieces of evidence, the General Bayesian Theorem that permits belief inversion just as the Bayesian theorem is used for probability inversion, and the pignistic transformation that permits to build the probability measure needed once decision is required (but that does not 'represent' the beliefs themselves). For multisensor data fusion, the likelihoods are the conditional plausibility functions , the General Bayesian Theorem is used to derive the beliefs on the hypothesis space from the likelihoods, the conjunctive combination rule is used to combine the *a posteriori* beliefs induced on the hypothesis space by the various observations (admitted as usual to be conditionally independent), and the pignistic probabilities are used to transform the overall beliefs into a probability measure in order to make decision by maximizing expected utilities.

A nice property brought by the TBM when compared with the classical Bayesian approach consists in the ability to use an *a priori* belief that represents really total ignorance, avoiding thus the classical problems encountered by the Bayesians when it comes to select and justify the required *a priori* probabilities. An example is given (see Section II) that illustrates the problem of choosing between the TBM and the Bayesian approaches, as the results are diametrically opposed depending on the used approach. The divergence results in fact from the introduction of the *a priori* probabilities needed in the Bayesian approach. This example is given in order to enhance that the choice of the model is a serious matter as it might strongly influence on the results.

We present also some illustrative examples on how the TBM should be used in multisensor data fusion for target identification problems.

Authors have often raised the issue that the use of belief functions, as done in the TBM, is computationally inefficient when compared with the probability model. These criticisms were often based on what we feel were inappropriate uses of the TBM. We reproduce these studies and show that, in the examples used by Buede and Girardi [2], the TBM is *never* less efficient and sometimes even more efficient that its contender, the probability model.

For the problem of multisensor data fusion, the TBM seems to offer a serious alternative to the probability model. To decide which model is the best is delicate as the term 'best' is hardly clearly defined. Nevertheless some very encouraging results have been obtained in real life contexts by [14], [15], [16], [17] for a problem of multisensor antipersonnel mine detection.

Other applications related to the use of the TBM for data fusion problems and detections have been developed in [18], [19], [20], [21], [22], [23], [24], [25] and a very fruitful TBM based method for discriminant analysis has been introduced by Denœux and his group [26]. Ayoun and Smets [27] study the problem of the number of targets under observation, a problem that precedes the identification phase.

VIII. Appendix

A. Computing pl^H_{123}

This computation is very easy. We have by (18):

$$
\begin{aligned}
pl_{123}^{H}(h_{ijk}) &= \sum_{h \subseteq H} pl_3^H(h_{ijk} \cap h)\ m_{12}^{H'\uparrow H}(h) \\
&= \sum_{h \subseteq H'} pl_3^H(h_{ijk} \cap h^{\uparrow H})\ m_{12}^{H'}(h) \\
&= pl_3^H(h_{ijk}) \sum_{h_{ijk} \in h \subseteq H'} m_{12}^{H'}(h) \\
&= pl_3^H(h_{ijk})\ pl_{12}^{H'}(h_{ij})
\end{aligned}
$$

and

$$
\begin{aligned}
pl_{12}^{H'}(h_{ij}) &= \sum_{h \subseteq H'} pl_2^{H'}(h_{ij} \cap h)\ m_1^{H''\uparrow H'}(h) \\
&= \sum_{h \subseteq H''} pl_2^{H'}(h_{ij} \cap h^{\uparrow H'})\ m_1^{H''}(h) \\
&= pl_2^{H'}(h_{ij}) \sum_{h_{ij} \in h \subseteq H''} m_1^{H''}(h) \\
&= pl_2^{H'}(h_{ij})\ pl_1^{H''}(h_i)
\end{aligned}
$$

hence

$$
pl_{123}^{H}(h_{ijk}) = pl_3^H(h_{ijk})\ pl_2^{H'}(h_{ij})\ pl_1^{H''}(h_i)
$$

Expressed in likelihoods, it becomes:

$$
pl_{123}^{H}(h_{ijk}) = l_3(h_{ijk}) l_2(h_{ij})\ l_1(h_i).
$$

This computation is so direct that if there is a time problem, one could discard all those h_{ijk} where $pl_{123}^H(h_{ijk})/\max_{h_\nu \in H} pl_{123}^H(h_\nu)$ is less than a given threshold. Then the computation for bel_{123}^H and $m_{123}^H(\emptyset)$ is done on a smaller frame, thus faster.

B. Computing $m_{123}^H(\emptyset)$

Using relations (16) and (12), we get:

$$\begin{aligned} m_{123}^H(\emptyset) &= \sum_{h \subseteq H} m_3^H[h](\emptyset)\ m_{12}^{H'\uparrow H}(h) \\ &= \sum_{h \subseteq H'} m_3^H[h^{\uparrow H}](\emptyset)\ m_{12}^{H'}(h) \\ &= \sum_{h \subseteq H'} b_3^H(\overline{h}^{\uparrow H})\ m_{12}^H(h) \end{aligned}$$

where

$$b_3^H(\overline{h}^{\uparrow H}) = \prod_{h_{ijk} \in h} (1 - l_3(h_{ijk}))$$

The last terms are efficiently computed by storing the terms

$$t(n) = \prod_{h_{ijk} \in h_{ij}} (1 - l_3(h_{ijk}),\ n = \sum_{i'=1}^{i-1} n_{i'} + j,\ \forall h_{ij} \in H',$$

and then combining them with the next MatLab algorithm.

```
bel = 1;
for i =1:cardinalH'
        bel=[bel bel*t(i)];
end
```

The computation of m_{12}^H may be done by expanding vacuously $m_1^{H''}$ on H' and combining it with $m_2^{H'}$, using the Fast Möbius Transforms (see Section I-III-C) and relation (10). Their computation corresponds to the heavier task in the whole computation and requires the application of the Fast Möbius Transforms on a vector of length $2^{|H'|}$, (which is much smaller than the $2^|H|$ encountered if one was not using the present algorithm). Computation is trivial for $|H'| \leq 15$. For much larger values, the algorithm must be adapted in order to benefit from the special structure of the bba's derived from the GBT.

C. Computing of bel_{123}^H

We have:

$$bel_{123}^H(h_{ijk}) = \sum_{h\subseteq H} bel_3^H[h](h_{ijk})\ m_{12}^{H'\uparrow H}(h)$$
$$= \sum_{h\subseteq H'} (bel_3^H(h_{ijk}\cup \overline{h}^{\uparrow H}) - bel_3^H(\overline{h}^{\uparrow H}))\ m_{12}^{H'}(h)$$

where

$$bel_3^H(h_{ijk}\cup \overline{h}^{\uparrow H}) - bel_3^H(\overline{h}^{\uparrow H}) = 0 \qquad \text{if } h_{ijk}\in \overline{h}$$
$$= x(h) \text{ if } h_{ijk}\in h$$

and

$$x(h) = \prod_{h_{ijk}\neq h_\nu\in h}(1-l_3(h_\nu)) - \prod_{h_\nu\in h}(1-l_3(h_\nu))$$
$$= (\frac{1}{1-l_3(h_{ijk})} - 1)\prod_{h_\nu\in h}(1-l_3(h_\nu))$$
$$= \frac{l_3(h_{ijk})}{1-l_3(h_{ijk})}\prod_{h_\nu\in h}(1-l_3(h_\nu))$$

and thus

$$bel_{123}^H(h_{ijk}) = \frac{l_3(h_{ijk})}{1-l_3(h_{ijk})}\sum_{h:h_{ijk}\in h\subseteq H'} m_{12}^H(h)\ \prod_{h_\nu\in h}(1-l_3(h_\nu))$$

The terms $m_{12}^H(h)$ were computed is Section VIII-B.

D. Computation of Normalized Bel_{123}^H and Pl_{123}^H

It is done by dividing all terms in bel_{123}^H and pll_{123}^H by $1 - m_{123}^H(\emptyset)$.

References

[1] E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston, 1990.

[2] D. M. Buede and P. Girardi, “A target identification comparison of Bayesian and Dempster-Shafer multisensor fusion,” *IEEE Trans. SMC: A*, vol. 27, pp. 569–577, 1997.

[3] Ph. Smets, “Information content of an evidence,” *International Journal of Machine Studies*, vol. 19, pp. 33–43, 1983.

[4] Ph. Smets and P. Magrez, "Additive structure of the measure of information content.," in *Approximate Reasoning in Expert Systems.*, M.M. Gupta, A. Kandel, W. Bandler, and J.B. Kiszkaed, Eds., 1985, pp. 195–197.

[5] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *International Journal of General Systems*, vol. 9, no. 4, pp. 249–260, 1983.

[6] G. J. Klir, "Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like," *Int. J. Fuzzy Sets and Systems*, vol. 24, pp. 141–160, 1987.

[7] G. J. Klir, "Measures of uncertainty in the Dempster-Shafer theory of evidence," in *Advances in the Dempster-Shafer theory of evidence*, R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., pp. 35–49. John Wiley and Sons, New-York, 1994.

[8] G. Klir and M. J. Wierman, *Uncertainty-based-information : elements of generalized information theory*, Physica-Verlag, Heidelberg, NY, 1998.

[9] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press. Princeton, NJ, 1976.

[10] F. Janez, *Fusion de sources d'information définies sur des référentiels non exhaustifs différents*, Ph.D. thesis, Université d'Angers, France, 1996.

[11] F. Janez and A. Appriou, "Fusion of sources defined on different non-exhaustive frames.," in *Information Processing and Management of Uncertainty*, IPMU-96, Ed., 1996.

[12] F. Janez and A. Appriou, "Théorie de l'évidence et cadres de discernements non exhaustifs.," *Revue Traitement du Signal.*, vol. 13, 1996.

[13] P. Smets, "Practical uses of belief functions," in *Uncertainty in Artificial Intelligence 99*, K. B. Laskey and H. Prade, Eds. 1999, pp. 612–621, Morgan Kaufman, San Francisco, Ca.

[14] N. Milisavljevic, I. Bloch, and M. Acheroy, "Characterization of mine detection sensors in terms of belief function and their fusion," in *Proc. 3rd Int. Conf. on Information Fusion (FUSION 2000)*, 2000, pp. ThC3.15–ThC3.22.

[15] N. Milisavljevic, I. Bloch, and M. Acheroy, "Modeling, combining and discounting mine detection sensors within Dempster-Shafer framework," in *Detection Technologies for Mines and Minelike Targets*. 2000, vol. 4038, pp. 1461–1472, SPIE Press, Orlando, USA.

[16] N. Milisavljevic, *Analysis and Fusion Using Belief Functions Theory of Multisensor Data for Close-Range Humanitarian Mine Detection*, Ph.D. thesis, Ecole Normale Supérieure de Télécomminications, Paris, France, 2001.

[17] N. Milisavljevic, S. van den Broek, I. Bloch, P. B. W. Schwering, H. A. Lensen, and M. Acheroy, "Comparison of belief function and voting method for fusion of mine detection sensors," in *Detection and Remediation Technologies for Mines and Minelike Targets VI*. 2001, vol. 4394, SPIE Press, Orlando, USA.

[18] I. Jarkass, *Reconnaissance de l'état d'un système dynamique à l'aide d'un réseau de Petri crédibiliste*, Ph.D. thesis, Université de Technologie de Compiègne, France, 1998.

[19] G. Lohmann, *An evidential reasoning approach to the classification of satellite images*, Ph.D. thesis, Deutsche Forschungsanstalt fur Luft- und Raumfahrt. Distributed through Wissenschaftliches Berichtswesen der DLR, Postfach 90 60 58, 5000 Koln 90, Germany, 1991.

[20] G. Lohmann, "An evidential reasoning approach to the classification of satellite images," in *Symbolic and Quantitative Approaches to Uncertainty*, R. Kruse and P Siegel, Eds. 1991, pp. 227–231, Springer-Verlag, Berlin.

[21] A. Nifle, *Modélisation comportementale en fusion de données. Application la détection et l'identification d'objets et de situations*, Ph.D. thesis, Université de Paris-Sud, 1997.

[22] L. M. Zouhal, *Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes*, Ph.D. thesis, Université de Compiègne, 1997.

[23] L. M. Zouhal and T. Denœux, "An adaptive k-NN rule based on Dempster-Shafer theory," in *Proc. of the 6th Int. Conf. on Computer Analysis of Images and Patterns (CAIP'95)*, Prague, September 1995, pp. 310–317, Springer Verlag.

[24] L. M. Zouhal and T. Denœux, "A comparison between fuzzy and evidence-theoretic k-NN rules for pattern recognition," in *Proceedings of EUFIT'95*, Aachen, August 1995, vol. 3, pp. 1319–1325.

[25] A. Appriou, "Multisensor signal processing in the framework of the theory of evidence," in *NATO/RTA, SCI Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, 1999.

[26] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Transactions on Systems, Man and Cybernetics:*, vol. 25, pp. 804–813, 1995.

[27] A. Ayoun and Ph. Smets, "Data association in multi-target detection using the transferable belief model.," *Inter. J. Intell. Systems*, vol. 16, pp. 1167–1182, 2001.