Kalman Filter and Joint Tracking and Classification in the TBM framework

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

Branko Ristic
ISR division
DSTO
PO Box 1500
Edinburgh SA 5111, Australia
branko.ristic@dsto.defence.gov.au

Abstract – *The paper presents an approach to joint tracking and classification based on belief functions as understood in the transferable belief model (TBM). For the tracking phase, a Kalman filter in the TBM framework is derived. This filter is essentially the same as the classical Kalman filter with a diffuse prior, although it is derived in a more general context. For the classification phase, the TBM solution provides more reasonable results than the corresponding Bayesian classifier in situations where no one-to-one mapping between target behaviours and classes can be established.*

Keywords: Transferrable belief model, target classification, Kalman filter, belief function theory

1 Introduction

The paper is devoted to joint tracking and classification of targets based on kinematic data. The optimal Bayesian estimator for this problem [4, 9, 10, 12] consists of a bank of filters that match the expected dynamic behaviour of each class (class-matched filters). In some numerical examples, however, this type of classifier may produce unreasonable performance due to inadequate mapping between behaviour and class. This was a motivation to consider the problem of joint tracking and classification (JTC) in the framework of transferable belief model (TBM) [18, 19, 20, 23], which is based on Shafer's belief functions [15]. In order to be consistent, this requires to formulate both the class-matched filtering and classification in the TBM framework. Hence in this paper we first present the derivation of the Kalman filter (as a typical building block of class-matched filters) in the TBM framework. In particular, we show that the classical Kalman filtering relations [1] can be derived within the TBM even after relaxing several assumptions underlying the probabilisitic analysis. The classification results, however, differ significantly between the TBM and the Bayesian model, and we argue that those obtained using the TBM are more satisfactory.

The Kalman filter (KF) have been described as an example of evidential networks by [8]. In fact the special nature of belief functions involved in the KF as considered here make it possible to solve the problems without using the whole machinery of the TBM. Nevertheless for the more general models, we would use these evidential networks. Note that Kohlas and Monney [11, Ch.10] also present a formal solution for handling dynamic uncertainty as in the KF. Their solution is similar to the one presented here but it is derived within the hint model framework which is still based on the probability theory, whereas the TBM does not assume the existence of any underlying probabilities.

For the classification phase, we apply the General Bayesian Theorem which is detailed in [6, 5, 16]. The concept of doxastic independence is detailed in [2, 3]. Belief functions are applied on continuous spaces, a topic covered in [13, 22]. All proofs are omitted from this paper and can be found in a longer version [14].

2 JTC in the probabilistic framework

2.1 Problem formulation

The target state at time $t = 0, 1, 2, \dots$ is represented by n-dimensional vector $\boldsymbol{x}_t \in R^n$. Target class is a time-invariant attribute which takes values from a discrete set: $c \in C = \{c_1, \dots, c_s\}$. For simplicity let us assume that the state vector evolves according to the linear target motion model:

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}(c_i)\boldsymbol{x}_t + \boldsymbol{v}_t \tag{1}$$

where $\boldsymbol{F}(c_i)$ is $n \times n$ class dependent transition matrix and $\boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ is i.i.d. process noise. It is assumed that prior class probabilities $P(c_i)$ and prior conditional state densities $p(\boldsymbol{x}_0|c_i)$ are known and Gaussian for $i = 1, \dots, s$, i.e.

$$\boldsymbol{x}_0|c_i \sim \mathcal{N}(\boldsymbol{\mu}_0^i, \boldsymbol{\Sigma}_0^i)$$

Again for simplicity let us assume that kinematic measurements $\boldsymbol{z} \in R^m$ are linearly related to the target state:

$$\boldsymbol{z}_{t+1} = \boldsymbol{H}\boldsymbol{x}_{t+1} + \boldsymbol{w}_{t+1} \tag{2}$$

where $\boldsymbol{H}$ is $m \times n$ measurement matrix and $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$ is i.i.d. measurement noise. Furthermore, process and measurement noises are mutually independent. The optimal joint state and class estimator in the

Bayesian sense requires to construct at time t the posterior density $p(\boldsymbol{x}_t, c_i|\boldsymbol{Z}_t)$, where $\boldsymbol{Z}_t = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$ is the sequence of observations up to time t.

2.2 The Bayesian solution

The optimal Bayesian solution for JTC [10, 12] consists of a bank of class-matched filters as shown in Fig.1. Class probabilities are computed recursively as follows:

$$P(c_i|\boldsymbol{Z}_t) = \alpha\, p(\boldsymbol{z}_t|\boldsymbol{Z}_{t-1}, c_i)\, P(c_i|\boldsymbol{Z}_{t-1}) \qquad (3)$$

where $\Lambda_t^i = p(\boldsymbol{z}_t|\boldsymbol{Z}_{t-1}, c_i)$ is the likelihood of class i at time t and α is a normalising constant. Class-matched

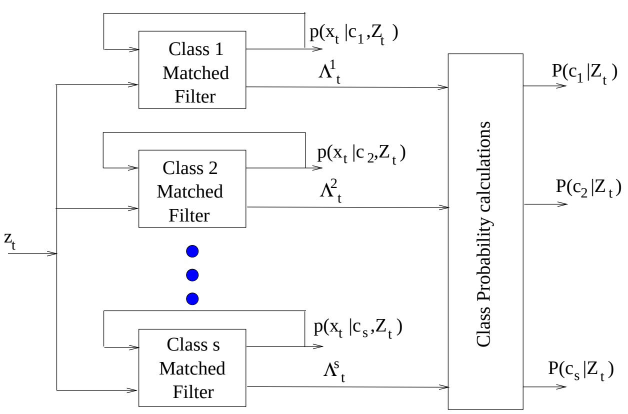

Fig. 1: JTC scheme

filters are typically built as Kalman filters or a weighted sum of Kalman filters, as described in the next example.

2.3 A motivating example with discussion

Suppose we have $s = 3$ classes of targets, and for classification we use their acceleration (that is manoeuvering) capabilities. Class 1 (c_1) are for example commercial planes with modest acceleration, class 2 (c_2) are large military aircraft (e.g. bomber) which can perform medium level acceleration and class 3 (c_3) are light and agile military aircraft (fighter planes) which are able to attain very large levels of acceleration.

To simplify analysis let us consider 1D geometry with state vector $\boldsymbol{x} = [x \quad \dot{x}]'$, and with a linear target motion model (for all three classes):

$$\boldsymbol{x}_{t+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} a + \boldsymbol{w}_t$$

where T is the sampling interval and a represents the input acceleration with limits $|a| \leq L_i$, where $L_i = 1$g, 3g and 5g for class $i = 1, 2, 3$ respectively (g= 9.81 m/s^2 is acceleration due to gravity). The measurements equation is given by (2) where $\boldsymbol{H} = [1 \;\; 0]$. The motion with small acceleration ($|a| < 1$g) we will refer to nearly constant velocity (CV) motion.

The question is how to implement the JTC scheme shown in Fig.1 for this example. The bank of filters in JTC scheme can be more easily tuned to detect the *behavior* of targets, rather than their class. Let $B = \{b_1, \ldots, b_n\}$ be the set of possible behaviors and $p^B(b_i|\boldsymbol{Z}_t)$ be a vector of behaviour probabilities. In our example $C = \{$commercial, bomber, fighter$\}$, while $B = \{$nearly CV, slow turn, sharp turn$\}$. The mapping rule between B and C can be: 'sharp turn implies fighter', 'slow turn implies fighter or bomber', and 'nearly CV implies fighter or bomber or commercial'. In general, behaviors are related to the classes by a matrix $\boldsymbol{M}$, where $\boldsymbol{M}_{ij} = p(c_i|b_j)$. The beliefs p^C on C are constructed using $p^C(.|\boldsymbol{Z}_t) = \boldsymbol{M} \cdot p^B(.|\boldsymbol{Z}_t)$.

How to choose matrix $\boldsymbol{M}$? When behaviors and classes are in one-to-one correspondence, indexes can be organized so that $b_i \equiv c_i$, in which case $\boldsymbol{M} = \boldsymbol{I}$ and $p^C(.|\boldsymbol{Z}_t) = p^B(.|\boldsymbol{Z}_t)$. In our example, however, it may be better to use the matrix

$$\boldsymbol{M} = \begin{bmatrix} 1/3 & 0 & 0 \\ 1/3 & 1/2 & 0 \\ 1/3 & 1/2 & 1 \end{bmatrix} \qquad (4)$$

since in this case if $p^B(b_1) = 1$, all three target classes will be of equal probability.

Next we analyse the classification results obtained using the JTC scheme of Fig.1 with identity matrix and the matrix of equation (4) for $\boldsymbol{M}$. The target is moving with CV in the first 36 scans, followed by 4 scans of acceleration with 2g, and then another 40 scans of CV motion. The sampling interval is $T = 3$ s.

The filters in JTC scheme are tuned to the behaviour set B: a KF is used for behaviour b_1; an IMM [1] with three modes (corresponding to $a \in \{-2a, 0, 2a\}$) for behaviour b_2; an IMM with 5 modes (corresponding to $a \in \{-4a, -2a, 0, 2a, 4a\}$) for behaviour b_3.

Fig.2 shows the classification results averaged over 20 Monte Carlo runs: (a) for $\boldsymbol{M} = \boldsymbol{I}$ and equi-priors; (b) for $\boldsymbol{M}$ given by (4) and equi-priors; (c) for $\boldsymbol{M}$ given by (4) and priors $p_0^B(b_1) = 0.9999$, $p_0^B(b_2) = p_0^B(b_3) = 0.00005$.

All classification results shown in Fig.2 appear unsatisfactory. The classifier of Fig.2(a) tends to classify a target moving with a CV as a class 1 target (commercial), while in reality all three classes can and often fly with (nearly) CV. The classifiers of Fig.2(b) and Fig.2(c) always tend to classify the target as a class 3 (fighter). Other choices of matrix $\boldsymbol{M}$ would not improve classification. This example serves as a motivation to consider the classification problem using the transferrable belief model. However, we need to look at the JTC problem in its entirety and therefore in the next section we derive the Kalman filter (the building block of class-matched filters) in the TBM framework.

3 Kalman filter in the TBM framework

We begin this section by presenting the background relations in the probabilistic framework which lead to the KF equations. Then we introduce the TBM background and re-derive the KF in the TBM framework.

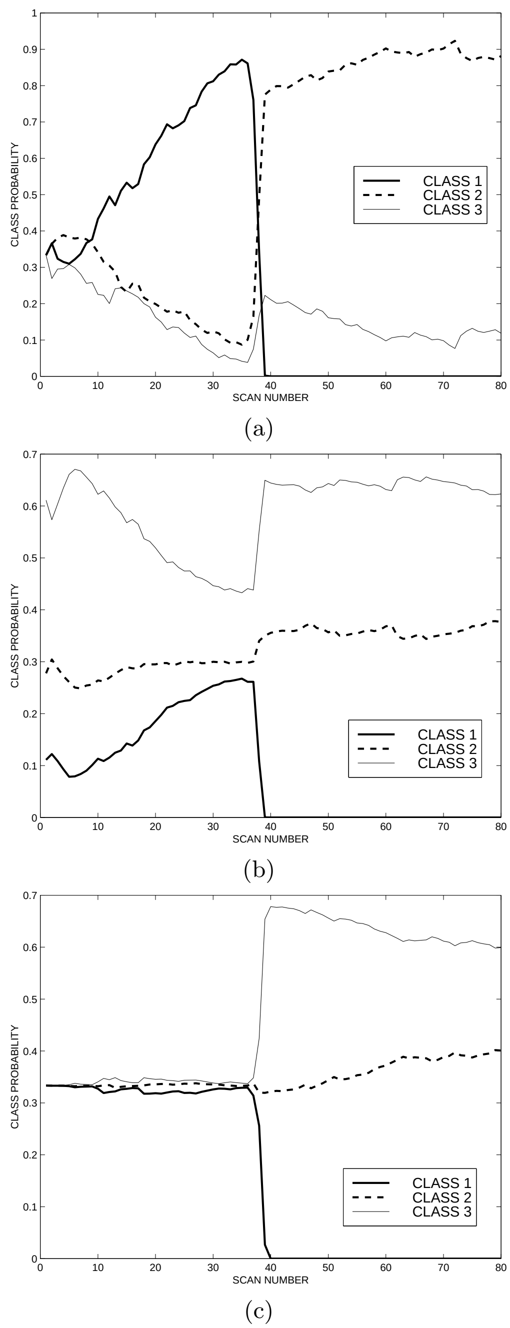

Fig. 2: Classification results using JTC scheme in the probabilistic framework

3.1 Background probabilistic relations

Relations and notation are strongly influenced by [1]. We have simplified the general model in the sense that the transition matrices and the noise parameters are not time dependant. Extension to the generalization (time-dependent) case is very simple, but notation becomes so cumbersome that the major underlying ideas get hidden.

Let $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the n-dimensional normal distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The $\sim$ in expressions like $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that the variable Y is a random variable which probability density function is Gaussian.

Lemma 3.1 *Let $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\boldsymbol{F}$ be a $m \times n$ matrix. Then $\boldsymbol{FY} \sim \mathcal{N}_m(\boldsymbol{F\mu}, \boldsymbol{F\Sigma F}')$.*

Lemma 3.2 *Let $\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) : i = 1, 2$ and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $m \times n$ matrices. If $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are stochastically independent, then $\boldsymbol{AY}_1 + \boldsymbol{BY}_2 \sim \mathcal{N}_m(\boldsymbol{A\mu}_1 + \boldsymbol{B\mu}_2, \boldsymbol{A\Sigma}_1\boldsymbol{A}' + \boldsymbol{B\Sigma}_2\boldsymbol{B}')$.*

Lemma 3.3 *Let $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Z}\end{bmatrix} \sim \mathcal{N}(\begin{bmatrix}\boldsymbol{\mu_X}\\ \boldsymbol{\mu_Z}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma_{XX}} & \boldsymbol{\Sigma_{XZ}}\\ \boldsymbol{\Sigma_{ZX}} & \boldsymbol{\Sigma_{ZZ}}\end{bmatrix})$. Then the probability density function about $\boldsymbol{X}$ given $\boldsymbol{Z} = \boldsymbol{z}$ is given by $\boldsymbol{X}|\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu_{X|z}}, \boldsymbol{\Sigma_{X|z}})$ where*

$$\boldsymbol{\mu_{X|z}} = \boldsymbol{\mu_X} + \boldsymbol{\Sigma_{XZ}\Sigma_{ZZ}^{-1}}(\boldsymbol{z} - \boldsymbol{\mu_Z})$$

and

$$\boldsymbol{\Sigma_{X|z}} = \boldsymbol{\Sigma_{XX}} - \boldsymbol{\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}}.$$

3.2 Kalman Filter Equations

Consider the dynamic and measurement equations introduced by (1) and (2), respectively. Suppose there is only one class and the prior state is Gaussian distributed, i.e.

$$\boldsymbol{x}_0 \sim N_n(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{5}$$

For simplicity sake, we assume that $m = n$, that the noise components are stationary (the $\boldsymbol{R}$ and $\boldsymbol{Q}$ matrices are time independent), and that the parameters of the system are not time-varying (the $\boldsymbol{F}$ and $\boldsymbol{H}$ matrices are time independent).

The KF performs filtering in two phases (prediction and update):

- the predicted state $\hat{\boldsymbol{x}}_{t|t-1}$ which concerns the prediction of $\boldsymbol{x}$ at time t given the measurement sequence $\boldsymbol{Z}_{t-1}$,
- the update state $\hat{\boldsymbol{x}}_{t|t}$ which concerns the update of $\boldsymbol{x}$ at time t given the new measurement $\boldsymbol{z}_t$.

These predictions are computed using the next relations.

Definition 3.1

$$\hat{\boldsymbol{x}}_{j|t} \triangleq E(\boldsymbol{x}_j|\boldsymbol{z}_t) \tag{6}$$

$$\boldsymbol{\Sigma}_{j|t} \triangleq E((\boldsymbol{x}_j - \hat{\boldsymbol{x}}_{j|t})(\boldsymbol{x}_j - \hat{\boldsymbol{x}}_{j|t})'|\boldsymbol{z}_t) \tag{7}$$

The relations at $t = 0$ are:

$$\hat{\boldsymbol{x}}_{0|0} = \boldsymbol{\mu}_0, \tag{8}$$
$$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0 \tag{9}$$

For $t = 1, 2, \ldots$, the relation are:

Predicted state

$$\boldsymbol{x}_{t|t-1} \sim \mathcal{N}(\boldsymbol{F}\hat{\boldsymbol{x}}_{t-1|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \tag{10}$$
$$\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{F}\hat{\boldsymbol{x}}_{t-1|t-1} \tag{11}$$
$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{F}\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q} \tag{12}$$

Update state

$$\hat{\boldsymbol{z}}_{t|t-1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{t|t-1} \tag{13}$$
$$\boldsymbol{\nu}_t = \boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1} \tag{14}$$
$$\boldsymbol{S}_t = \boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R} \tag{15}$$
$$\boldsymbol{W}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'\boldsymbol{S}_t^{-1} \tag{16}$$
$$\boldsymbol{x}_{t|t} \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \tag{17}$$
$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{W}_t\boldsymbol{\nu}_t \tag{18}$$
$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{W}_t\boldsymbol{S}_t\boldsymbol{W}_t' \tag{19}$$

The likelihood of the measurement sequence is required for classification. It is given by:

$$p(\boldsymbol{Z}_t) = \prod_{i=1,\ldots,t} p(\boldsymbol{z}_i|\boldsymbol{Z}_{i-1}) \tag{20}$$

where $p(\boldsymbol{z}_i|\boldsymbol{Z}_{i-1}) = \mathcal{N}(\boldsymbol{z}_i; \hat{\boldsymbol{z}}_{i|i-1}, \boldsymbol{S}_i)$. Thus:

$$p(\boldsymbol{Z}_t) = p(\boldsymbol{Z}_{t-1})\mathcal{N}(\boldsymbol{z}_i; \hat{\boldsymbol{z}}_{i|i-1}, \boldsymbol{S}_i).$$

3.3 TBM Background

The transferable belief model (TBM) [23, 17, 18] is a model to represent quantified beliefs based on the belief function theory developed by Shafer [15], but completely unrelated to any underlying probabilistic constraints as it is the case with the models of Dempster [7] and with the hint models [11]. These differences are not important here.

The essential tool is the basic belief assignment (bba) m^Ω which maps subsets of its domain Ω to $[0, 1]$. Its value $m^\Omega(A)$ for $A \subseteq \Omega$, called the basic belief mass (bbm), is the amount of belief that specifically supports that the actual value of the variable on which beliefs are expressed belongs to A, and that supports nothing more specific due to a lack of information, but that might support any strict subset of A if further information justify it.

3.3.1 Notations and Definitions

A credal variable is a variable for which we have defined a bba.

The symbol $m^{R^n}\{\boldsymbol{w}\}$ denotes the bba about the credal variable $\boldsymbol{w}$ which domain is R^n. The R^n subscript is often omitted.

A credal space is a triple $(\Omega, \mathcal{A}, m)$ where Ω is a set, $\mathcal{A}$ an algebra on Ω (closed under union, intersection, complement, with $\emptyset$ and Ω), and m is a bba on Ω. If the credal variable $\boldsymbol{x}$ is defined on a credal space $(\Omega, \mathcal{A}, m)$, we write $\boldsymbol{x} \sim m$.

A probability space is a particular credal space.

If Ω is finite or countable, $\mathcal{A} = 2^\Omega$.

If $\Omega = R$, $\mathcal{A}$ is a special subset of the power set of R [22, 13]: the Borel sigma-algebra on the set of real numbers, thus the sigma-algebra generated by the collection of closed intervals on the real numbers. As every sigma-algebra, it is closed under complements, countable unions and countable intersections. One can prove that it contains all open intervals, closed intervals, countably infinite unions or intersections of either.

If $\Omega = R^n$, $\mathcal{A}$ is the cross product of n Borel sigma-algebra as just defined.

A vacuous credal variable $\boldsymbol{x}$ on Ω is a credal variable $\boldsymbol{x}$ which $bel^\Omega\{\boldsymbol{x}\}$ function is a vacuous belief function. Thus its related plausibility function $pl^\Omega\{\boldsymbol{x}\}$ satisfies $pl^\Omega\{\boldsymbol{x}\}(A) = 1, \forall A \subseteq \Omega, A \neq \emptyset$. We denote it $\boldsymbol{x} \sim VBF$.

3.3.2 The TBM with Bayesian Belief Functions

To differentiate between a Gaussian pdf and a Gaussian Bayesian belief function, we use notation $\mathcal{N}$ for the former and NB for the latter. We neglect the n index as the dimension is clear from the context.

Definition 3.2 *$\boldsymbol{x} \sim NB(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\boldsymbol{x}$ is a n-dimensional credal variable on R^n whose basic belief density allocates non zero densities only to the singletons of R^n and these densities are those of the n-dimensional Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.*

Let $m\{\boldsymbol{x}\}$ denote the basic belief density (bbd) that represents the basic belief assignment (bba) relative to $\boldsymbol{x}$. Then:

$$m\{\boldsymbol{x}\}(A) = \begin{cases} \mathcal{N}(\boldsymbol{a}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) & \textit{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ 0 & \textit{otherwise.} \end{cases} \tag{21}$$

We also use the notation $m\{\boldsymbol{x}\} = NB(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and we refer to $m\{\boldsymbol{x}\}$ as to the Gaussian bbd.

Next we present a few useful lemmas, which are proved in [14].

Lemma 3.4 *If $\boldsymbol{x} \sim NB(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{F}\boldsymbol{x} \sim NB(\boldsymbol{F}\boldsymbol{\mu}, \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}')$.*

Lemma 3.5 *Let $\boldsymbol{x} \sim NB(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $\boldsymbol{y} \sim NB(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$. Let $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}$. Then $\boldsymbol{z} \sim NB(\boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{B}\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{B}')$.*

We use $\bigcirc\!\!\!\!\cap$ to denote the conjunctive combination rule. We use $\oplus$ to denote the Dempster's rule , thus the normalized conjunctive combination rule.

Lemma 3.6 *Let $\boldsymbol{x} \in R^n$. Let $m_1\{\boldsymbol{x}\} = NB(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $m_2\{\boldsymbol{x}\} = NB(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ be two bbd relative to $\boldsymbol{x}$. Then (after normalization)*

$$m_{1\oplus 2}\{\boldsymbol{x}\} = m_1\{\boldsymbol{x}\} \oplus m_2\{\boldsymbol{x}\} = NB(\boldsymbol{\nu}, \boldsymbol{S})$$

where $\boldsymbol{S}^{-1} = \boldsymbol{\Sigma_1}^{-1} + \boldsymbol{\Sigma_2}^{-1}$ and $\boldsymbol{\nu} = \boldsymbol{S}(\boldsymbol{\Sigma_1}^{-1}\boldsymbol{\mu_1} + \boldsymbol{\Sigma_2}^{-1}\boldsymbol{\mu_2})$.

Lemma 3.7 *Let $\boldsymbol{x} \sim NB(\boldsymbol{\mu_X}, \boldsymbol{\Sigma_X})$ and $\boldsymbol{w} \sim NB(\boldsymbol{\mu_W},\ \boldsymbol{R})$. Let $\boldsymbol{z} = \boldsymbol{Ax} + \boldsymbol{w}$. Then the conditional bbd on $\boldsymbol{x}$ given $\boldsymbol{z}$ is given by*

$$\boldsymbol{x}|\boldsymbol{z} \sim NB(\boldsymbol{A^{-1}z}, \boldsymbol{A^{-1}}\ \boldsymbol{RA^{-1'}}).$$

Lemma 3.8 *Let $\boldsymbol{x} \in R^n$. Let f_1 and f_2 are two pdf on R^n. For $i = 1, 2$, let $m_i\{\boldsymbol{x}\} = f_iB$ be two bbd relative to $\boldsymbol{x}$ with*

$$m_i\{\boldsymbol{x}\}(A) = \begin{cases} f_i(\boldsymbol{a}) & \text{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Then

$$m_{1 \cap 2}\{\boldsymbol{x}\}(A) = m_1\{\boldsymbol{x}\} \cap m_2\{\boldsymbol{x}\}(A) \tag{23}$$

$$= \begin{cases} f_1(\boldsymbol{a})f_2(\boldsymbol{a}) & \text{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Lemma 3.9 *When bbd's are Gaussian, the TBM degrades into probability theory.*

3.4 Decision making in TBM

When a decision must be made and uncertainty is represented by the TBM, the decision maker derives the so called pignistic probability function using the pignistic transformation. The result is just a probability function that is used to make decisions using the classical expected utility theory. Given a bba m^Ω on Ω, the pignistic probability function (denoted $BetP$) is defined as :

$$BetP(Y) = \sum_{X \subseteq \Omega} \frac{|Y \cap X|}{|X|} \frac{m^\Omega(X)}{1 - m^\Omega(\emptyset)},\ \forall Y \subseteq \Omega$$

The nature and justification of the pignistic transformation are presented in [21, 23].

When the credal variable X is defined on R, we end up with a pignistic probability density function $Betf$ which is defined as:

$$Betf(a) = \lim_{\varepsilon \to 0} \int_{x=-\infty}^{x=a} \int_{y=a+\varepsilon}^{y=\infty} \frac{m^R(X \in [x, y])}{y - x} dydx. \tag{25}$$

The next lemma shows that the major property of stochastic independence holds between pignistic probabilities.

Lemma 3.10 *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two doxasticaly independent credal variables defined on R^{n_1} and R^{n_2}, respectively. Let $Betf\{\boldsymbol{X}\}$ and $Betf\{\boldsymbol{Y}\}$ be their pignistic transformations. The pignistic transformation of their joint belief on $R^{n_1+n_2}$ satisfies $\forall \boldsymbol{x} \subseteq \boldsymbol{X}, \boldsymbol{y} \subseteq \boldsymbol{Y}$:*

$$Betf\{\boldsymbol{X}, \boldsymbol{Y}\}(\boldsymbol{x}, \boldsymbol{y}) = Betf\{\boldsymbol{X}\}(\boldsymbol{x}) \cdot Betf\{\boldsymbol{Y}\}(\boldsymbol{y}).$$

3.5 Kalman Filter in the TBM Framework

We will

1. keep assumptions (1) and (2),
2. relax (5) into $\boldsymbol{x}_0 \sim VBF$,
3. accept that $\boldsymbol{v}_t$ is a credal variable; its bbd is unknown but its pignistic transformation $Betf\{\boldsymbol{v}\}$ is a Gaussian pdf: $Betf\{\boldsymbol{v}\} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$,
4. accept that $\boldsymbol{w}_t \sim NB(\boldsymbol{0}, \boldsymbol{R})$,
5. select the most committed admissible bbd for the credal variable $\boldsymbol{x}_{t+1|t}$
6. use the General Bayesian Theorem for the final classification phase with a vacuous *a priori* belief on B and a $\boldsymbol{M}$ matrix that better translates the implication rules.

The whole prediction phase will be essentially the same as with the probability approach. The gain obtained by relaxing (5) about $\boldsymbol{x}_0$ is really not that essential. All it does is avoiding the "quarrel" about the choice of an adequate prior on $\boldsymbol{x}_0$. Relaxing constraints on $\boldsymbol{v}_t$ is surely more impressive, as we do not even require knowing its underlying bbd, all we use is its pignistic transformation. Selecting the most committed admissible bbd is arguable, but our purpose is to show what are the assumptions underlying the Kalman filter relations within the TBM, not to build a new set of Kalman filter relations.

What is interesting is that the classical relations used in the Kalman filter can be justified within the TBM.

Major discrepancies appear in the classification phase. For the previous example of section 2.3, we will use conditional belief functions that represent exactly the rules between classes and behaviors. Such a representation is not achievable in probability theory.

Given the relaxed assumptions, in particular the third one about $\boldsymbol{v}_t$, the bbd's of $\boldsymbol{x}_{t+1|t}$ and $\boldsymbol{x}_{t+1|t+1}$ cannot be derived. Still we know their related pignistic probabilities, and as far as this happens to be all we need in practice, ignoring the bbd's is not a real issue.

As the initial assumptions on $\boldsymbol{x}_0$ are not those of the classical Kalman filters, we must derive the properties of $\boldsymbol{x}_{1|0}$, $\boldsymbol{z}_1$ and $\boldsymbol{x}_{1|1}$.

We then proceed with time $t = 2$ and derive the properties of $\boldsymbol{x}_{2|1}$, $\boldsymbol{z}_2$ and $\boldsymbol{x}_{2|2}$ which turns out to be those of the classical KF for what concerns their pignistic probabilities. We can just proceed then as with the classical KF relations.

3.5.1 *Predicted state for $t = 1$: bba on $\boldsymbol{x}_1$ induced by $\boldsymbol{x}_0$*

From (1) we have:

$$\boldsymbol{x}_1 = \boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{v}_1$$

Since $\boldsymbol{x}_0 \sim VBF$, then $\boldsymbol{F}\boldsymbol{x}_0 \sim VBF$.

We ignore $m^{R^n}\{\boldsymbol{v}_1\}$. Still whatever $m^{R^n}\{\boldsymbol{v}_1\}$, $\boldsymbol{x}_1 \sim VBF$. This reflects the natural rule that if one adds two terms, one being completely unknown, the result is also completely unknown. Formally, we combine the two bbd's after vacuously extending them on their product space.

$$m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_1\} = m^{R^n}\{\boldsymbol{F}\boldsymbol{x}_0\}^{\uparrow R^{2n}} \circledcap m^{R^n}\{\boldsymbol{v}_1\}^{\uparrow R^{2n}}$$

The bbd satisfies:

$$m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_1\}(\boldsymbol{x}, \boldsymbol{v}) = \begin{cases} m^{R^n}\{\boldsymbol{v}_1\}(\boldsymbol{v}) & \text{if } \boldsymbol{x} = R^n, \\ & \text{and } \boldsymbol{v} \subseteq R^n \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

The bbd $m^{R^n}\{\boldsymbol{x}_1\}$ is the result of a coarsening of $m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_1\}$ which is a vacuous belief function ($\boldsymbol{x}_1 \sim VBF$) as every mass of $m^{R^n}\{\boldsymbol{v}_1\}^{\uparrow R^{2n}}$ intersects every cylindrical extension of $\boldsymbol{x}_1$ in R^{2n}.

3.5.2 *Updated state for $t = 1$: bba on $\boldsymbol{x}_1$ induced by $\boldsymbol{x}_0$ and $\boldsymbol{z}_1$*

From (2) we have:

$$\boldsymbol{z}_1 = \boldsymbol{H}\boldsymbol{x}_1 + \boldsymbol{w}_1$$

The argument used to show $\boldsymbol{x}_1 \sim VBF$ leads similarly to $\boldsymbol{z}_1 \sim VBF$.

For $\boldsymbol{x}_{1|1}$, we have $\boldsymbol{x}_{1|1} = \boldsymbol{H}^{-1}(\boldsymbol{z}_1 - \boldsymbol{w}_1)$ where $\boldsymbol{z}_1$ is the observed value and $\boldsymbol{w}_1 \sim NB(\boldsymbol{0}, \boldsymbol{R})$. Therefore $\boldsymbol{x}_{1|1} \sim NB(\hat{\boldsymbol{x}}_{1|1}, \boldsymbol{\Sigma}_{1|1})$ where $\hat{\boldsymbol{x}}_{1|1} = \boldsymbol{H}^{-1}\boldsymbol{z}_1$ and $\boldsymbol{\Sigma}_{1|1} = \boldsymbol{H}^{-1}\boldsymbol{R}\boldsymbol{H}'^{-1}$.

3.5.3 *Predicted state for $t = 2$: bba on $\boldsymbol{x}_{2|1}$ induced by $\boldsymbol{x}_{1|1}$*

The input is $\boldsymbol{x}_{1|1} \sim NB(\hat{\boldsymbol{x}}_{1|1}, \boldsymbol{\Sigma}_{1|1})$ and $\boldsymbol{v}_2$ for which we ignore its bbd but know its pignistic probabilities. Thus we know the pignistic probabilities of both variables, and thus we know the pignistic probability function on their joint space, out of which we compute (by convolution) the pignistic probabilities relative to their sum. For what concerns the pignistic probabilities, probability theory applies and we have: $Betf\{\boldsymbol{x}_{2|1}\} = N(\hat{\boldsymbol{x}}_{2|1}, \boldsymbol{\Sigma}_{2|1})$ where $\hat{\boldsymbol{x}}_{2|1} = F\hat{\boldsymbol{x}}_{1|1}$ and $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{F}\boldsymbol{\Sigma}_{1|1}\boldsymbol{F}' + \boldsymbol{Q}$. Still, the bbd of $\boldsymbol{x}_{2|1}$ is unknown because the bbd of $\boldsymbol{v}_2$ is unknown.

3.5.4 *Updated state for $t = 2$: bba on $\boldsymbol{x}_{2|2}$ induced by $\boldsymbol{x}_{2|1}$ and $\boldsymbol{z}_2$*

The input is $\boldsymbol{x}_{2|1}$ which bbd is unknown but which pignistic probability function is gaussian and $\boldsymbol{w}_1$ which bbd is is a gaussian Bayesian belief function.

For $\boldsymbol{z}_2$, we ignore its bbd but we know its pignistic probability function: $Betf\{\boldsymbol{z}_2\} = N(\hat{\boldsymbol{z}}_{2|1}, \boldsymbol{S}_2)$ where $\hat{\boldsymbol{z}}_{2|1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{2|1}$ and $\boldsymbol{S}_2 = \boldsymbol{H}\boldsymbol{\Sigma}_{2|1}\boldsymbol{H}' + \boldsymbol{R}$.

For $\boldsymbol{x}_{2|2}$, if we knew the bbd $m^{R^n}\{\boldsymbol{x}_{2|1}\}$, we would vacuously extend it on the R^{2n} space. We would also vacuously extend the belief about $\boldsymbol{w}_2$ on the R^{2n} space, and combine these two bbd's with the conjunctive combination rule. Then we condition the result on the observation $\boldsymbol{z}_2$, and marginalise the result on $\boldsymbol{x}_{2|2}$ domain. This last bbd would be the bbd on $\boldsymbol{x}_{2|2}$ induced by $\boldsymbol{x}_{2|1}$ and $\boldsymbol{z}_2$.

We can then invoke the principle of maximum commitment that states: keep as much information as possible, and select for $\boldsymbol{x}_{2|1}$ the most committed normalized bbd among those which pignistic transformation is $Betf\{\boldsymbol{x}_{2|1}\}$. The solution is the Bayesian belief function corresponding to the pignistic probability function. Thus $\boldsymbol{x}_{2|1} \sim NB(\hat{\boldsymbol{x}}_{2|1}, \boldsymbol{\Sigma}_{2|1})$. In that case, both belief functionson R^{2n} are bbfs, and probability theory just applies.

We get $\boldsymbol{x}_{2|2} \sim NB(\hat{\boldsymbol{x}}_{2|2}, \boldsymbol{\Sigma}_{2|2})$ where $\hat{\boldsymbol{x}}_{2|2}$ and $\boldsymbol{\Sigma}_{2|2}$ are given by relations (18) and (19), respectively.

Being back into the classical KF relations setting, we can proceed just as with a classical KF for the prediction phases. The only particularity is that the pignistic probability function on $\boldsymbol{z}_t$ is known but not its bbd. So for the classification phase, we will have to reconsider the procedure.

3.6 Diffuse prior and TBM solution

It is worth mentioning that the TBM solution is the same solution one would obtained if one used the probabilistic approach with a diffuse (also called improper or uninformative) prior on $\boldsymbol{x}_0$. So one might ask why bother with the TBM, if all it achieves is what probability theory produces with a diffuse prior.

An answer is that a 'diffuse pdf' is not a pdf, and its use violates the foundation of probability theory, even thought its users claim to be using probability theory. The TBM also violates the foundation of probability theory, but purposely and indeed we never claim to be using probability theory.

Another answer is that even though diffuse priors produce the TBM solution, we can also consider other priors, and the flexibility of the TBM allows us to use any prior on $\boldsymbol{x}_0$, as well as for $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$. Our presentation is limited to a context very similar to the one used with probability theory, but can be generalized, what will be studied in future works.

4 JTC in the TBM framework

JTC in the TBM framework is done conceptually in a similar manner as in the Bayesian framework. Again we have a bank of tracking filters matched to target behaviour or class, which (according to the previous section) can be based on classical Kalman filters. The main difference, however, is in the way TBM performs classification. Let us denote by $l_i = p(\boldsymbol{Z}_t|b_i)$ the measurement likelihoods which are output from filter i matched to behaviour b_i, see eq.(20).

We consider a vacuous a priori on B. Then the General Bayesian Theorem (GBT) permits us to compute the posterior belief on B. A suitable formula for GBT is [6, 5, 16]:

$$m^B[\boldsymbol{Z}_t](b) = \prod_{i:b_i \in b} l_i \prod_{j:b_j \notin b} (1 - l_j), \ \forall b \subseteq B.$$

The bbd of the credal variable $\boldsymbol{z}_t$ is unknown, the likelihood cannot be directly derived. But as we know the pignistic probability function of $\boldsymbol{z}_t$, we can applied the least committed prionciple that states: never allocate more belief than necessary. It means we select the q-least committed belief function among those which pignistic transformation is the known one. The solution is based on:

$$\begin{aligned} pl([x, y]) =& 2(x - \mu)\mathcal{N}(x; \mu, \Sigma) \\ &+ \int_{t=x}^{t=\infty} 2\mathcal{N}(t; \mu, \Sigma)dt, \ y \geq x \geq \mu \end{aligned}$$

and for $\mu \in [x, y]$, $pl([x, y]) = 1$ (see details in [22]).

The relation between B and C can be established in a precise manner on the power set. This is explained using the example of section 2.3. In this example the relation between B and C is described by three conditional belief functions:

$$m^C[b_1](\{c_1, c_2, c_3\}) = 1, \quad (27)$$
$$m^C[b_2](\{c_2, c_3\}) = 1, \text{ and} \quad (28)$$
$$m^C[b_3](\{c_3\}) = 1. \quad (29)$$

Let us represent the bba's as vectors whose elements are ordered as follows (for m^B):

$$\emptyset, \{b_1\}, \{b_2\}, \{b_1, b_2\}, \{b_3\}, \{b_1, b_3\}, \{b_2, b_3\}, \{b_1, b_2, b_3\},$$

and similarly for m^C. Then, the derivation of the bba on C given the bba on B and the three conditional belief functions (27)–(29) is achieved using matrix $\overline{\boldsymbol{M}}$ as follows:

$$m^C = \overline{\boldsymbol{M}} m^B,$$

where:

$$\overline{\boldsymbol{M}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

This transformation means that

- $m^B(\emptyset)$ is transferred to $m^C(\emptyset)$
- $m^B(\{b_1\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_2\})$ is transferred to $m^C(\{c_2, c_3\})$ as b_2 implies $\{c_2, c_3\}$
- $m^B(\{b_1, b_2\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_2 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_3\})$ is transferred to $m^C(\{c_3\})$ as b_3 implies $\{c_3\}$
- $m^B(\{b_1, b_3\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_3 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_2, b_3\})$ is transferred to $m^C(\{c_2, c_3\})$ as b_2 or b_3 implies $\{c_2, c_3\}$
- $m^B(\{b_1, b_2, b_3\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_2 or b_3 implies $\{c_1, c_2, c_3\}$

These rules are based on the logical property: if a implies x and b implies y, then a or b implies x or y.

Next we reconsider the example of section 2.3. If $l_1 = 1$, $l_2 = l_3 = 0$, then $m^B[\boldsymbol{Z}_t](b_1) = 1$ and $m^C[\boldsymbol{Z}_t](C) = 1$. Thus the a posteriori on C is vacuous, as it should be.

If $l_1 = l_2 = l_3 = 1$, then $m^B[\boldsymbol{Z}_t](B) = 1$ and $m^C[\boldsymbol{Z}_t](C) = 1$. Thus the a posteriori on C is also vacuous, as it should be.

To further illustrate the theory, we repeat the same experimental setup from section 2.3, with the KF, IMM-3 and IMM-5 filters tuned to the target behaviour. We show the resulting pignistic class probabilities, obtained using the described TBM classifier, in Fig.3 (obtained by averaging over 20 Monte Carlo runs). The classification results appear reasonable: before the manoeuver all three target classes are equally probable, while after the manoeuver (which can be performed only by class 2 and 3 targets), the probability of class 1 drops to zero while the probability of class 2 and 3 target jumps to 1/2. The TBM classifier thus resolves the issues raised in section 2.3.

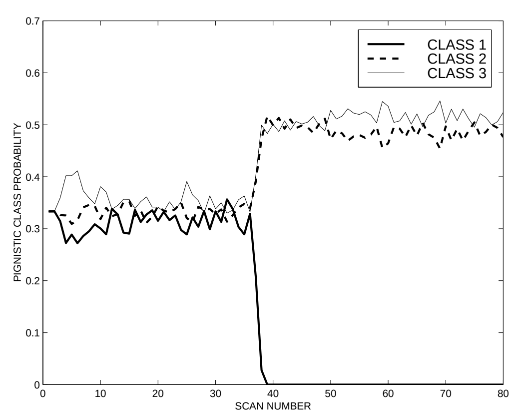

Fig. 3: Classification results using JTC scheme in the TBM framework

5 Conclusions

We have shown that it is possible to derive the Kalman filter within the TBM framework. The TBM solution for the tracking (filtering) phase of JTC is essentially the same as the one achieved within the probability framework. In derivation, however, several assumptions have been already seriously relaxed, and further generalisation is perfectly possible. For the classification phase of JTC, where the probabilistic approach essentially fails, the TBM offers a solution which is intuitively satisfactory. The overall TBM solution is theoretically sound and coherent, and provides a seemingly better framework for joint tracking and classification.

Acknowledgements

B. Ristic would like to thank IRIDIA, Université libre de Bruxelles, for being his host while the research presented in this paper has been carried out.

References

[1] V. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with applications to tracking and navigation.* Wiley, New York, 2001.

[2] B. Ben Yaghlane, Ph. Smets, and K. Mellouli. Belief function independence: I. The marginal case. *Int. J. Approximate Reasoning*, 29:47–70, 2001.

[3] B. Ben Yaghlane, Ph. Smets, and K. Mellouli. Belief function independence: II. The conditional case. *Int. J. Approximate Reasoning*, 31:31–75, 2002.

[4] S. Challa and G. W. Pulford. Joint target tracking and classification using radar and ESM sensors. *IEEE Trans. Aerospace and Electronic Systems*, 37(3):1039–1055, 2001.

[5] F. Delmotte and P. Smets. Target identification based on the transferable belief model interpretation of Dempster-Shafer model. Part 2: Applications. *Submitted*, 2002.

[6] F. Delmotte and P. Smets. Target identification based on the transferable belief model interpretation of Dempster-Shafer model. *IEEE Trans. Systems, Man, and Cybernetics - Part A*, 2004. Accepted for publication.

[7] A. P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Ann. Math. Statistics*, 38:325–339, 1967.

[8] A.P. Dempster. Construction and local computation aspects of network belief functions. In R. M. Oliver and J. Q. Smith, editors, *Influencce diagrams, belief nets and decision analysis*, pages 121–141. John Wiley and Sons, Chichester, 1990.

[9] A. Farina, P. Lombardo, and M. Marsella. Joint tracking and identification algorithms for multisensor data. *IEE Proc. Radar Sonar Navig.*, 149(6):271–280, 2002.

[10] N. Gordon, S. Maskell, and T. Kirubarajan. Efficient particle filters for joint tracking and classification. In *Proc. SPIE*, volume 4728, pages 439–449, 2002.

[11] J. Kohlas and P. A. Monney. *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence.* Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag, 1995.

[12] B. Ristic, N. Gordon, and A. Bessell. On target classification using kinematic data. *Information Fusion*, 5:15–21, 2004.

[13] B. Ristic and Ph. Smets. Belief function theory on the continuous space with an application to model based classification. In *Proc. Information Processing and Management of Unceratinty (IPMU)*, Perugia, Italy, July 2004. Submitted.

[14] B. Ristic and Ph. Smets. Kalman filters for tracking and classification and the transferable belief model. Technical report, IRIDIA, Université Libre de Bruxelles, Belgium., 50 av. Roosevelt, CP194/6, 1050 Brussels, Belgium, 2004. In preparation.

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

[16] Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *Int. J. Approximate Reasoning*, 9:1–35, 1993.

[17] Ph. Smets. What is Dempster-Shafer's model? In R. R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 5–34. Wiley, New York, 1994.

[18] Ph. Smets. The transferable belief model for quantified belief representation. In D. M. Gabbay and Ph. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer, Doordrecht, The Netherlands, 1998.

[19] Ph. Smets. Practical uses of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 99*, pages 612–621. Morgan Kaufman, San Francisco, Ca, 1999.

[20] Ph. Smets. Data fusion in the transferable belief model. In *Proc. 3rd Int. Conf. Information Fusion*, pages PS 21–33. FUSION 2000, Paris, France, 2000.

[21] Ph. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava and T.J. Mock, editors, *Belief Functions in Business Decisions*, pages 17–61. Physica-Verlag, Heidelberg, Germany, 2002.

[22] Ph. Smets. Belief functions on continuous spaces. *International Journal of Approximate Reasoning*, 2004.

[23] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.