This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3024858, IEEE Signal Processing Letters

Classification-Aided Multitarget Tracking Using the Sum-Product Algorithm

Domenico Gaglione*, Giovanni Soldi*, Paolo Braca, *Senior Member, IEEE*, Giovanni De Magistris, Florian Meyer, *Member, IEEE*, and Franz Hlawatsch, *Fellow, IEEE*

***Abstract*—Multitarget tracking (MTT) is a challenging task that aims at estimating the number of targets and their states from measurements of the target states provided by one or multiple sensors. Additional information, such as imperfect estimates of target classes provided by a classifier, can facilitate the target-measurement association and thus improve MTT performance. In this letter, we describe how a recently proposed MTT framework based on the sum-product algorithm can be extended to efficiently exploit class information. The effectiveness of the proposed approach is demonstrated by simulation results.**

***Index Terms*—Multitarget tracking, probabilistic data association, sum-product algorithm, classification, factor graph.**

I. Introduction

Multitarget tracking (MTT) [1], [2] aims at estimating the number of targets and their states from measurements provided by one or multiple sensors. A major challenge in MTT is posed by measurement origin uncertainty (MOU) [1], i.e., the fact that it is not known if a measurement is produced by a target, and by which target. A Bayesian message passing algorithm that efficiently addresses MOU was presented in [3]. This algorithm has been used to develop an MTT tracking method [4], [5] that employs the sum-product algorithm (SPA) [6]–[8] to approximate the marginal posterior probability density functions (pdfs) of the target states. SPA-based MTT was demonstrated to be highly scalable and to outperform several previously proposed MTT methods [4], [5]. Moreover, the flexibility of the SPA approach allows for several extensions such as integration of data provided by heterogenous sensors [9], [10] and adaptation to time-varying parameters [11].

Compared to processing only the sensor measurements, as is done by most existing MTT methods (e.g., [4], [5], [11]–[23]), exploiting additional target-related information generally leads to improved MTT performance. Here, we consider the exploitation of imperfect *class* information [24], [25], which assigns a target to one among several categories. In the maritime domain, for example, such categories could be commercial ships, military ships, and fishing boats. This class information is generally the output of a classifier and allows, e.g., the use of class-dependent motion or measurement models.

Feature-aided and classification-aided tracking techniques [24]–[31] have lately encountered a growing interest, in particular following recent advances in deep learning classification methods [32]–[34]. In this letter, we show how imperfect target class information can be integrated into the SPA-based MTT method of [4], [5]. More concretely, we propose an SPA-based MTT method that takes into account the output of a classifier distinguishing between target-related classes as well as between target- and clutter-originated measurements. Our simulation results show that the proposed classification-aided MTT method outperforms the MTT method of [4].

The remainder of this letter is organized as follows. Section II presents the system model and its statistical formulation. Section III describes the proposed method. Section IV reports simulation results.

II. System Model and Statistical Formulation

The system model described in this section recalls the one introduced in [4] and, in addition, establishes statistical characterizations of the target class and of the imperfect class information provided by the classifier.

A. Target Model

We consider K *potential targets* (PTs) indexed by $k \in \mathcal{K} \triangleq \{1,\ldots,K\}$, whose existence at time n is indicated by $r_{n,k} \in \{0,1\}$, i.e., $r_{n,k} = 1$ if PT k exists and $r_{n,k} = 0$ otherwise. Note that K is an upper limit of the number of actual targets that can be tracked simultaneously, i.e., the (time-varying) number of actual targets need not be known in advance except that it is not larger than K. (Scalable SPA-based multisensor MTT methods with a time-varying number of PTs K are presented in [5].) The state of PT k at time n is represented by the vector $\boldsymbol{x}_{n,k}$ and consists of the PT's position and possibly further parameters, such as the PT's velocity; a "dummy" state is formally considered also if $r_{n,k} = 0$. Each PT k belongs to one of C distinct classes $c \in \{1,\ldots,C\}$ at time n; the class of PT k at time n is specified by the class index variable $\ell_{n,k} \in \{1, \ldots, C\}$, which is unknown just as $\boldsymbol{x}_{n,k}$ and $r_{n,k}$. We allow $\ell_{n,k}$ to be time-dependent, although in most applications it is constant. The time evolution of the state of a PT k that exists at times $n-1$ and n (i.e., $r_{n-1,k} = r_{n,k} = 1$) is modeled as $\boldsymbol{x}_{n,k} = \boldsymbol{\theta}_{\ell_{n,k}}\big(\boldsymbol{x}_{n-1,k}, \boldsymbol{u}_{n,k}^{(\ell_{n,k})}\big)$, where $\boldsymbol{u}_{n,k}^{(\ell_{n,k})}$ is a process noise that is independent and identically distributed (iid) across n and k. The state transition function $\boldsymbol{\theta}_{\ell_{n,k}}(\cdot,\cdot)$ is selected from a set $\{\boldsymbol{\theta}_c(\cdot,\cdot)\}_{c=1}^{C}$ by the class index variable $\ell_{n,k}$.

This work was supported in part by the NATO Allied Command Transformation (ACT) under the DKOE project and by the Austrian Science Fund (FWF) under grants J 3886-N31 and P 32055-N31. D. Gaglione, G. Soldi, P. Braca, and G. De Magistris are with the NATO Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy (e-mail: [domenico.gaglione, giovanni.soldi, paolo.braca, giovanni.demagistris]@cmre.nato.int). F. Meyer is with the Scripps Institution of Oceanography and the Electrical and Computer Engineering Department, University of California San Diego, La Jolla, CA, USA (e-mail: flmeyer@ucsd.edu). F. Hlawatsch is with the Institute of Telecommunications, TU Wien, Vienna, Austria (e-mail: franz.hlawatsch@tuwien.ac.at).

* Co-first authors.

Furthermore, also the statistics of $\boldsymbol{u}_{n,k}^{(\ell_{n,k})}$ generally depend on $\ell_{n,k}$. The function $\boldsymbol{\theta}_{\ell_{n,k}}(\cdot,\cdot)$ and the statistics of $\boldsymbol{u}_{n,k}^{(\ell_{n,k})}$ define the state transition pdf $f(\boldsymbol{x}_{n,k}|\boldsymbol{x}_{n-1,k}, \ell_{n,k})$. In addition to the dynamics of the PTs, also other PT characteristics (e.g., color, size, type) may be related to the PT class.

For convenience, we define the *augmented state* vector $\boldsymbol{y}_{n,k} \triangleq [\boldsymbol{x}_{n,k}^{\mathrm{T}}, r_{n,k}, \ell_{n,k}]^{\mathrm{T}}$ as well as the stacked vector $\boldsymbol{y} \triangleq [\boldsymbol{y}_0^{\mathrm{T}}, \ldots, \boldsymbol{y}_n^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{y}_n \triangleq [\boldsymbol{y}_{n,1}^{\mathrm{T}}, \ldots, \boldsymbol{y}_{n,K}^{\mathrm{T}}]^{\mathrm{T}}$. The time evolution of the augmented state of a PT k is statistically described by the transition pdf $f(\boldsymbol{y}_{n,k}|\boldsymbol{y}_{n-1,k})$. Assuming that $\boldsymbol{x}_{n,k}$ and $r_{n,k}$ are conditionally independent of $\ell_{n-1,k}$ given $\boldsymbol{x}_{n-1,k}$, $r_{n-1,k}$, and $\ell_{n,k}$, this transition pdf can be obtained as

$$\begin{aligned} f(\boldsymbol{y}_{n,k}|\boldsymbol{y}_{n-1,k}) &= f(\boldsymbol{x}_{n,k}, r_{n,k}, \ell_{n,k}|\boldsymbol{x}_{n-1,k}, r_{n-1,k}, \ell_{n-1,k}) \\ &= f(\boldsymbol{x}_{n,k}, r_{n,k}|\ell_{n,k}, \boldsymbol{x}_{n-1,k}, r_{n-1,k}, \ell_{n-1,k}) \\ &\quad \times p(\ell_{n,k}|\boldsymbol{x}_{n-1,k}, r_{n-1,k}, \ell_{n-1,k}) \\ &= f(\boldsymbol{x}_{n,k}, r_{n,k}|\ell_{n,k}, \boldsymbol{x}_{n-1,k}, r_{n-1,k}) \\ &\quad \times p(\ell_{n,k}|\boldsymbol{x}_{n-1,k}, r_{n-1,k}, \ell_{n-1,k}). \end{aligned} \tag{1}$$

Here, an expression of $f(\boldsymbol{x}_{n,k}, r_{n,k}|\ell_{n,k}, \boldsymbol{x}_{n-1,k}, r_{n-1,k})$ is provided in [11, Sec. II-C]. If the PT's dynamic model is independent of $\ell_{n,k}$, then $f(\boldsymbol{x}_{n,k}, r_{n,k}|\ell_{n,k}, \boldsymbol{x}_{n-1,k}, r_{n-1,k}) = f(\boldsymbol{x}_{n,k}, r_{n,k}|\boldsymbol{x}_{n-1,k}, r_{n-1,k})$; an expression of this pdf is provided in [4, Sec. II-A]. Furthermore, the transition probability mass function (pmf) $p(\ell_{n,k}|\boldsymbol{x}_{n-1,k}, r_{n-1,k}, \ell_{n-1,k})$ is given as follows. If PT k did not exist at time $n-1$, i.e., $r_{n-1,k} = 0$, then $p(\ell_{n,k}|\boldsymbol{x}_{n-1,k}, 0, \ell_{n-1,k}) = 1/C$. Else, if PT k existed at time $n-1$, i.e., $r_{n-1,k} = 1$, then $p(\ell_{n,k}|\boldsymbol{x}_{n-1,k}, 1, \ell_{n-1,k})$ is described by the transition matrix $\boldsymbol{D}(\boldsymbol{x}_{n-1,k}) \in [0,1]^{C\times C}$, with $[\boldsymbol{D}(\boldsymbol{x}_{n-1,k})]_{i,j} = p(\ell_{n,k} = i|\boldsymbol{x}_{n-1,k}, 1, \ell_{n-1,k} = j)$, $i, j \in \{1, \ldots, C\}$. We note that $\sum_{i=1}^{C}[\boldsymbol{D}(\boldsymbol{x}_{n-1,k})]_{i,j} = 1$.

B. Measurement and Classifier Model

There are S sensors indexed by $s \in \{1, \ldots, S\}$. Each sensor s provides, at time n, $M_n^{(s)}$ measurements $\boldsymbol{q}_{n,m}^{(s)}$, $m \in \mathcal{M}_n^{(s)} \triangleq \{1, \ldots, M_n^{(s)}\}$. An existing PT k (i.e., with $r_{n,k} = 1$) is detected by sensor s—in the sense that it generates a measurement at sensor s—with probability $P_{\mathrm{d}}^{(s)}(\boldsymbol{x}_{n,k}, \ell_{n,k})$. A measurement originating from PT k follows the measurement model $\boldsymbol{q}_{n,m}^{(s)} = \boldsymbol{\psi}_s(\boldsymbol{x}_{n,k}, \boldsymbol{v}_{n,m}^{(s)})$. Here, $\boldsymbol{v}_{n,m}^{(s)}$ is measurement noise that is iid across n and m and independent across s. The function $\boldsymbol{\psi}_s(\cdot,\cdot)$ and the statistics of $\boldsymbol{v}_{n,m}^{(s)}$ define the likelihood function $f(\boldsymbol{q}_{n,m}^{(s)}|\boldsymbol{x}_{n,k})$. A "false alarm" (clutter) measurement is distributed according to pdf $f_0(\boldsymbol{q}_{n,m}^{(s)})$. The number of false alarms at sensor s is Poisson distributed with mean $\mu^{(s)}$.

Each measurement $\boldsymbol{q}_{n,m}^{(s)}$ is accompanied by an estimate $\zeta_{n,m}^{(s)}$ of the class index, which is provided by a classifier. Here, $\zeta_{n,m}^{(s)} = 0$ expresses the classifier's belief that measurement $\boldsymbol{q}_{n,m}^{(s)}$ is clutter-generated, and $\zeta_{n,m}^{(s)} = c \in \{1, \ldots, C\}$ that it is generated by a target that belongs to class c. For convenience, we define the *augmented measurement* vector $\boldsymbol{z}_{n,m}^{(s)} \triangleq [\boldsymbol{q}_{n,m}^{(s)\mathrm{T}}, \zeta_{n,m}^{(s)}]^{\mathrm{T}}$ as well as the stacked vectors $\boldsymbol{z}_n^{(s)} \triangleq [\boldsymbol{z}_{n,1}^{(s)\mathrm{T}}, \ldots, \boldsymbol{z}_{n,M_n^{(s)}}^{(s)\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{z}_n \triangleq [\boldsymbol{z}_n^{(1)\mathrm{T}}, \ldots, \boldsymbol{z}_n^{(S)\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{z} \triangleq [\boldsymbol{z}_1^{\mathrm{T}}, \ldots, \boldsymbol{z}_n^{\mathrm{T}}]^{\mathrm{T}}$.

The statistical dependency of a target-generated augmented measurement $\boldsymbol{z}_{n,m}^{(s)}$ on the underlying PT state $\boldsymbol{x}_{n,k}$ and PT class $\ell_{n,k}$ is described by the likelihood function $f(\boldsymbol{z}_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k})$. Assuming that $\zeta_{n,m}^{(s)}$ is conditionally independent of $\boldsymbol{q}_{n,m}^{(s)}$ given $\boldsymbol{x}_{n,k}$ and $\ell_{n,k}$, we obtain

$$\begin{aligned} f(\boldsymbol{z}_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) &= f(\boldsymbol{q}_{n,m}^{(s)}, \zeta_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) \\ &= p(\zeta_{n,m}^{(s)}|\boldsymbol{q}_{n,m}^{(s)}, \boldsymbol{x}_{n,k}, \ell_{n,k}) f(\boldsymbol{q}_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) \\ &= p(\zeta_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) f(\boldsymbol{q}_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) \\ &= p(\zeta_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k}) f(\boldsymbol{q}_{n,m}^{(s)}|\boldsymbol{x}_{n,k}). \end{aligned} \tag{2}$$

Here, in the last step, we used that the measurement model $\boldsymbol{\psi}_s(\cdot,\cdot)$ does not depend on $\ell_{n,k}$. The pmf $p(\zeta_{n,m}^{(s)}|\boldsymbol{x}_{n,k}, \ell_{n,k})$ models the performance of the classifier; it is described by a confusion matrix $\boldsymbol{G}^{(s)}(\boldsymbol{x}_{n,k}) \in [0,1]^{(C+1)\times C}$. Here, $[\boldsymbol{G}^{(s)}(\boldsymbol{x}_{n,k})]_{i,j} = p(\zeta_{n,m}^{(s)} = i|\boldsymbol{x}_{n,k}, \ell_{n,k} = j)$ is the probability that the classifier output is $i \in \{0, \ldots, C\}$ when the measurement is generated by a target belonging to class $j \in \{1, \ldots, C\}$. We note that $\sum_{i=0}^{C}[\boldsymbol{G}^{(s)}(\boldsymbol{x}_{n,k})]_{i,j} = 1$.

The pdf of a false alarm (clutter-generated) augmented measurement is given by $f_{\mathrm{FA}}(\boldsymbol{z}_{n,m}^{(s)}) = p_0(\zeta_{n,m}^{(s)}|\boldsymbol{q}_{n,m}^{(s)}) f_0(\boldsymbol{q}_{n,m}^{(s)})$, where $p_0(\zeta_{n,m}^{(s)} = i|\boldsymbol{q}_{n,m}^{(s)})$ is the probability conditioned on $\boldsymbol{q}_{n,m}^{(s)}$ that the classifier output is $i \in \{0, \ldots, C\}$ when the measurement $\boldsymbol{q}_{n,m}^{(s)}$ is clutter-generated.

We note that the proposed model can be extended to classifiers providing "soft" probabilistic information. Here, instead of a class estimate $\zeta_{n,m}^{(s)}$, the classifier output is a probability vector $\boldsymbol{p}_{n,m}^{(s)} = [p_{n,m,1}^{(s)}, \ldots, p_{n,m,C}^{(s)}]^{\mathrm{T}}$ where $p_{n,m,i}^{(s)} \in [0,1]$ is the classifier's estimate of the probability that measurement m is generated by a target belonging to class $i \in \{1, \ldots, C\}$.

C. MOU Model

The association between the measurements and the existing PTs is unknown, and it is also possible that a measurement $\boldsymbol{q}_{n,m}^{(s)}$ does not originate from any existing PT (false alarm) or an existing PT does not generate any measurement (missed detection). We make the assumption—known as point target assumption—that an existing PT can generate at most one measurement at a given sensor and a measurement can originate from at most one existing PT [1]. Let us define the *PT-oriented association variable* $a_{n,k}^{(s)}$, $k \in \mathcal{K}$ to be $m \in \mathcal{M}_n^{(s)}$ if PT k generates measurement m at sensor s, and zero if PT k is missed by sensor s. Similarly, the *measurement-oriented association variable* $b_{n,m}^{(s)}$, $m \in \mathcal{M}_n^{(s)}$ is $k \in \mathcal{K}$ if measurement m at sensor s originates from PT k, and zero if it is a false alarm. Following [3], [35], we define the indicator function $\Psi_{km}^{(s)}(a_{n,k}^{(s)}, b_{n,m}^{(s)})$ to be one if the values of $a_{n,k}^{(s)}$ and $b_{n,m}^{(s)}$ are consistent, i.e., if they do not describe different PT-measurement associations, and zero otherwise. More formally, $\Psi_{km}^{(s)}(a_{n,k}^{(s)}, b_{n,m}^{(s)}) = 0$ if either $a_{n,k}^{(s)} = m$ and $b_{n,m}^{(s)} \neq k$ or $a_{n,k}^{(s)} \neq m$ and $b_{n,m}^{(s)} = k$, and $\Psi_{km}^{(s)}(a_{n,k}^{(s)}, b_{n,m}^{(s)}) = 1$ otherwise. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ comprise, respectively, all the $a_{n',k}^{(s)}$, $k \in \mathcal{K}$ and all the $b_{n',m}^{(s)}$, $m \in \mathcal{M}_n^{(s)}$ for all the sensors s and all the times $n' = 1, \ldots, n$.

A Bayesian network showing the statistical dependencies among $\boldsymbol{y}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{z}$, and $M_{n'}^{(s)}$ for all sensors $s \in \{1, \ldots, S\}$ and all times $n' = 1, \ldots, n$ is presented in the supplementary material manuscript [36].

III. The Proposed Method

MTT aims to determine if a PT $k \in \mathcal{K}$ exists and, if it exists, to estimate its state $\boldsymbol{x}_{n,k}$. This essentially amounts to calculating the posterior existence probability $p(r_{n,k}=1|\boldsymbol{z})$ and the posterior state pdf $f(\boldsymbol{x}_{n,k}|r_{n,k}=1,\boldsymbol{z})$. PT k is detected—i.e., declared to exist—if $p(r_{n,k}=1|\boldsymbol{z})$ is larger than a suitably chosen threshold P_{th} [37, Ch. 2]. Then, an estimate of $\boldsymbol{x}_{n,k}$ is given by $\hat{\boldsymbol{x}}_{n,k} \triangleq \int \boldsymbol{x}_{n,k} f(\boldsymbol{x}_{n,k}|r_{n,k}=1,\boldsymbol{z}) \mathrm{d}\boldsymbol{x}_{n,k}$ [37, Ch. 4]. The statistics $p(r_{n,k}=1|\boldsymbol{z})$ and $f(\boldsymbol{x}_{n,k}|r_{n,k}=1,\boldsymbol{z})$ can be obtained from the posterior pdf $f(\boldsymbol{x}_{n,k},r_{n,k},\ell_{n,k}|\boldsymbol{z}) = f(\boldsymbol{y}_{n,k}|\boldsymbol{z})$ essentially via marginalization. Thus, it remains to calculate the posterior pdfs $f(\boldsymbol{y}_{n,k}|\boldsymbol{z})$ for all $k \in \mathcal{K}$.

The posterior pdf $f(\boldsymbol{y}_{n,k}|\boldsymbol{z})$ is a marginal density of the joint posterior pdf $f(\boldsymbol{y},\boldsymbol{a},\boldsymbol{b}|\boldsymbol{z})$. With the assumptions made in Section II and those made in [4] (also stated in the supplementary material manuscript [36]), one can show that

$$f(\boldsymbol{y},\boldsymbol{a},\boldsymbol{b}|\boldsymbol{z}) \propto \prod_{k=1}^{K} f(\boldsymbol{y}_{0,k}) \prod_{n'=1}^{n} f(\boldsymbol{y}_{n',k}|\boldsymbol{y}_{n'-1,k}) \times \prod_{s=1}^{S} \upsilon^{(s)}\big(\boldsymbol{x}_{n',k}, r_{n',k}, \ell_{n',k}, a_{n',k}^{(s)}; \boldsymbol{z}_{n'}^{(s)}\big) \prod_{m=1}^{M_{n'}^{(s)}} \Psi\big(a_{n',k}^{(s)}, b_{n',m}^{(s)}\big). \tag{3}$$

Here, $f(\boldsymbol{y}_{n,k}|\boldsymbol{y}_{n-1,k})$ is given by (1), and $f(\boldsymbol{y}_{0,k})$ is the prior pdf of the augmented state of PT k at time $n=0$, which generally includes also some prior probabilistic information on the PT class. Furthermore, $\upsilon^{(s)}\big(\boldsymbol{x}_{n,k}, r_{n,k}, \ell_{n,k}, a_{n,k}^{(s)}; \boldsymbol{z}_n^{(s)}\big)$ is given for $r_{n,k}=1$ by $P_{\mathrm{d}}^{(s)}(\boldsymbol{x}_{n,k},\ell_{n,k}) f\big(\boldsymbol{z}_{n,m}^{(s)}|\boldsymbol{x}_{n,k},\ell_{n,k}\big) / \big(\mu^{(s)} f_{\mathrm{FA}}\big(\boldsymbol{z}_{n,m}^{(s)}\big)\big)$ if $a_{n,k}^{(s)}=m \in \mathcal{M}_n^{(s)}$ and $1-P_{\mathrm{d}}^{(s)}(\boldsymbol{x}_{n,k},\ell_{n,k})$ if $a_{n,k}^{(s)}=0$, and for $r_{n,k}=0$ by $\delta_{a_{n,k}^{(s)},0}$, where $f\big(\boldsymbol{z}_{n,m}^{(s)}|\boldsymbol{x}_{n,k},\ell_{n,k}\big)$ is given by (2) and $\delta_{a_{n,k}^{(s)},0}$ is one if $a_{n,k}^{(s)}=0$ and zero otherwise. A detailed derivation of (3) is provided in [4]; the resulting factorization formally differs from (3) only by the definitions of the transition pdf $f(\boldsymbol{y}_{n,k}|\boldsymbol{y}_{n-1,k})$ and the function $\upsilon^{(s)}\big(\boldsymbol{x}_{n,k}, r_{n,k}, \ell_{n,k}, a_{n,k}^{(s)}; \boldsymbol{z}_n^{(s)}\big)$. A sketch of that derivation is included in the supplementary material manuscript [36]. If some of the sensors are not accompanied by a classifier providing estimates of the target class, then for these sensors the definition of the function $\upsilon^{(s)}(\cdot)$ from [4] has to be used.

The factor graph [6]–[8] describing the factorization (3) is shown for one time step in Fig. 1. Following the approach in [4], approximations of the marginal posterior pdfs $f(\boldsymbol{y}_{n,k}|\boldsymbol{z}) = f(\boldsymbol{x}_{n,k},r_{n,k},\ell_{n,k}|\boldsymbol{z})$, known as *beliefs* and denoted as $\tilde{f}(\boldsymbol{y}_{n,k}) = \tilde{f}(\boldsymbol{x}_{n,k},r_{n,k},\ell_{n,k})$, can be calculated efficiently by running iterative SPA message passing on this factor graph. Since the factor graph contains loops, there is no unique order of calculating the messages, and different orders may result in different beliefs. We define the order by the following rules: first, messages are not sent backward in time, and second, *iterative* message passing is only performed for probabilistic data association, and separately at each time step and at each sensor. The second rule implies that for loops involving different sensors, only a single message passing iteration is performed. The structure of the factor graph in Fig. 1 equals that in [4], even though the underlying Bayesian model—i.e., the functions represented by the factor nodes "f_k" and "υ_k" and the variables represented by the variable nodes "$\boldsymbol{y}_k$"—are different. The derivation and expressions of the messages are thus analogous to [4] and are omitted because of space restrictions. However, a detailed statement of our method is provided in the supplementary material manuscript [36].

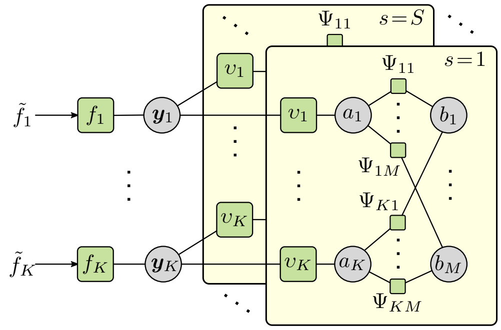

Fig. 1. Factor graph describing the factorization of $f(\boldsymbol{y},\boldsymbol{a},\boldsymbol{b}|\boldsymbol{z})$ in (3) for one time step n. For simplicity, the sensor index s and the time index n are omitted, and the following short notations are used: $\tilde{f}_k \triangleq \tilde{f}(\boldsymbol{y}_{n,k})$, $f_k \triangleq f(\boldsymbol{y}_{n,k}|\boldsymbol{y}_{n-1,k})$, $\upsilon_k \triangleq \upsilon^{(s)}\big(\boldsymbol{y}_{n,k}, a_{n,k}^{(s)}; \boldsymbol{z}_n^{(s)}\big)$, $a_k \triangleq a_{n,k}^{(s)}$, $b_m \triangleq b_{n,m}^{(s)}$, $\Psi_{km} \triangleq \Psi_{km}\big(a_{n,k}^{(s)}, b_{m,n}^{(s)}\big)$, and $M \triangleq M_n^{(s)}$.

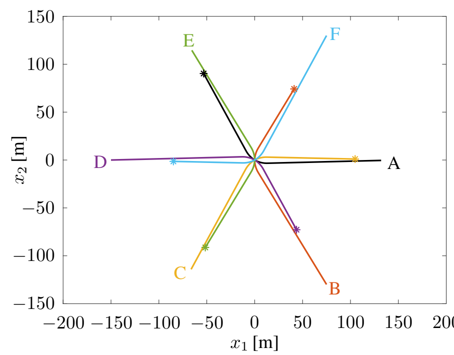

Fig. 2. Trajectories of the six targets. The stars mark the final target positions.

IV. Numerical Study

A. Simulation Setup

We simulated six targets "A" through "F" that move in a rectangular region of interest (ROI) during 140 time steps with a constant speed of 1 m/s. The target trajectories and the ROI are shown in Fig. 2. The targets move toward the ROI center starting from positions uniformly placed on a circle of radius 150 m, and then, approximately at time $n=75$, they perform a right turn of 60 degrees. Targets A, C, and E start and stop to exist at times $n=10$ and $n=130$, respectively, and targets B, D, and F at times $n=1$ and $n=120$, respectively. There are $C=3$ target classes; targets A and D belong to class $c=1$, targets B and E to class $c=2$, and targets C and F to class $c=3$. However, the tracking method has no prior knowledge about these class affiliations.

The PT states are $\boldsymbol{x}_{n,k} = [\check{\boldsymbol{x}}_{n,k}^{\mathrm{T}}, \dot{\check{\boldsymbol{x}}}_{n,k}^{\mathrm{T}}]^{\mathrm{T}}$ with two-dimensional (2D) position $\check{\boldsymbol{x}}_{n,k}$ and 2D velocity $\dot{\check{\boldsymbol{x}}}_{n,k}$. The dynamic model used by the tracking methods is a nearly constant velocity model, i.e., $\boldsymbol{x}_{n,k} = \boldsymbol{\theta}(\boldsymbol{x}_{n-1,k},\boldsymbol{u}_{n,k}) = \boldsymbol{A}\boldsymbol{x}_{n-1,k} + \boldsymbol{W}\boldsymbol{u}_{n,k}$, where $\boldsymbol{A} \in \mathbb{R}^{4\times4}$ and $\boldsymbol{W} \in \mathbb{R}^{4\times2}$ are as in [38, Sec. 6.3.2] (with a time step duration of 2 s) and $\boldsymbol{u}_{n,k}$ is iid zero-mean Gaussian with a per-component standard deviation of 0.1 m/s^2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3024858, IEEE Signal Processing Letters

Note that the PTs' dynamic model does not depend on the PT class, but the PT class may be related to some other target characteristic(s) such as color, size, or type. In the tracking methods, the number of PTs is chosen as $K = 20$. The elements of the class transition matrix $\boldsymbol{D}(\boldsymbol{x}_{n-1,k}) = \boldsymbol{D}$ are chosen as $[\boldsymbol{D}]_{i,j} = 0.95$ if $i = j$ and $[\boldsymbol{D}]_{i,j} = 0.025$ if $i \neq j$.

There are $S = 1$ or 2 sensors equally spaced on a circle of radius 3 km around $(0, 0)$. The sensors measure range and bearing. The target-generated measurements are thus modeled as $\boldsymbol{q}^{(s)}_{n,m} = \boldsymbol{\psi}_s(\boldsymbol{x}_{n,k}, \boldsymbol{v}^{(s)}_{n,m}) = \big[\|\tilde{\boldsymbol{x}}_{n,k} - \boldsymbol{p}^{(s)}\|, \phi(\tilde{\boldsymbol{x}}_{n,k}, \boldsymbol{p}^{(s)})\big]^{\mathrm{T}} + \boldsymbol{v}^{(s)}_{n,m}$, where $\boldsymbol{p}^{(s)}$ is the position of sensor s, $\phi(\tilde{\boldsymbol{x}}_{n,k}, \boldsymbol{p}^{(s)})$ is the angle of the vector $\tilde{\boldsymbol{x}}_{n,k} - \boldsymbol{p}^{(s)}$, and $\boldsymbol{v}^{(s)}_{n,m}$ is iid—also across s—zero-mean Gaussian with covariance matrix $\mathrm{diag}(\sigma_{\mathrm{r}}^2, \sigma_{\mathrm{b}}^2)$, where $\sigma_{\mathrm{r}} = 5$ m and $\sigma_{\mathrm{b}} = 0.1°$. The false alarm pdf $f_0(\boldsymbol{q}^{(s)}_{n,m})$ is linearly increasing with respect to range and uniform with respect to bearing within the ROI, and zero outside the ROI. The mean number of false alarms, $\mu^{(s)}$, is 5, 10, or 20. The probability of detection is $P^{(s)}_{\mathrm{d}}(\boldsymbol{x}_{n,k}, \ell_{n,k}) = P_{\mathrm{d}} = 0.9$. Each sensor includes a classifier whose output $\zeta^{(s)}_{n,m} \in \{0, 1, 2, 3\}$ accompanies measurement $\boldsymbol{q}^{(s)}_{n,m}$, where $\zeta^{(s)}_{n,m} = 0$ expresses the classifier's belief that $\boldsymbol{q}^{(s)}_{n,m}$ is clutter-generated and $\zeta^{(s)}_{n,m} = 1$, 2, and 3 that it is generated by a target belonging to class 1, 2, and 3, respectively. The elements of the classifiers' confusion matrix $\boldsymbol{G}^{(s)}(\boldsymbol{x}_{n,k}) = \boldsymbol{G}$ are chosen as $[\boldsymbol{G}]_{i,j} = 0.85$ if $i = j$ and $[\boldsymbol{G}]_{i,j} = 0.05$ if $i \neq j$, with $i \in \{0, 1, 2, 3\}$, $j \in \{1, 2, 3\}$. The pmf $p_0\big(\zeta^{(s)}_{n,m} \big| \boldsymbol{q}^{(s)}_{n,m}\big)$ is chosen independently of $\boldsymbol{q}^{(s)}_{n,m}$ as $p_0\big(\zeta^{(s)}_{n,m} = 0\big) = 0.85$ and $p_0\big(\zeta^{(s)}_{n,m} = i\big) = 0.05$ for $i \in \{1, 2, 3\}$.

B. Results

We compare the performance of the proposed classifier-aided method with that of the baseline method of [4], which does not use the classifier output $\zeta^{(s)}_{n,m}$. The performance is assessed in terms of the time-averaged mean generalized optimal sub-pattern assignment (MGOSPA) error [39], the time-averaged mean optimal sub-pattern assignment (MOSPA) error [40], the time-averaged MOSPA-for-tracks (MOSPA-T) error [41] (all three with order 1 and cutoff parameter 20 m; the time-averaged MOSPA-T error additionally with label error penalty 20 m), and the false alarm rate (FAR). The MGOSPA and MOSPA errors take into account estimation errors for correctly detected targets and errors due to incorrect target detections. The MOSPA-T error additionally penalizes incorrect switches of the estimated tracks; it equals the MOSPA error when there are no switches. The FAR is the number of false tracks per unit of space and unit of time.

Tables I and II report these metrics, averaged over 200 simulation runs, for $S = 1$ and $S = 2$ sensors, respectively, and for three different values of $\mu^{(s)}$. It can be seen that the proposed method consistently outperforms the baseline method; the gain in performance increases with the clutter level. In particular, for $\mu^{(s)} = 20$, the FAR is about 4.5 times lower than that obtained with the baseline method, and the time-averaged MOSPA-T error is about 53 % (for $S = 1$) and 62 % (for $S = 2$) lower. Fig. 3 displays the MOSPA-T error versus time, for $S = 1$ or 2 sensors and mean number of false alarms $\mu^{(s)} = 10$. Again, the proposed method consistently outperforms the baseline method. In particular, the MOSPA-T error of the baseline method noticeably increases right after the turn of the targets (approximately at time $n = 78$), which indicates the occurrence of switches among the estimated tracks. The MOSPA-T error of the proposed method, instead, keeps approximately the same level, which indicates that no or almost no track switches occur. Note that the peaks observed at various times are due to target appearance and disappearance.

These results demonstrate the benefit of using class information to improve the performance of the SPA-based MTT methodology. Indeed, the class information allows a more reliable association between measurements and targets, which reduces track switches and thus increases tracking accuracy. Additional results assessing the performance of the proposed method for different values of the number of classes C and different choices of the confusion matrix $\boldsymbol{G}$ are provided in the supplementary material manuscript [36].

TABLE I
TIME-AVERAGED MGOSPA, MOSPA, AND MOSPA-T ERRORS AS WELL AS FAR FOR $S = 1$ SENSOR.

$\mu^{(s)}$	MGOSPA [m]		MOSPA [m]		MOSPA-T [m]		FAR [$\mathrm{km}^{-2}\mathrm{s}^{-1}$]	
	Basel.	Prop.	Basel.	Prop.	Basel.	Prop.	Basel.	Prop.
5	22.4	20.8	5.0	4.4	10.0	4.7	0.64	0.31
10	25.8	21.9	6.0	4.8	10.7	5.2	1.38	0.53
20	39.1	24.2	8.6	5.6	12.7	6.0	4.62	1.01

TABLE II
TIME-AVERAGED MGOSPA, MOSPA, AND MOSPA-T ERRORS AS WELL AS FAR FOR $S = 2$ SENSORS.

$\mu^{(s)}$	MGOSPA [m]		MOSPA [m]		MOSPA-T [m]		FAR [$\mathrm{km}^{-2}\mathrm{s}^{-1}$]	
	Basel.	Prop.	Basel.	Prop.	Basel.	Prop.	Basel.	Prop.
5	16.7	15.7	3.5	3.3	9.0	3.5	0.37	0.21
10	19.5	16.4	4.4	3.5	9.6	3.8	1.05	0.36
20	30.0	18.4	7.1	4.2	11.6	4.5	3.69	0.83

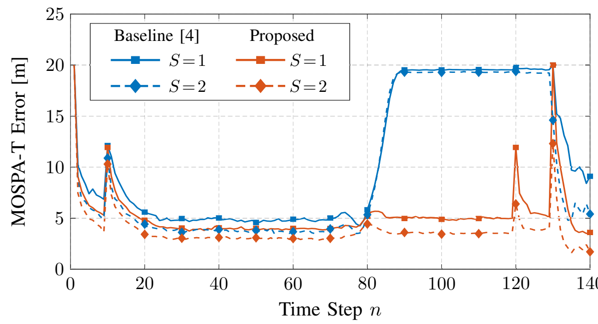

Fig. 3. MOSPA-T error versus time for $\mu^{(s)} = 10$.

V. Conclusion

A challenging issue in MTT is the unknown association between measurements and targets. The use of class information in addition to sensor measurements can improve probabilistic target-measurement association and, in turn, overall MTT performance. In this letter, we showed how the output of a classifier can be integrated into the SPA-based MTT framework recently proposed in [4], and we demonstrated experimentally significant performance advantages of the resulting classifier-aided method over the method of [4].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3024858, IEEE Signal Processing Letters

References

[1] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT, USA: YBS Publishing, 2011.

[2] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.

[3] J. L. Williams and R. A. Lau, "Approximate evaluation of marginal association probabilities with belief propagation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2942–2959, Oct. 2014.

[4] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, "A scalable algorithm for tracking an unknown number of targets using multiple sensors," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3478–3493, Jul. 2017.

[5] F. Meyer, T. Kropfreiter, J. L. Williams, R. A. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, Feb. 2018.

[6] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[7] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[8] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[9] D. Gaglione, P. Braca, and G. Soldi, "Belief propagation based AIS/radar data fusion for multi-target tracking," in *Proc. FUSION-18*, Cambridge, UK, Jul. 2018, pp. 2143–2150.

[10] G. Soldi, D. Gaglione, F. Meyer, F. Hlawatsch, P. Braca, A. Farina, and M. Z. Win, "Heterogeneous information fusion for multitarget tracking using the sum-product algorithm," in *Proc. IEEE ICASSP-19*, Brighton, UK, May 2019, pp. 5471–5475.

[11] G. Soldi, F. Meyer, P. Braca, and F. Hlawatsch, "Self-tuning algorithms for multisensor-multitarget tracking using belief propagation," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3922–3937, Aug. 2019.

[12] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: Direct derivation and implementation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1883–1901, Aug. 2018.

[13] Y. Xia, K. Granström, L. Svensson, and Á. F. García-Fernández, "Performance evaluation of multi-Bernoulli conjugate priors for multi-target filtering," in *Proc. FUSION-17*, Xi'an, China, Jul. 2017, pp. 1–8.

[14] K. Granström, L. Svensson, Y. Xia, J. L. Williams, and Á. F. García-Fernández, "Poisson multi-Bernoulli mixture trackers: continuity through random finite sets of trajectories," in *Proc. FUSION-18*, Cambridge, UK, Jul. 2018, pp. 1–5.

[15] J. Vermaak, S. J. Godsill, and P. Perez, "Monte Carlo filtering for multi target tracking and data association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 309–332, Jan. 2005.

[16] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.

[17] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3553–3567, Jul. 2007.

[18] G. Battistelli, L. Chisci, S. Morrocchi, F. Papi, A. Farina, and A. Graziano, "Robust multisensor multitarget tracker with application to passive multistatic radar tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3450–3472, Oct. 2012.

[19] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the Bayes multi-target tracking filter," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6554–6567, Dec. 2014.

[20] J. L. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA and association-based MeMBer," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1664–1687, Jul. 2015.

[21] S. Nannuru, S. Blouin, M. Coates, and M. Rabbat, "Multisensor CPHD filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 4, pp. 1834–1854, Aug. 2016.

[22] P. Sharma, A. Saucan, D. J. Bucci, and P. K. Varshney, "Decentralized Gaussian filters for cooperative self-localization and multi-target tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5896–5911, Nov. 2019.

[23] T. Li, X. Wang, Y. Liang, and Q. Pan, "On arithmetic average fusion and its application for distributed multi-Bernoulli multitarget tracking," *IEEE Trans. Signal Process.*, vol. 68, pp. 2883–2896, Apr. 2020.

[24] O. E. Drummond, "Feature, attribute, and classification aided target tracking," in *Proc. SPIE-01*, vol. 4473, San Diego, CA, USA, Jul. 2001, pp. 542–558.

[25] Y. Bar-Shalom, T. Kirubarajan, and C. Gokberk, "Tracking with classification-aided multiframe data association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 868–878, Jul. 2005.

[26] V. C. Ravindra, Y. Bar-Shalom, and T. Damarla, "Feature-aided localization of ground vehicles using passive acoustic sensor arrays," in *Proc. FUSION-09*, Seattle, WA, USA, Jul. 2009, pp. 70–77.

[27] S. Singh, H. Tu, W. Donat, K. Pattipati, and P. Willett, "Anomaly detection via feature-aided tracking and hidden Markov models," *IEEE Trans. Syst., Man, Cybern. A*, vol. 39, no. 1, pp. 144–159, Jan. 2009.

[28] C. Ying, C. Zhen, and W. Shuliang, "Feature aided Gaussian mixture probability hypothesis density filter with modified 2D assignment," in *Proc. IEEE CIE RADAR-11*, vol. 1, Chengdu, China, Oct. 2011, pp. 800–803.

[29] R. Georgescu and P. Willett, "The GMCPHD tracker applied to the Clutter09 dataset," in *Proc. FUSION-13*, Istanbul, Turkey, Jul. 2013, pp. 530–537.

[30] G. R. Mellema, "Feature-aided tracking in dense clutter using the Clutter09 data set," in *Proc. FUSION-14*, Salamanca, Spain, Jul. 2014, pp. 1–8.

[31] S. Mori, K. Chang, and C. Chong, "Performance prediction of feature-aided track-to-track association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2593–2603, Oct. 2014.

[32] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[33] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[35] M. Bayati, D. Shah, and M. Sharma, "Max-product for maximum weight matching: Convergence, correctness, and LP duality," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1241–1251, Mar. 2008.

[36] D. Gaglione, G. Soldi, P. Braca, G. De Magistris, F. Meyer, and F. Hlawatsch, "Classification-aided multitarget tracking using the sum-product algorithm — Supplementary material." [Online]. Available: https://arxiv.org/pdf/2008.01667.pdf

[37] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer, 1994.

[38] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001.

[39] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proc. FUSION-17*, Xi'an, China, Jul. 2017, pp. 1–8.

[40] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.

[41] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3452–3457, Jul. 2011.