14th International Conference on Information Fusion
Chicago, Illinois, USA, July 5-8, 2011

Measurement-to-Track Association for Nontraditional Measurements

Ronald Mahler
Unified Data Fusion Sciences, Inc.
Eagan, MN, 55122, USA.
USDFRPSM@gmail.com

Abstract – Data fusion algorithms must typically address not only kinematic issues—that is, target tracking—but also nonkinematics—for example, target identification, threat estimation, intent assessment, etc. Whereas kinematics involves traditional measurements such as radar detections, nonkinematics typically involves non-traditional measurements such as quantized data, attributes, features, natural-language statements, and inference rules. The kinematic vs. nonkinematic chasm is often bridged by grafting some expert-system approach (fuzzy logic, Dempster-Shafer, rule-based inference) into a single- or multi-hypothesis multitarget tracking algorithm, using ad hoc methods. The purpose of this paper is to show that conventional measurement-to-track association theory can be **directly** ***extended to nontraditional measurements in a Bayesian manner. Concepts such as association likelihood, association distance, hypothesis probability, and global nearest-neighbor distance are defined, and explicit formulas are derived for specific kinds of nontraditional evidence.***

Keywords: Data association, measurement-to-track association, non-traditional measurements, random sets, generalized likelihood function.

1 Introduction

Recent years have seen the emergence of multitarget detection and tracking algorithms that avoid explicit measurement-to-track association (MTA). These algorithms include a complete Bayesian formulation of the problem: the general multitarget Bayes filter [3, Chapter 14], along with its approximations, the PHD, CPHD, and multi-Bernoulli filters [3, Chapters 16,17]. Using the techniques of Chapters 3-6 of [3], these filters also provide a *complete Bayesian formulation for processing nontraditional measurement types* such as quantized data, attributes, features, natural-language statements, and even inference rules (see Section 14.4.2 of [3]).

This complete Bayesian formulation is not, however, the subject of this paper. The dominant multitarget tracking methodology—for both legacy algorithms and for most ongoing data fusion algorithm R&D—is not the MTA-free paradigm just mentioned, but rather MTA itself. Thus one might ask: Is it possible to adapt the complete Bayesian formulation so that it, or at least its major aspects, can be incorporated into legacy MTA-based algorithms? I claim that this question can be answered in the affirmative, but to answer it one must address a major gap in MTA theory.

Although MTA has been applied to a range of information fusion problems, its theoretical foundations (see Section 3 and Chapter 10 of [3]) remain those of multitarget tracking and, specifically, those associated with a particular kind of input data: kinematic measurements. MTA arose as a way of extending the purview of the Kalman filter to the multitarget realm.

In MTA, a set of measurement-vectors—positions or bearing angles, for example—is collected. Then the following question is posed: Which target tracks generated which of these measurements, and which measurements cannot be attributed to any track? Various techniques—gating, nearest-neighbor, global nearest-neighbor, etc.—are used to determine the best possible association. In this case MTA leads to what is known as a single-hypothesis tracker (SHT).

SHTs tend to exhibit non-robust performance. Thus, more generally, one can propagate a table of suboptimal associations, along with their probabilities of being true. In this case MTA leads to the current workhorse of practical multitarget detection and tracking, the multi-hypothesis tracker (MHT).

All of these approaches depend on the ability to define what it means for a measurement to be "near" a track—and thus to have been plausibly generated by that track. That is, they all depend on the ability to define "distance" between measurements and tracks.

The theoretical foundations of MTA are based on the presumption that collected measurements, like tracks, are kinematic entities that can be easily represented in Gaussian terms—i.e., in terms of vectors and covariance matrices. In this case, it is easy to define the concept of an "association distance." Most commonly, this takes the form of a Mahalanobis distance and its multitarget generalizations.

But what if the measurements are not kinematic—they pertain, for example, to target identity? Typical examples of such "nontraditional" measurements are attributes extracted by human operators, features extracted by digital signal processing (DSP) algorithms, natural-language statements, and inference rules. What does "distance between tracks and measurements" mean in this case?

The most typical approaches to this problem have consisted of bottom-up, *ad hoc* integrations of familiar expert systems methods—fuzzy logic, Dempster-Shafer theory, etc. into MTA-based algorithms such as MHT. (See [4] for summaries of some of these techniques.) It remains the case, however, that such methods remain controversial, especially among Bayesians, who often describe them as inherently "heuristic" or worse.

The purpose of this paper is to *extend standard kinematic MTA theory to nontraditional measurements in a systematic, top-down, theoretically disciplined, and as-Bayesian-as-possible manner*. I am not aware of any previous work that has even attempted to do this—which consequentially means that I know of no meaningful precedents in the literature. (If readers know of such prededents, I would be happy to learn of them.)

My approach is to review traditional kinematic MTA theory (Section 3), summarize my Bayesian theory of generalized measurements and generalized likelihood functions (Section 4), and then show how to integrate the two (Sections 5, 6). Specifically, *the systematic treatment of conventional MTA in Section 3 is, in Section 5, repeated essentially verbatim*—except that generalized likelihoods (for nontraditional measurements) are substituted in place of conventional likelihoods (for conventional measurements). What results is *an conceptually parsimonious and straightforward generalization of conventional MTA theory to nontraditional measurements.*

Thus I will define concepts such as association likelihood, association distance, hypothesis probability, and global nearest-neighbor distance for non-traditional measurements. I will derive explicit formulas for specific kinds of nontraditional measurements. Special attention will be devoted to the case when probability of detection is unity, since relatively simple closed-form formulas can be derived under this assumption. The main results of the paper are the formulas for association likelihood and hypothesis probability, Eqs. (49,54), and the explicit closed-form formulas for generalized likelihoods for fuzzy and fuzzy Dempster-Shafer measurements in Section 6.[1]

But it is necessary to also point out what this paper does *not* do. First, *my viewpoint is unreservedly Bayesian*. As such, this is not the proper venue for a survey or assessment of proposed applications of Dempster-Shafer, fuzzy logic, DSmT, etc., to multitarget tracking. Second, *I am not proposing a new formulation of MTA*. Rather, I am proposing an extension of existing MTA theory to nontraditional data. Third, my purpose is to *report a new theoretical advance, not to engage in tutorial expositions of existing results*. The Bayesian theory of non-traditional measurements developed in [3] is summarized in Section 4. I present brief examples of the process of representing nontraditional measurements as random sets and then specifying their likelihood functions (e.g., Eqs. (30,31)). But for more concrete examples, the reader should consult Chapter 3 of [3]. Fourth, my purpose is to *propose an inherently mathematically complex theoretical advance*. While implementation and simulation are appropriate subjects for future research, the reader will discover that the entire page limit is required just to systematically elucidate the theoretical approach.[2] Fifth and finally, *the purpose of this paper is not to provide an overview*. The approach requires systematic adherence to a mathematically precise Bayesian methodology—with the aim of deriving explicit closed-form formulas that can be used in practice for different kinds of non-traditional data. As such, the exposition *is no more and no less mathematical than it must be* if valid results are to be reported in sufficient detail for the engaged (as opposed to cursory) reader.

The paper is organized as follows. It expands upon concepts briefly introduced in Section 10.8, pp. 341-342, of [3]. Single- and multi-hypothesis trackers are briefly reviewed in Section 2. The basic theory of measurement-to-track association is reviewed in Section 3. Generalized measurements and generalized likelihood functions are briefly reviewed in Section 4. Section 5 describes the extension of the material in Section 3 to such measurements. In Section 6, this theory is applied to arrive at closed-form formulas under certain assumptions. Section 7 describes how to extend it to joint kinematic and nonkinematic processing. Conclusions are in Section 8.

2 Single- and multi-hypothesis trackers

Probably the two most common multitarget tracking algorithms in practical application are the *single-hypothesis tracker* (SHT) and the *multi-hypothesis tracker* (MHT). A MHT recursively propagates two items through time: a track table $\mathcal{T}_{k|k}$ and a hypothesis table $\mathcal{H}_{k|k}$, using a time-update followed by a measurement-update:

$$\ldots \to \mathcal{T}_{k|k}, \mathcal{H}_{k|k} \to \mathcal{T}_{k+1|k}, \mathcal{H}_{k+1|k} \to \mathcal{T}_{k+1|k+1}, \mathcal{H}_{k+1|k+1} \to \ldots$$

Each track τ is a model of a possible real-world target, based on accumulated evidence. It has the form $\tau = (\mathbf{x}, P)$ where $\mathbf{x}$ is the estimated target state and P the associated error-covariance matrix. Each hypothesis θ in

[1] While the result of Eq. (87-89) might seem like the main result of the paper, in actuality it is a near-triviality. It would have little value without the existence of the actual main results.

[2] I might add: The founding IEEE PHD and CPHD filter papers [5,6] contained no implementations or simulations. Despite these heresies, progress in information fusion does not appear to have suffered.

$\mathcal{H}_{k|k}$ is an association—i.e., an assumption about how any given target generated any given measurement—if indeed a particular target generated any measurement, or if a particular measurement was generated by any target. It is assumed that

$$\sum_{\theta} p_{k|k}(\theta) = 1 \cdot \quad (1)$$

In terms of the most typical implementations seen in the literature, an MHT consists of a bank of extended Kalman filters (EKFs) or unscented Kalman filters (UKFs). Given any particular association θ, the measurements are used to data-update the predicted tracks with which they are associated. Thus each association corresponds to a particular set of updated tracks—that is, to some subset of the track table. In this sense, each θ is a model of what "ground truth" looks like, and $p_{k+1|k+1}(\theta)$ is the probability that this model accurately represents "ground truth."

It is possible to avoid the multi-hypothesis structure by propagating only the highest-probability hypothesis rather than the entire hypothesis table. This is accomplished by using a "global association distance" to determine the optimal measurement-to-track association at any given time-step. Association distances are defined from the formulas for hypothesis probabilities. The result, various kinds of SHTs, are computationally more tractable but also typically exhibit poorer tracking performance.

The basic theory is introduced in Section 10.3.1 of [3] and is reviewed in Section 3.

3 Measurement-to-track association (review)

This section draws heavily from Section 10.3.1 of [3]. Suppose that, at time-step k+1,

$$X = \{(\mathbf{x}_1,P_1),\ldots,(\mathbf{x}_n,P_n)\} \quad (2)$$

is the set of tracks predicted from the previous time-step, where $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are their states and $P_1,\ldots,P_n$ are their error-covariance matrices. The track density corresponding to $(\mathbf{x}_i,P_i)$ is the Gaussian distribution

$$f(\mathbf{x} \mid i) = N_{P_i}(\mathbf{x} - \mathbf{x}_i) \cdot \quad (3)$$

Also, let the sensor likelihood function at time k+1 be

$$f_i(\mathbf{z} \mid \mathbf{x}) = N_R(\mathbf{z} - H\mathbf{x}_i) \quad (4)$$

and let the measurement-set collected at time-step k+1 be

$$Z = \{\mathbf{z}_1,\ldots,\mathbf{z}_m\} \quad (5)$$

with $|Z| = m$.

3.1 Definition of an association

A *measurement-to-track association* is a function

$$\theta : \{1,\ldots,n\} \to \{0,1,\ldots,m\} \quad (6)$$

which has the following property: $\theta(i) = \theta(i') > 0$ implies that $i = i'$. This function has the following intuitive interpretation:

(a) for every $i = 1,\ldots,n$, if $\theta(i) > 0$ then the measurement $\mathbf{z}_{\theta(i)}$ is uniquely associated with the track $\mathbf{x}_i$;

(b) if $\theta(i) = 0$ then no measurement is associated with $\mathbf{x}_i$ (i.e., the track was not detected); and

(c) in either case, those measurements in Z which are not equal to any $\mathbf{z}_{\theta(i)}$ with $\theta(i) > 0$ are false detections.

At one extreme, if $\theta(i) = 0$ for all i then no target generated a measurement and so every measurement in Z must be a false detection. In this case, abbreviate $\theta = 0$ and adopt the convention $d_{\theta} = 0$. At the other extreme, if $\theta(i) > 0$ for all i then every track generated a measurement.

3.2 Local association distance

Suppose for the moment that a single kinematic measurement $\mathbf{z}$ has been collected. We are to estimate which of the predicted tracks $(\mathbf{x}_1,P_1),\ldots,(\mathbf{x}_n,P_n)$ most likely generated it, assuming that it was not a false detection. The total likelihood that $\mathbf{z}$ was generated by $(\mathbf{x}_i,P_i)$—i.e., the *association likelihood*—is

$$\ell(\mathbf{z} \mid i) = \int f(\mathbf{z} \mid \mathbf{x}) \cdot f(\mathbf{x} \mid i) d\mathbf{x} \quad (7)$$

$$= \int N_R(\mathbf{z} - H\mathbf{x}) \cdot N_{P_i}(\mathbf{x} - \mathbf{x}_i) d\mathbf{x} \quad (8)$$

$$= N_{R+HP_iH^T}(\mathbf{z} - H\mathbf{x}_i) \quad (9)$$

$$= \frac{1}{\sqrt{\det 2\pi(R + HP_iH^T)}} \cdot \exp\left(-\tfrac{1}{2} d(\mathbf{z} \mid \mathbf{x}_i)^2\right) \quad (10)$$

where

$$d(\mathbf{z}|\mathbf{x}_i)^2 = (\mathbf{z}-H\mathbf{x}_i)^T(R+HP_iH^T)^{-1}(\mathbf{z}-H\mathbf{x}_i) . \quad (11)$$

3.3 Global association distance

Suppose that an entire measurement-set $Z = \{\mathbf{z}_1,\ldots,\mathbf{z}_m\}$ with $|Z| = m$ has been collected. Assume that the false alarm process is Poisson with clutter rate λ and spatial distribution $c(\mathbf{z})$, where I am suppressing the time-index k for notational clarity.

Case 1: $p_D = 1$. Begin by assuming that probability of detection is unity, in which case an association is a one-to-one function

$$\chi : \{1,\ldots,n\} \to \{1,\ldots,m\}. \quad (12)$$

Let $W_{\chi} \subseteq Z$ be the subset of detections—that is, those $\mathbf{z}$ that equal $\mathbf{z}_{\chi(i)}$ for some $i = 1,\ldots,n$.

Given this, the *global association likelihood*—that is, the likelihood that χ matches the measurements with the tracks, taking false alarms into account—is, from Eqs. (10.58) and (10.94) of [3],

$$\ell_{\chi}(Z \mid X) = c(\chi) \cdot \prod_{i=1}^{n} \ell(\mathbf{z}_{\chi(i)} \mid i) \quad (13)$$

where

$$c(\chi) = e^{-\lambda} \prod_{\mathbf{z} \in Z - W_{\chi}} \lambda c(\mathbf{z}) \cdot \quad (14)$$

The value of the global association likelihood tends to be larger for those χ which more plausibly associate measurements to predicted tracks.

Let $p_{\chi,0}$ be the prior probability of the association χ. Since there is no *a priori* reason to prefer one association over another, $p_{\chi,0}$ is a uniform distribution.

Thus the posterior probability that the association χ is the correct one is

$$p_\chi = \frac{\ell_\chi(Z \mid X)}{\sum_{\theta'} \ell_{\chi'}(Z \mid X)} \propto \ell_\chi(Z \mid X) \cdot \quad (15)$$

The optimal association is given by the MAP estimate

$$\hat{\chi} = \arg\max_\theta p_\chi = \arg\max_\chi \ell_\chi(Z \mid X) \cdot \quad (16)$$

The global association distance for an association χ is defined from Eq. (15) by

$$d_\chi^2 = -2\log \ell_\chi(Z \mid X) \cdot \quad (17)$$

Thus the optimal association is the one that minimizes this association distance.

From Eq. (10.91) of [3], we find that the specific formula for Eq. (15) is

$$p_\chi \propto c(\chi) \cdot \exp(-\tfrac{1}{2} d_\chi(Z \mid X)^2) \quad (18)$$

where

$$d_\chi(Z \mid X)^2 = \sum_{i=1}^{n} (\mathbf{z}_{\chi(i)} - H\mathbf{x}_i)^T (R + HP_i H^T)^{-1} (\mathbf{z}_{\chi(i)} - H\mathbf{x}_i) \quad (19)$$

Assume that $c(\mathbf{z}) = c$ is uniform over some region of space. Then the quantity $c(\chi) = e^{-\lambda} (\lambda c)^{m-n}$ is no longer functionally dependent on χ and thus

$$p_\chi \propto \exp(-\tfrac{1}{2} d_\chi(Z \mid X)^2) \cdot \quad (20)$$

Thus the global association distance can be taken to be as in Eq. (19):

$$d_\chi = d_\chi(Z \mid X) \cdot \quad (21)$$

This is the definition most commonly employed in data association algorithms.

Case 2: $p_D < 1$. Now consider the more general case, in which the probability of detection is constant but not unity, and thus θ is an arbitrary association. Then the global association likelihood is (see Eq. (G.238) of [3])

$$\ell_\theta(Z \mid X) = (1 - p_D)^{n - n_\theta} \cdot c(\chi) \cdot \prod_{i=1}^{n} p_D \ell(\mathbf{z}_{\chi(i)} \mid i) \quad (22)$$

where n_θ denotes the number of detections associated with θ. In this case the formula for the association probability p_θ is (see Eqs. (10.115) and (10.116) of [3]):

$$p_\theta \propto K^{n_\theta} \cdot c(\theta) \cdot F_\theta \cdot \exp(-\tfrac{1}{2} d_\theta(Z \mid X)^2) \quad (23)$$

where

$$K = p_D^{\,n_\theta} (1 - p_D)^{n - n_\theta} \quad (24)$$

$$c_{k+1}(\theta) = e^{-\lambda} \prod_{\mathbf{z} \in Z - W_\chi} \lambda c(\mathbf{z}) \quad (25)$$

$$F_\theta = \prod_{i:\theta(i)>0} \frac{1}{\sqrt{\det 2\pi (R + HP_i H^T)}} \quad (26)$$

$$d_\theta(Z \mid X)^2 = \sum_{i:\theta(i)>0} (\mathbf{z}_{\chi(i)} - H\mathbf{x}_i)^T (R + HP_i H^T)^{-1} (\mathbf{z}_{\chi(I)} - H\mathbf{x}_i) \quad (27)$$

Even if the clutter spatial distribution is uniform, association distance is much more complex than was the case for unity probability of detection.

4 Nontraditional measurements (review)

In Chapter 5 of [3], I introduced the concept of "unambiguously generated ambiguous (UGA) measurements" and their "generalized likelihood functions." Nontraditional measurements—e.g., attributes, features, natural-language statements, and inference rules—are represented as random (closed) subsets Θ of a measurement space $\mathfrak{Z}_0$—that is as *generalized measurements*. In turn, generalized measurements are mediated by *generalized likelihood functions*, which are defined as

$$f_{k+1}(\Theta|\mathbf{x}) = \Pr(\eta_{k+1}(\mathbf{x}) \in \Theta) \quad (28)$$

where $\mathbf{z} = \eta_{k+1}(\mathbf{x})$ is a deterministic measurement model. I showed how various expert-system formalisms for representing nontraditional measurements—fuzzy logic, Dempster-Shafer theory, rule-based inference—can be used to construct random set models.

For example, consider an *imprecise measurement*, which is just a (closed) subset S of the measurement space. (A quantized measurement is a typical example of an imprecise measurement. This type of measurement is explicitly considered in a companion paper [2].) If the imprecise measurement is understood as being deterministic, then $\Theta = S$ and the corresponding generalized likelihood function is

$$f_{k+1}(S|\mathbf{x}) = \Pr(\eta_{k+1}(\mathbf{x}) \in S) = \mathbf{1}_S(\eta_{k+1}(\mathbf{x})) \quad (29)$$

where $\mathbf{1}_S(\mathbf{z})$ is the set indicator function of S.

As another example, consider a "fuzzy measurement"—i.e., a fuzzy membership function $g(\mathbf{z})$ on $\mathfrak{Z}_0$. Such a measurement can be understood as an "imprecisely specified imprecise measurement"—meaning that the measurement is imprecise, but that the specific form of the imprecision is unclear, having many possible forms $S_a = \{\mathbf{z}|\ a \leq g(\mathbf{z})\}$ for $0 \leq a \leq 1$. Using the random set representation

$$\Sigma_g = \{\mathbf{z}|\ A \leq g(\mathbf{z})\} \quad (30)$$

of $g(\mathbf{z})$, where A is a uniformly distributed random number on the unit interval $[0,1]$, I showed that its corresponding generalized likelihood is ([3], Eq. 5.29):

$$f_{k+1}(g|\mathbf{x}) = g(\eta_{k+1}(\mathbf{x})) \,. \quad (31)$$

As a third example, consider a "fuzzy Dempster-Shafer (FDS) measurement"—i.e., a basic mass assignment $\mu(g)$ on the fuzzy subsets g of $\mathfrak{Z}_0$, defined by the properties:

- $\mu(g) \geq 0$ for all g;
- $\mu(g) = 0$ if $g = 0$;
- $\mu(g) = 0$ for all but a finite number of g (the "focal fuzzy sets" of μ); and
- $\Sigma_g\ \mu(g) = 1$.

(When the focal fuzzy sets of μ are all crisp, then μ is a conventional Dempster-Shafer basic mass assignment.) A FDS measurement is a further generalization of the concept of an imprecise measurement, in which multiple hypotheses are required to represent the uncertainty in the choice of imprecise measurement. Employing a random set representation Σ_μ of μ, I showed that its corresponding generalized likelihood is ([3], Eq. (5.73)):

$$f_{k+1}(\mu \mid \mathbf{x}) = \sum_g \mu(g) \cdot g(\eta_{k+1}(\mathbf{x})) \cdot \quad (32)$$

I derived generalized likelihood functions for other nontraditional measurements, such as fuzzy inference rules $g \Rightarrow g'$ ([3], Eq. (5.80)):

$$f_{k+1|}(g \Rightarrow g' | \mathbf{x}) = (g \wedge g')(\eta_{k+1}(\mathbf{x})) + \frac{1}{2}\left(1 - g'(\eta_{k+1}(\mathbf{x}))\right) \quad (33)$$

using a random set representation $\Sigma_{g \Rightarrow g'}$ for $g \Rightarrow g'$.

Now, all of the above generalized likelihood function formulas are based on a common assumption: that the generalized measurement is deterministic. Even though random sets Σ_g, Σ_μ, and $\Sigma_{g \Rightarrow g'}$, are used to represent the nontraditional measurements g, μ, and $g \Rightarrow g'$, these measurements are themselves not random. That is, they do not arise as specific instantiations of some random variable. (For a more complete discussion see, for example, Eq. (4.23) of [3].)

In the companion paper [2], I showed how to extend this formulation to include nontraditional measurements that are instantiations of a random variable. In this case, the generalized measurement has the form

$$\Theta - \mathbf{V}_{k+1}. \quad (34)$$

where Θ is the random set model of the deterministic nontraditional measurement, and $\mathbf{V}_{k+1}$ is an additive random noise vector. In this case the generalized likelihood of the nontraditional measurement, taking randomness into account, is

$$f_{k+1}(\Theta|\mathbf{x}) = \Pr(\eta(\mathbf{x}) \in \Theta - \mathbf{V}_{k+1}). \quad (35)$$

$$= \int \mu_\Theta(\mathbf{z}) \cdot f_{k+1}(\mathbf{z} | \mathbf{x}) d\mathbf{z} \quad (36)$$

where

$$\mu_\Theta(\mathbf{z}) = \Pr(\mathbf{z} \in \Theta) \quad (37)$$

is Goodman's one-point covering function of Θ (see [3], Eq. (4.20)).

As a special case, let $\Theta = \Theta_g$ model a fuzzy measurement $g(\mathbf{z})$. Then Eq. (35) reduces to

$$f_{k+1}(g | \mathbf{x}) = \int g(\mathbf{z}) \cdot f_{k+1}(\mathbf{z} | \mathbf{x}) d\mathbf{z} \cdot \quad (38)$$

As a simple closed-form example of Eq. (38), let

$$f_{k+1}(\mathbf{z} | \mathbf{x}) = N_R(\mathbf{z} - H\mathbf{x}) \quad (39)$$

$$g(\mathbf{z}) = \sqrt{\det 2\pi C} \cdot N_C(\mathbf{z} - \mathbf{c}) \cdot \quad (40)$$

Then Eq. (38) becomes

$$f_{k+1}(g | \mathbf{x}) = \sqrt{\det 2\pi C} \cdot N_{C+R}(\mathbf{c} - H\mathbf{x}) \cdot \quad (41)$$

5 Measurement-to-track association (nontraditional measurements)

Suppose that, at time-step $k+1$,

$$X = \{(\mathbf{x}_1, P_1), \ldots, (\mathbf{x}_n, P_n)\} \quad (42)$$

is the set of tracks predicted from the previous time-step, where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are their states and $P_1, \ldots, P_n$ are their error-covariance matrices. The track densities corresponding to the $(\mathbf{x}_i, P_i)$ are not necessarily linear-Gaussian distributions (though the $(\mathbf{x}_n, P_n)$ are computed from them):

$$f(\mathbf{x} | i) \cdot \quad (43)$$

These track densities arise though standard MTA track propagation. In general, the propagation is accomplished via the single-target Bayes filter rather than, say, an EKF or UKF (see Section 5.3).

The source likelihood function at time $k+1$ is defined on generalized measurements. It is therefore a generalized likelihood function, and in particular it will not necessarily be linear-Gaussian:

$$f_i(\Theta | \mathbf{x}) \cdot \quad (44)$$

Assume that the measurement-set collected at time-step $k+1$ consists of nontraditional measurements,

$$Z = \{\Theta_1, \ldots, \Theta_m\} \quad (45)$$

with $|Z| = m$.

5.1 Local association distance (nontraditional measurements)

I proceed as in Section 3.2. Suppose that a single nontraditional measurement Θ has been collected. We are to estimate which of the predicted tracks $(\mathbf{x}_1, P_1), \ldots, (\mathbf{x}_n, P_n)$ most likely generated it, assuming that it was not a false detection. The association likelihood for Θ is then

$$\ell(\Theta | i) = \int f(\Theta | \mathbf{x}) \cdot f(\mathbf{x} | i) d\mathbf{x} \cdot \quad (46)$$

5.2 Global association distance (nontraditional measurements)

Suppose that an entire set $Z = \{\Theta_1, \ldots, \Theta_m\}$ of nontraditional measurements with $|Z| = m$ has been collected. Because the measurements are generalized, they must be mediated by a generalized likelihood function $f(\Theta|\mathbf{x})$. Consequently, the clutter process must be defined on generalized measurements, and its spatial function $c(\Theta)$ must have the form of a generalized likelihood function—that is, it must be unitless. For example, if $\Theta = \Theta_g$ represents a fuzzy measurement $g(\mathbf{z})$, one might choose

$$c(g) = c(\Theta_g) = \int g(\mathbf{z}) \cdot c(\mathbf{z}) d\mathbf{z} \cdot \quad (47)$$

For current purposes and in what follows, however, I will assume that $c(\Theta) = c$ is constant.

Case 1: $p_D = 1$. Given this, as in Section 3.3 begin by assuming that probability of detection is unity, in which case an association is a one-to-one function

$$\chi : \{1, \ldots, n\} \rightarrow \{1, \ldots, m\}. \quad (48)$$

Let W_χ denote the set of Θ in Z that equal $\Theta_{\chi(i)}$ for some $i = 1, \ldots, n$.

Then the global association likelihood, which is a generalized likelihood function, is

$$\ell_\chi(Z | X) = c(\chi) \cdot \prod_{i=1}^{n} \ell(\Theta_{\chi(i)} | i) \quad (49)$$

where

$$c(\chi) = \prod_{\mathbf{z} \in Z - W_\chi} \lambda c(\Theta) = (\lambda c)^{m-n} \quad (50)$$

which does not functionally depend on χ. In this case the probability of the association χ is

$$p_\chi = \frac{\ell_\chi(Z | X)}{\sum_{\theta'} \ell_{\chi'}(Z | X)} \quad (51)$$

$$\propto \prod_{i=1}^{n} \ell(\Theta_{\chi(i)} \mid i) \tag{52}$$

and so

$$d_\chi^2 = -\tfrac{1}{2}\sum_{i=1}^{n} \log \ell(\Theta_{\chi(i)} \mid i) \tag{53}$$

can be taken to be a global association distance.

Case 2: $p_D < 1$. If probability of detection is not unity, then we have no choice but to employ the more complex formula of Eq. (23):

$$p_\theta \propto K^{n_\theta} \cdot c(\theta) \cdot F_\theta \cdot \exp(-\tfrac{1}{2} d_\theta(Z \mid X)^2). \tag{54}$$

5.3 Track propagation

The formulas described in Sections 5.1 and 5.2 permit measurement-to-track association to be accomplished when measurements are nontraditional. Given this, how are individual tracks propagated? Under certain simplifying assumptions—e.g., when the source likelihood has a linear-Gaussian form as in Eq. (41)—the track densities $f(\mathbf{x}|i)$ in (43) will be linear-Gaussian. In this case EKF/UKF propagation can be retained.

In general, however, the $f(\mathbf{x}|i)$ will be non-Gaussian. As is well known, MTA algorithms such as MHT can employ any single-target filter, not just EKFs or UKFs. In particular, they can employ the general single-target Bayes filter (or, more precisely, some approximating implementation of it).

In this case, tracks in the track table are first time-updated using the Bayes filter prediction integral

$$f_{k+1|K}(\mathbf{x} \mid Z^k) = \int f_{k+1|k}(\mathbf{x} \mid \mathbf{x}') \cdot f_{k|k}(\mathbf{x}') d\mathbf{x}'.$$

Once MTA is applied, particular measurements have been associated with particular tracks. Given this, the Bayes filter measurement-update equation (i.e., Bayes' rule) is used to update each track using its associated measurement:

$$f_{k+1|k+1}(\mathbf{x} \mid Z^{k+1}) = \frac{f_{k+1}(\mathbf{z}_{k+1} \mid \mathbf{x}) \cdot f_{k+1|k}(\mathbf{x})}{f_{k+1}(\mathbf{z}_{k+1} \mid Z^k)}$$

where

$$f_{k+1|}(\mathbf{z}_{k+1} \mid Z^k) = \int f_{k+1}(\mathbf{z}_{k+1} \mid \mathbf{x}) \cdot f_{k+1|k}(\mathbf{x}) d\mathbf{x}$$

Then the process is repeated.

6 Closed-form examples

In this section I briefly use the results in Section 5 to derive closed-form formulas in two special cases: fuzzy measurements (Section 6.1) and fuzzy Dempster-Shafer (FDS) measurements (Section 6.2).

6.1 Fuzzy measurements

I consider three cases in turn: (1) single nonrandom fuzzy measurements, (2) multiple nonrandom fuzzy measurements, and (3) single or multiple random fuzzy measurements.

Case 1: *Single Nonrandom Fuzzy Measurements*. Let $g(\mathbf{z})$ be a fuzzy measurement. Then because of Eq. (31) the local association likelihood, Eq. (46), becomes

$$\ell(g \mid i) = \int f(g \mid \mathbf{x}) \cdot f(\mathbf{x} \mid i) d\mathbf{x} \tag{55}$$

$$= \int g(\eta_{k+1}(\mathbf{x})) \cdot f(\mathbf{x} \mid i) d\mathbf{x}. \tag{56}$$

Case 2: *Multiple Nonrandom Fuzzy Measurements*. Now assume that $\eta_{k+1} = H$ is linear and that the tracks and the fuzzy measurement are linear-Gaussian in form:

$$f(\mathbf{x} \mid i) = N_{P_i}(\mathbf{x} - \mathbf{x}_i) \tag{57}$$

$$g(\mathbf{z}) = \sqrt{\det 2\pi C} \cdot N_C(\mathbf{z} - \mathbf{c}). \tag{58}$$

Then Eq. (56) becomes

$$\ell(g \mid i) = \sqrt{\det 2\pi C} \int N_C(H\mathbf{x} - \mathbf{c})) \cdot N_{P_i}(\mathbf{x} - \mathbf{x}_i) d\mathbf{x} \tag{59}$$

$$= \sqrt{\det 2\pi C} \cdot N_{C+HP_iH^T}(\mathbf{c} - H\mathbf{x}_i). \tag{60}$$

Now assume that we have a set $Z = \{g_1,\ldots,g_m\}$ of fuzzy measurements with $m = |Z|$, and that these measurements have the form

$$g_j(\mathbf{z}) = \sqrt{\det 2\pi C_j} \cdot N_{C_j}(\mathbf{z} - \mathbf{c}_j) \tag{61}$$

in which case

$$\ell(g_j \mid i) = \sqrt{\det 2\pi C_j} \cdot N_{C_j + HP_iH^T}(\mathbf{c}_j - H\mathbf{x}_i). \tag{62}$$

For the sake of conceptual clarity, further assume that probability of detection is unity and that the clutter spatial function is constant: $c(g) = c$. Then because of Eq. (62), Eq. (52) reduces to

$$p_\chi \propto \prod_{i=1}^{n} \ell(g_{\chi(i)} \mid i) \tag{63}$$

$$= \prod_{i=1}^{n} \sqrt{\frac{\det 2\pi C_{\chi(i)}}{\det 2\pi (C_{\chi(i)} + HP_iH^T)}} \tag{64}$$

$$\cdot \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n} (\mathbf{c}_{\chi(i)} - H\mathbf{x}_i)(C_{\chi(i)} + HP_iH^T)^{-1}(\mathbf{c}_{\chi(i)} - H\mathbf{x}_i)\right)$$

This leads to a simpler expression if, finally, we assume that $C_j = C$ for all $j = 1,\ldots,m$. (In other words, the fuzzy measurements all have the same shape.[3]) Note that, by Eq. (47), $c(g)$ in this case will automatically be constant if the clutter spatial distribution $c(\mathbf{z})$ is constant.) In this case

$$p_\chi \propto \exp\left(-\tfrac{1}{2} d_\chi(Z \mid X)^2\right) \tag{65}$$

where association distance is defined by

$$d_\chi(Z \mid X)^2 = \sum_{i=1}^{n} (\mathbf{c}_{\chi(i)} - H\mathbf{x}_i)(C + HP_iH^T)^{-1}(\mathbf{c}_{\chi(i)} - H\mathbf{x}_i). \tag{66}$$

If the assumption $C_j = C$ is not made then we end up with a more complex formula for association distance,

$$d_\chi^2 = \sum_{i=1}^{n} \left(\begin{array}{l} \log \dfrac{\det(C_{\chi(i)} + HP_iH^T)}{\det C_{\chi(i)}} \\ + (\mathbf{c}_{\chi(i)} - H\mathbf{x}_i)(C_{\chi(i)} + HP_iH^T)^{-1}(\mathbf{c}_{\chi(i)} - H\mathbf{x}_i) \end{array} \right). \tag{67}$$

Case 3: *Single or Multiple Random Fuzzy Measurements*. We can generalize this analysis further by assuming that fuzzy measurements have a random component. In this case, by Eq. (38) the local association likelihood, Eq. (46), becomes

$$\ell(g \mid i) = \int f(g \mid \mathbf{x}) \cdot f(\mathbf{x} \mid i) d\mathbf{x} \tag{68}$$

$$= \int \int g(\mathbf{z}) \cdot f_{k+1}(\mathbf{z} \mid \mathbf{x}) \cdot f(\mathbf{x} \mid i) d\mathbf{x} d\mathbf{z} \tag{69}$$

$$= \int g(\mathbf{z}) \cdot \ell(\mathbf{z} \mid i) d\mathbf{z}. \tag{70}$$

[3] Note that, by Eq. (47), $c(g)$ in this case will automatically be constant if the clutter spatial distribution $c(\mathbf{z})$ is constant.

Thus using Eqs. (9) and (58), we get a closed-form example:

$$\ell(g\,|\,i) = \int f(g\,|\,\mathbf{x})\cdot f(\mathbf{x}\,|\,i)d\mathbf{x} \tag{71}$$

$$= \sqrt{\det 2\pi C}\int N_C(\mathbf{z}-\mathbf{c})\cdot N_{R+HPH^T}(\mathbf{z}-H\mathbf{x}_i)d\mathbf{z} \tag{72}$$

$$= \sqrt{\det 2\pi C}\cdot N_{C+R+HP_iH^T}(\mathbf{c}-H\mathbf{x}_i)\cdot \tag{73}$$

Note that this differs from Eq. (60) only in the presence of R in the covariance matrix. Thus Eqs. (65-67) also can be generalized in the obvious manner.

6.2 FDS measurements

I consider three cases in turn: (1) single nonrandom FDS measurements, (2) multiple nonrandom FDS measurements, and (3) single or multiple random FDS measurements.

Case 1: *Single Nonrandom FDS Measurements*. Let μ be an FDS measurement. Then because of Eq. (32) the local association likelihood, Eq. (46), becomes

$$\ell(\mu\,|\,i) = \int f(\mu\,|\,\mathbf{x})\cdot f(\mathbf{x}\,|\,i)d\mathbf{x} \tag{74}$$

$$= \sum_g \mu(g)\int g(\eta_{k+1}(\mathbf{x}))\cdot f(\mathbf{x}\,|\,i)d\mathbf{x}\,\cdot \tag{75}$$

Let the focal fuzzy subsets of μ be $g_1,\ldots,g_m$ with associated weights $w_j = \mu(g_j)$; and assume that they have the form

$$g_j(\mathbf{z}) = \sqrt{\det 2\pi C_j}\cdot N_{C_j}(\mathbf{z}-\mathbf{c}_j)\cdot \tag{76}$$

Then Eq. (69) becomes

$$\ell(\mu\,|\,i) = \sum_{j=1}^{m} w_j\sqrt{\det 2\pi C_j}\int N_{C_j}(H\mathbf{x}-\mathbf{c}_j))\cdot N_{P_i}(\mathbf{x}-\mathbf{x}_i)d\mathbf{x} \tag{77}$$

$$= \sum_{j=1}^{m} w_j\sqrt{\det 2\pi C_j}\cdot N_{C_j+HP_iH^T}(\mathbf{c}_j-H\mathbf{x}_i)\cdot \tag{78}$$

Case 2: *Multiple Nonrandom FDS Measurements*. Now assume that we have a set $Z = \{\mu_1,\ldots,\mu_m\}$ of FDS measurements with $m = |Z|$, and that these measurements have focal fuzzy subsets

$$g_j^l(\mathbf{z}) = \sqrt{\det 2\pi C_j^l}\cdot N_{C_j^l}(\mathbf{z}-\mathbf{c}_j^l)\cdot \tag{79}$$

with respective weights

$$w_j^l \tag{80}$$

for all $j = 1,\ldots,m$ and $l = 1,\ldots,m_j$. Then the generalized likelihoods for these FDS measurements are

$$\ell(\mu_j\,|\,i) = \sum_{l=1}^{m_j} w_j\sqrt{\det 2\pi C_j^l}\cdot N_{C_j^l+HP_iH^T}(\mathbf{c}_j^l-H\mathbf{x}_i)\cdot \tag{81}$$

Thus the global association probability, Eqs. (50-51), becomes

$$p_\chi \propto \prod_{i=1}^{n}\ell(\mu_{\chi(i)}\,|\,i) \tag{82}$$

$$= \prod_{i=1}^{n}\left(\sum_{l=1}^{m_{\chi(i)}} w_j\sqrt{\det 2\pi C_{\chi(i)}^l}\cdot N_{C_{\chi(i)}^l+HP_iH^T}(\mathbf{c}_{\varphi(i)}^l-H\mathbf{x}_i)\right) \tag{83}$$

If we assume that the covariance matrices of all fuzzy focal sets are equal to C, then this simplifies to

$$p_\chi \propto \prod_{i=1}^{n}\left(\sum_{l=1}^{m_{\chi(i)}} w_j\cdot\exp\left(-\tfrac{1}{2}d(\mathbf{c}_{\chi(i)}^l,H\mathbf{x}_i)^2\right)\right) \tag{84}$$

where

$$d(\mathbf{y},\mathbf{w})^2 = (\mathbf{y}-\mathbf{w})^T(C+HP_iH^T)^{-1}(\mathbf{y}-\mathbf{w})\cdot \tag{85}$$

Case 3: *Single or Multiple Random FDS Measurements*. The reasoning is similar to that for Case 3 in Section 6.1.

7 Joint kinematic/nonkinematic association

Practical applications will most typically involve states of the form $(c,\mathbf{x})$ where $\mathbf{x}$ is the kinematic state and c is a discrete nonkinematic state variable, such as target class. In this case measurements will typically have the form $(\phi,\mathbf{z})$ where $\mathbf{z}$ is a kinematic measurement and ϕ is a nonkinematic feature pertaining to c. This section shows how to apply the results in the previous sections to this more general case.

7.1 Local association likelihood

Under these assumptions, Eq. (46) will have the form

$$\ell(\phi,\mathbf{z}\,|\,i) = \sum_c\int f(\Theta_\phi,\mathbf{z}\,|\,c,\mathbf{x})\cdot f(c,\mathbf{x}\,|\,i)d\mathbf{x} \tag{86}$$

where Θ_ϕ is the random set representation of the feature ϕ. *Assume that kinematic and nonkinematic information is approximately statistically independent*. Then this can be simplified to

$$\ell(\phi,\mathbf{z}\,|\,i) = \sum_c\int f(\Theta_\phi\,|\,c)\cdot f(\mathbf{z}\,|\,\mathbf{x})\cdot f(c\,|\,i)\cdot f(\mathbf{x}\,|\,i)d\mathbf{x} \tag{87}$$

$$= \left(\sum_c f(\Theta_\phi\,|\,c)\cdot f(c\,|\,i)\right)\left(\int f(\mathbf{z}\,|\,\mathbf{x})\cdot f(\mathbf{x}\,|\,i)d\mathbf{x}\right) \tag{88}$$

$$= \ell(\phi,i)\cdot\ell(\mathbf{z}\,|\,i)\,. \tag{89}$$

This means that the kinematic and nonkinematic association likelihoods can be computed separately, and then used in more general association-related quantities such as association distance and hypothesis probability. This result requires only the independence assumption.

7.2 Global association likelihood

Under the same assumptions, the global association likelihood similarly separates into kinematic and nonkinematic parts. From Eqs. (51-52) we get

$$p_\chi \propto \prod_{i=1}^{n}\ell(\phi_{\chi(i)},\mathbf{z}_{\chi(i)}\,|\,i) \tag{90}$$

$$= \left(\prod_{i=1}^{n}\ell(\phi_{\chi(i)}\,|\,i)\right)\left(\prod_{i=1}^{n}\ell(\mathbf{z}_{\chi(i)}\,|\,i)\right) \tag{91}$$

$$\propto p_\chi^{nonkinematic}\cdot p_\chi^{kinematic} \tag{92}$$

This means that the kinematic and nonkinematic hypothesis probabilities can be computed separately and then multiplied to get the joint hypothesis probability.

7.3 Jointly processing kinematic and nonkinematic measurements

The formulas described in Sections 7.1 and 7.2 permit measurement-to-track association to be accomplished when both kinematic and nonkinematic measurements are available. Given this, how are individual tracks propagated?

Time-Update (Prediction). Both the kinematic and nonkinematic parts of each track in the track table must be time-updated. The prediction step for the kinematic part of a track is, for typical MHT implementations, accomplished using the EKF or UKF prediction equations.

The nonkinematic part of the i^{th} track at time-step k, however, consists not of a state-vector and covariance matrix but, instead, a probability distribution $p_{k|k}(c) = f(c|i)$ on target type c.[4] In this case the prediction step for c is the Bayes filter prediction equation:

$$p_{k+1|k}(c) = \sum_{c'} p_{k+1|k}(c \mid c') \cdot p_{k|k}(c') \cdot \tag{93}$$

Here the quantity $p_{k+1|k}(c|c')$ is a Markov transition matrix, which describes the probability that a target will have identity c at time-step k+1 if it had identity c' at time-step k.

Remark: Ordinarily, a target does not change identity over time, in which case $p_{k+1|k}(c|c') = \delta_{c,c'}$. However, this is not always the case. Diesel-electric submarines, for example, have quite different sensor phenomenologies depending on whether they are snorkeling or submerged. Thus both $p_{k+1|k}(snork|submg)$ and $p_{k+1|k}(submg|snork)$ will typically be nonzero. If these transitions are not taken into account, substantially reduced filter performance can be the result.

Measurement-update (Correction). Both the kinematic and nonkinematic parts of each track in the track table must be measurement-updated. The correction step for the kinematic part of a track is, for typical MHT implementations, accomplished using the EKF or UKF correction equations.

Once again, the nonkinematic part of a track at time-step k consists not of a state-vector and covariance matrix but, instead, a time-updated probability distribution $p_{k+1|k}(c)$ on target type c. In this case the measurement-update for the nonkinematic state is just Bayes' rule:

$$p_{k+1|k+1}(c) = \frac{p_{k+1}(\phi \mid c) \cdot p_{k+1|k}(c)}{\sum_{c'} p_{k+1}(\phi \mid c') \cdot p_{k+1|k}(c')} \tag{94}$$

where, as usual, $p_{k+1}(\phi|c)$ denotes the likelihood function of the nonkinematic measurement ϕ.

8 Conclusion

Measurement-to-track association (MTA) remains the dominant conceptual paradigm for multisource-multitarget detection, tracking, and identification—and for most legacy algorithms in particular. The theoretical foundations of MTA are based on the presumption that collected measurements, like tracks, are kinematic entities that can be easily represented in terms of vectors and covariance matrices.

It is not obvious how these same foundations might be applied when the measurements are not kinematic—when, for example, they pertain to target identity. In such cases, the measurements can take the form of attributes or features, or even natural-language statements and inference rules.

To address this fact, this paper has shown how to extend conventional MTA theory to nontraditional measurements in a systematic and theoretically disciplined, Bayesian, manner. The fundamental unifying conceptual bridge is the "generalized likelihood function," which allows conventional MTA reasoning to be extended to nontraditional measurement-types.

Concepts such as association likelihood, association distance, hypothesis probability, and global nearest-neighbor distance were defined. Closed-form formulas were derived for specific kinds of nontraditional evidence. Special attention was paid to the case when probability of detection is unity, since relatively simple closed-form formulas can be derived in this case.

As with any theoretical development, the practical efficacy of the approach in this paper is yet to be determined. Implementation and simulation-based testing of it is a natural subject of future research.

References

[1] R. Mahler, *Bayesian Fusion and Filtering of Nontraditional Data*, Tutorial No. 15, Proc. 13th Int'l Conf. on Information Fusion, Edinburgh, Scotland, July 26-29, 2010.

[2] R. Mahler, *General Bayes filtering of quantized measurements*, Proc. 14th Int'l Conf. on Information Fusion, Chicago, Illinois, July 5-9, 2011.

[3] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Artech House, Boston, 2007.

[4] S. Blackman and S. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 2000.

[5] R. Mahler, *Multitarget filtering via first-order multitarget moments*, IEEE Trans. Aerospace and Electronics Sys., Vol. 39, No. 4, pp. 1152-1178, 2003.

[6] R. Mahler, *PHD filters of higher order in target number,* IEEE Trans. Aerospace and Electronic Sys., Vol. 43, No. 4, pp. 1523-1543, 2007

[4] More generally, the target types can be organized into a multi-level taxonomy or ontology, in which case one must employ probability distributions on the different levels.