Track-to-Track Association Using Attributes

YAAKOV BAR-SHALOM
University of Connecticut

HUIMIN CHEN
University of New Orleans

The problem of track-to-track association—a prerequisite for the fusion of tracks—has been considered in the literature for tracks described by kinematic states and, more recently, has been generalized to include additional (continuous valued) feature and (discrete valued) attribute variables which pertain to those tracks. These approaches allow the search for the maximum likelihood (ML) or maximum a posteriori (MAP) association. However, while for kinematic variables there is a "gating" procedure based on a Gaussian distribution—which corresponds to a Neyman-Pearson test of "common origin" (actually, "same kinematic state") with selectable power—there is no simple counterpart of this for attributes. The sufficient statistic for the optimal association test (in the Neyman-Pearson sense) based on discrete-valued target classification information observables (attributes) is derived and its relationship with the class probability vector is discussed. Based on this, "attribute gates" are presented, which allow a Neyman-Pearson test for "same class" with the desired power.

Manuscript received December 8, 2006; revised August 1, 2007; released for publication January 30, 2008.

Refereeing of this contribution was handled by Benjamin Slocumb.

Authors' adresses: Y. Bar-Shalom, University of Connecticut, Dept. of Electrical & Computer Engineering, Storrs, CT 06269-2157, E-mail: (ybs@ee.uconn.edu); H. Chen, University of New Orleans, Dept. of Electrical Engineering, New Orleans, LA 70148, E-mail: (hchen2@uno.edu).

1557-6418/07/$17.00

1. INTRODUCTION

The problem of track-to-track association (T2TA)—a prerequisite for the fusion of tracks—has been considered initially in the literature for tracks described by kinematic states [1]. More recently, it has been generalized to include additional (continuous valued) feature and (discrete valued) attribute variables which pertain to those tracks.[1] These approaches allow the search for the maximum likelihood (ML) or maximum a posteriori (MAP) association.

It turns out that, under the Gaussian assumption on the estimation errors—which applies to the kinematic states and, possibly, the features—a simple sufficient statistic exists for the track association hypothesis testing and, consequently, it is easy to find the threshold for the desired Neyman-Pearson test of "common origin" (actually, "same kinematic state") with a selectable power. However, this does not apply for attributes, which are discrete valued. In this paper the sufficient statistic for the optimal association test in the Neyman-Pearson sense is derived for discrete-valued attribute/classification information and its relationship with the class probability vector is discussed.

Feature-aided T2TA was presented in [19, 20, 8]. A comprehensive procedure for incorporation of attributes and their possible dependence on the features was presented in [23, 14, 15] and shown to be amenable to obtain the MAP association of tracks from two sensors using linear programming. A multiple model approach for feature aided tracking (FAT) was presented in [22]. Classification-aided tracking with measurement-to-track association via multidimensional assignment (MDA) was discussed in [4]. However, while these approaches provide the ML or MAP association, they do not provide the means to set up a statistical hypothesis test with a desired power.

Target features and attributes/classification outputs are in general useful for track-to-track association especially when targets are closely spaced and the association based on the kinematic states only is unreliable. In some cases one deals with sensors that provide target attributes but the associated kinematic information is highly inaccurate. In such a case it is of interest to provide association decisions based on the attributes/classification information alone. This is the major motivation for the present work.

The rest of the paper is organized as follows. Section 2 discusses briefly the use of (continuous valued) kinematic and feature variables for track-to-track for association. The modeling and use of (discrete valued) attribute/classification information for track-to-track association is presented in Section 3. The modeling of the

[1] Examples of features are radar cross-section and target length. Examples of attributes are number of engines of an aircraft and type of emitter/waveform. Target classes can be, e.g., fighter vs. bomber or specific aircraft type. A detailed discussion of target features, attributes and classification can be found in [9, 10, 11, 12, 13].

classifier is discussed and the classification sufficient statistic is derived under the assumption of a constant "confusion matrix." The sufficient statistic calculation from the class probability vector is presented. To obtain the Neyman-Pearson test for "common class," the statistical characterization of the classification information sufficient statistic is developed in Section 4. Track-to-track association using all the information is discussed in Section 5. This section also presents a suboptimal test statistic for association based on classification information, as well as the (optimal in the Neyman-Pearson sense) likelihood ratio test and discusses the methodology for the performance evaluation. Numerical examples of the use of the likelihood function and likelihood ratio test are given in Section 6. Section 7 presents conclusions.

2. TRACK-TO-TRACK ASSOCIATION USING KINEMATIC AND FEATURE STATES

The problem of track-to-track association (T2TA) has been considered in the literature only for tracks described by kinematic states. If tracks also include continuous valued nonkinematic features which are estimated together with the kinematic states, they should be considered as part of the state in the process of track-to-track association. Once a decision on common origins of the tracks (actually, "same state") is made, the state estimates of those tracks deemed to correspond to the same target (have the same true state) can be fused to yield a more accurate target state estimate.

Assuming that the estimates of these features are obtained in a manner similar to the kinematic variables, their errors can be taken as zero mean Gaussian random variables. In this case, the association likelihood function based on the *augmented state*, consisting of the kinematic and feature components, can be expressed in the standard form [1]. This holds for continuous valued observations of both continuous as well as discrete valued features (the latter was discussed and illustrated in [13]). A special case is the target radar cross-section (RCS), which is positive and a Gaussian model is not appropriate. While a Swerling fluctuation model can be easily used for measurement to track association [1], for track-to-track association, there is no known sufficient statistic in this case for the hypothesis test, like the difference of state estimates when they have Gaussian errors. While a possible approach could be to use the difference of the RCS estimates (or their logarithms) with an (approximate) Gaussian assumption, the true RCS for different sensors is probably not the same due to the different aspect angles, which would make its use questionable. The situation of discrete valued observations, which is related to attributes/classifications, is discussed in Section 3.

For a pair of tracks, one can accept the "same state" hypothesis $H_1 : \mathcal{C}^{ij}$ for track i from one sensor and track j from another sensor if the normalized distance between their (augmented) local state estimates $\hat{x}^i$ and $\hat{x}^j$ (at a common time, not indicated in the above notation for simplicity), which is chi-square distributed, is "not too large." Specifically, the squared norm

$$D^{ij} = (\hat{x}^i - \hat{x}^j)'[T^{ij}]^{-1}(\hat{x}^i - \hat{x}^j) \quad (1)$$

has to be within the $1-\alpha$ probability region of the chi-square distribution with n_x degrees of freedom for acceptance of the same state hypothesis, i.e.,

$$D^{ij} \leq \chi^2(1-\alpha) \quad (2)$$

where the notation from [3] has been used. In (1)

$$T^{ij} = P^i + P^j - P^{ij} - P^{ji} \quad (3)$$

is the covariance of the difference between the local state estimation errors, which includes the local estimation error covariances P^i, P^j and the crosscovariance term P^{ij} due to the common process noise [1]. Note that for the feature part of the (augmented) state, which can be assumed in general time invariant without process noise, the crosscovariance of feature estimation errors between two local trackers is zero. For a practical way to obtain the crosscovariances for kinematic state components, see [7]. The acceptance region for H_1, defined by (2), is called *kinematic gate* [1].

According to the above, $\mathcal{C}^{ij}$ is rejected *if there is too much evidence against it*—the difference between the estimates is too large (relative to their accuracies, quantified by the covariance matrices) to accept that they are from the same true state. The hypothesis $H_1 : \mathcal{C}^{ij}$ is called in the literature "common origin," but it is more accurate to call it "same true kinematic state," as (4) indicates.

Due to the correlation in time of the track state estimation errors [3], the test (1) is based on the track estimates at a single (common) point in time. Typically, since sensors (and the corresponding local estimators) are not synchronized, one of the state estimates will be a (short interval) prediction.

The test statistic (1), while commonly used in the literature [1, 6] without proof of its validity, was proven for the first time in [18]. This proof is briefly outlined below, because it will be the basis of a similar approach for the case of track association with classification information to be presented in the next section.

The likelihood function of the *same kinematic state* hypothesis $\mathcal{C}^{ij}$ is the pdf of the track sufficient statistics (the kinematic state estimates) conditioned on $\mathcal{C}^{ij}$, namely,

$$\Lambda_{\text{kin}}(\mathcal{C}^{ij}) \triangleq p(\hat{x}^i, \hat{x}^j \mid \mathcal{C}^{ij}) = \int_{\mathcal{V}} p(\hat{x}^i, \hat{x}^j \mid x) p(x) dx \quad (4)$$

where x is the true common state of the two tracks and $\mathcal{V}$ is the region in which x takes values. The r.h.s. of (4) follows from the total probability theorem.

Assuming the joint pdf of the local state estimates to be

$$p(\hat{x}^i,\hat{x}^j \mid x) = \mathcal{N}\left(\begin{bmatrix}\hat{x}^i\\ \hat{x}^j\end{bmatrix};\begin{bmatrix}x\\ x\end{bmatrix},\begin{bmatrix}P^i & P^{ij}\\ P^{ji} & P^j\end{bmatrix}\right) \tag{5}$$

and using a *diffuse (noninformative) prior* [3] within a "sufficiently large" region $\mathcal{V}$, with volume V,

$$p(x) = V^{-1} \tag{6}$$

the likelihood function (4) becomes [18]

$$\Lambda_{\text{kin}}(\mathcal{C}^{ij}) = V^{-1}\mathcal{N}(\hat{x}^i - \hat{x}^j; 0, T^{ij}). \tag{7}$$

This shows the validity of using the test (2), which is based on the likelihood function (7) and states the following:

"Reject the same state (and thus the common origin) hypothesis if the normalized distance is in the α-tail of its distribution."

Note that the likelihood function (7) is a Gaussian pdf with the simple sufficient statistic (1). The Gaussian pdf makes it very easy to arrive at the test (2), which excludes its tail. This is the main reason to assume the estimation error of a continuous valued feature variable to be Gaussian.

REMARK The sufficient statistic (1) follows from the likelihood function of the "common origin" hypothesis. Assuming for the alternative hypothesis H_0: $\mathcal{C}^{i\neq j}$ ("different origin") a uniform diffuse distribution [3] i.e., a constant

$$\Lambda_{\text{kin}}(\mathcal{C}^{i\neq j}) = c \tag{8}$$

yields the likelihood ratio for the Neyman-Pearson test as (7) divided by the above constant c, which is irrelevant—the likelihood function and likelihood ratio tests are effectively the same (and have the same ROC curve). Consequently, (1), which can be seen to be the negative log-likelihood ratio, is optimal in the Neyman-Pearson sense and the power of the test (2) is $1-\alpha$.

The difference of the estimates, as in (7), will not be an exact sufficient statistic for the situation of continuous valued observations on a discrete valued feature. Nevertheless, as shown in Section 5.2, one can use such an approximate sufficient statistic effectively.

The generalization of (7) to an arbitrary number of tracks can be found in [5]. This allows the association of tracks from an arbitrary number of sensors based on their local estimates, covariances and crosscovariances.

3. TRACK-TO-TRACK ASSOCIATION USING DISCRETE ATTRIBUTE/CLASSIFICATION INFORMATION

This section deals with the modelling and use of (discrete valued) attribute/classification information for track-to-track association. The modelling of the classifier is discussed and the classification sufficient statistic is derived under the assumption of a constant "confusion matrix."

Consider the case where a track contains observations of discrete valued attributes from which one can infer the target's class. The following model will be assumed for the target classes. It is assumed that there are N_c classes of targets. Let the possible classes be

$$\kappa \in K = \{1,\ldots,N_c\}. \tag{9}$$

The target class is assumed to be time invariant.

If two tracks belong to different classes, they obviously cannot have a common origin. The converse, however, is not true: if two tracks belong to the same class, they do not necessarily originate from the same target, unless the class is a unique identity. Thus, what can be accomplished with class information is testing whether two tracks belong to the *same class*, rather than originating from the same target. This is similar to the test based on kinematic state estimates where the test is, rigorously speaking, "same kinematic state" rather than "same origin."

3.1. Modelling of the Classifier Output

Let ζ denote the output of the classifier. The classifier's output is an attribute (an element of a discrete set), which is related to the presumed class to which the target under consideration belongs, as discussed below. The output set can have, in general, a larger number of elements than the set of target classes.[2] Then

$$\zeta \in K_a = \{1,\ldots,N_a\} \supseteq K \tag{10}$$

and it is assumed that one has

$$c_{nm} = P\{\zeta = m \mid \kappa = n\}, \qquad n = 1,\ldots,N_c, \quad m = 1,\ldots,N_a \tag{11}$$

which are the elements of the "confusion matrix" (see, e.g., [17])

$$C = [c_{nm}]. \tag{12}$$

Note that c_{nm} is the **likelihood** (probability of the observable conditioned on the truth of interest, see, e.g., [3]) of the true class being n when the classifier output (the observable) is $\zeta = m$. Thus the class likelihood function for classifier output m is the mth column of the confusion matrix C. One can conceivably have the likelihood functions depend on additional variables, like target kinematic state (e.g., aspect angle, distance to target), lighting, etc.

In the sequel it is assumed that all elements in the confusion matrix are constant and the same across dif-

[2] For example, one can have an "undetermined" class or a "class n_1 or n_2." Following [13], we use the term attribute for the observable from which a probabilistic inference can be made on the target class.

ferent classifiers. These restrictions can be removed by using a time argument and/or a classifier index, in which case, for classifier i at time k one would have the likelihoods $c_{nm}^i(k)$. However, in this case there is no sufficient statistic as in Section 3.3. Furthermore, it will be assumed that the classifier outputs are, conditioned on the truth, independent across time (it has "white" errors) and independent of the kinematic and feature variables. This is the counterpart of the white measurement noise for the kinematic measurements.[3]

3.2. Update of the Classification Probabilities

Denote by μ_n^0 the prior probability of class n (prior to the observation under consideration). The posterior (or updated) probability of a target being in class n, given that the classifier's output is m, is

$$\mu_n = P\{\kappa = n \mid \zeta = m\} = \frac{c_{nm}\mu_n^0}{\sum_{l=1}^{N_c} c_{lm}\mu_l^0}. \qquad (13)$$

The corresponding class probability vector of the target under consideration can be written as

$$\mu = \frac{c_m \otimes \mu^0}{c_m' \mu^0} \qquad (14)$$

where c_m is the mth column of C, μ^0 is the prior probability vector and $\otimes$ is the Schur-Hadamard product (term by term) [16].

Similarly, for a track—*a sequence of associated measurements that includes classification information*—the updated class probability vector at time k, with classifier output m, is given by

$$\mu(k) \stackrel{\Delta}{=} \mathrm{col}[P\{\kappa = n \mid \zeta(k) = m, \zeta^{k-1}\}] = \frac{c_m \otimes \mu(k-1)}{c_m' \mu(k-1)} \qquad (15)$$

where ζ^{k-1} denotes the cumulative classification information at time $k-1$ and

$$\mu(0) = \mu^0 \qquad (16)$$

is the prior before getting any classification information.

It is worth mentioning that it is not via the probabilities (14) that the observations are used in the update (15), unless they are based on a uniform (i.e., noninformative) prior [3]. The update (15) requires the latest observation to enter via the *likelihood function* c_m. If one has only probabilities as in (14) with nonuniform priors, one can use the update/fusion procedure from Sec. 8.5.2 of [1], which avoids the "double counting" of the prior information.

3.3. The Sufficient Statistic for Classification

Denote the output of the classifier at time t as $m(t)$, $t = 1,\ldots,k$. The recursion (15) can be rewritten as follows

$$\mu(k) = \frac{1}{\alpha} c_{m(k)} \otimes c_{m(k-1)} \otimes \cdots \otimes c_{m(1)} \otimes \mu^0 \qquad (17)$$

where α is the normalizing constant.

Note that, since the target class was assumed to be time invariant, the actual times at which the classifier outputs are generated are not relevant. Furthermore, because of the commutativity of the Schur-Hadamard product in (17), it can rewritten as follows

$$\mu(k) = \frac{1}{\alpha} c_1^{[\nu_1]} \otimes c_2^{[\nu_2]} \otimes \cdots \otimes c_{N_a}^{[\nu_{N_a}]} \otimes \mu^0 \qquad (18)$$

where ν_m is the number of times the output of the classifier was m and $c_m^{[\nu_m]}$ is c_m raised to the power ν_m with the Schur-Hadamard product.

Since

$$c_m^{[\nu_m]} = [(c_{1m})^{\nu_m} \ (c_{2m})^{\nu_m} \cdots (c_{N_c m})^{\nu_m}]', \qquad m = 1,\ldots,N_a \qquad (19)$$

the sufficient statistic for the classifier output can be seen to be the number of times each output class was generated, i.e., it is the vector

$$\nu = [\nu_1,\ldots,\nu_{N_a}]'. \qquad (20)$$

The existence of the above sufficient statistic hinges on the assumption that the confusion matrix is constant. Otherwise it appears that there is no such sufficient statistic.

3.4. Calculation of the Sufficient Statistic from the Classification Probabilities

In practice it is more likely that the information provided by the classifier will be the class probability vector (18) rather than the sufficient statistic (20). In this case we need to recover the vector ν from the vector μ. Note that both ν and μ consist of N_a elements but the elements of μ sum up to unity after normalization by α, which is not known. The elements of ν sum up to N, the number of times the classifier provided an output i.e.,

$$\sum_{m=1}^{N_a} \nu_m = N. \qquad (21)$$

It is assumed that N and μ^0 are known. This allows the substitution

$$\nu_1 = N - \sum_{m=2}^{N_a} \nu_m \qquad (22)$$

[3]Just like the measurements are correlated because they observe (nearly) the same state, the classifier outputs will be correlated because the observe the same true variable, the class of the same target. However, the errors of the classifiers, like the errors of the sensor providing the kinematic measurements, are assumed to be white.

in (18), which can be written for component n as (the time index is omitted)

$$\mu_n = \frac{\mu_n^0}{\alpha}\prod_{m=1}^{N_a}(c_{nm})^{\nu_m} = \frac{\mu_n^0 (c_{n1})^N}{\alpha}\prod_{m=2}^{N_a}\left(\frac{c_{nm}}{c_{n1}}\right)^{\nu_m}, \quad n = 1,\ldots,N_c. \tag{23}$$

The above is a set of N_c equations in the unknowns consisting of ν_m, $m = 2,\ldots,N_a$, and α.

Taking the log of equations (23), one obtains a set of N_c linear equations in the N_a unknowns ν_m, $m = 2,\ldots,N_a$, and $\log\alpha$. If $N_a = N_c$, this set has a unique solution, allowing us to obtain ν_m, $m = 2,\ldots,N_a$ (note that α, while part of the solution, is of no interest); ν_1 follows from (22). This provides the complete solution in this case for the classification sufficient statistics ν_m, $m = 1,\ldots,N_a$. As it will be shown later, this vector ν (and not μ) will be needed for the test whether two tracks belong to the same class. If $N_a > N_c$, then in general one cannot find the classification sufficient statistics ν_m, $m = 1,\ldots,N_a$ uniquely by solving the above equations because there are not enough equations. In this case, if one has only the classification probabilities, then the likelihood function of the same class hypothesis cannot be fully specified. If $N_c > N_a$, then one can use a subset of N_a equations from (23) to obtain the likelihoods.

4. THE SAME CLASS LIKELIHOOD FUNCTION FROM CLASSIFICATION INFORMATION

This section presents the probability mass function (pmf) of the classifier's sufficient statistic, which is the basis of the Neyman-Pearson test.

The pmf, denoted as $P[\cdot]$, of the cumulative (local) classifier information using the sufficient statistic for the (local) track i is, for a total number of N^i classifier outputs, if the true class is n, given by the multinomial distribution [21]

$$P[\nu^i \mid \kappa^i = n] = P[\nu_1^i,\ldots,\nu_{N_a}^i \mid \kappa^i = n] = N^i!\prod_{m=1}^{N_a}\frac{c_{nm}^{\nu_m^i}}{\nu_m^i!} \tag{24}$$

where the total number of classifier outputs is

$$\sum_{m=1}^{N_a}\nu_m^i = N^i. \tag{25}$$

For simplicity, we assume that the local classifiers have the same confusion matrix C which is known at the fusion center. Furthermore, the outputs of the two classifiers are assumed, conditioned on the truth, independent.[4] Then the likelihood of the "same class" hypothesis H_1: $\mathcal{C}^{ij}$ can be written as follows

$$\begin{aligned}\Lambda_{\text{class}}(\mathcal{C}^{ij}) &\stackrel{\Delta}{=} P[\nu^i,\nu^j \mid \mathcal{C}^{ij}] = \sum_{n=1}^{N_c}P[\nu^i,\nu^j \mid \kappa^i = \kappa^j = n]\mu_n^0 \\ &= \sum_{n=1}^{N_c}P[\nu^i \mid \kappa^i = n]P[\nu^j \mid \kappa^j = n]\mu_n^0 \\ &= \sum_{n=1}^{N_c}N^i!N^j!\left[\prod_{m=1}^{N_a}\frac{c_{nm}^{\nu_m^i+\nu_m^j}}{\nu_m^i!\nu_m^j!}\right]\mu_n^0\end{aligned} \tag{26}$$

where κ^i and κ^j are the true classes of tracks i and j, respectively. Note that while in (4) the total probability theorem was used with the diffuse (or improper) prior (6) to yield (7), in (26) the proper (and not necessarily uniform) prior μ^0 was used (since κ takes values in a finite set). In (26) it is assumed that the classification errors are independent across time and sensors.

Using the same approach as in the case of the continuous valued states, we propose, based on the class information, to reject the common origin hypothesis *if there is too much evidence against it*: if the likelihood (26) is in the tail of its distribution.

We are faced with two problems here:

1) There is no simple sufficient statistic similar to the normalized distance in the continuous/Gaussian case. The difference of the two vectors ν^i and ν^j is not the exact sufficient statistic for the hypothesis test. Nevertheless, a suboptimal method based on this is explored in the next section and evaluated later.

2) While there is an expression for the likelihood function pmf, to find its "tail," an exhaustive evaluation of all its point masses is needed: these point masses have to be ordered and the "tail" identified.

An alternative approach would be to use a Monte Carlo method to determine whether a particular pair ν^i, ν^j is in the tail of the distribution. To evaluate a 5% tail probability with $2\sigma = 1\%$, one needs 2000 runs (of $N^i + N^j$ classifications), which is not too excessive.

In the examples presented later, the exhaustive evaluation of the likelihood function pmf is carried out, since it is not too expensive computationally. The number of points at which (26) has to be evaluated is obtained below.

The number of points (different outcomes) for the pmf (24) is

$$M(N_a,N^i) = (N_a)^{N^i} - \sum_{\substack{k_1,\ldots,k_{N_a}\in\{0,1,\ldots,N^i\}\\ k_1+\cdots+k_{N_a}=N^i}}\left(\frac{(N^i)!}{k_1!\cdots k_{N_a}!}-1\right). \tag{27}$$

The above result follows from Eq. (1.18) in Ch. 2 of [21]. Namely, it is the number of elements in the expansion of a multinomial of N_a elements raised to

[4]This is the counterpart of assuming independent measurement noises at different sensors.

the power N^i: it is $(N_a)^{N^i}$ less the number of outcomes which are equivalent because the order does not matter. The latter is given, for a certain outcome, by the corresponding multinomial coefficient less one because for each outcome we remove only the duplicates.

The number of points at which (26) has to be evaluated is then

$$M(N_a, N^i, N^j) = M(N_a, N^i)M(N_a, N^j). \tag{28}$$

5. TRACK-TO-TRACK ASSOCIATION TESTING USING ALL THE INFORMATION

This section proposes a procedure to test for common origin by decoupling the continuous variables from the discrete ones. Then the details of a simpler suboptimal test as well as the optimal (likelihood ratio) test using the discrete attribute variables are presented.

5.1. Test using the Likelihood Function of H_1

It seems reasonable to accept that tracks i and j are from the same target if

1) Their continuous valued state (kinematic and feature) estimates are "close enough" to accept that their kinenatic/feature stats are the same—they satisfy (2), and

2) Their classification does not present strong evidence that they cannot be the same, i.e., $\Lambda_{\text{class}}(\mathcal{C}^{ij})$, given in (26), falls into the $1-\alpha$ probability concentration region under H_1. In this case the "same class" hypothesis H_1 is accepted.

This sequence of tests assumes implicitly that the classification errors are independent of the kinematic and feature variables. While one can write a joint likelihood if there is a dependence [23], a joint test does not seem to be available at this point due to the mixed nature (continuous-discrete) of the likelihood function and the ensuing lack of a joint sufficient statistic.

5.2. Test using the Difference of Local Classification Sufficient Statistics

To reduce the complexity of the likelihood function given in (26) and thus simplify the computation of the probability region (the "attribute gate"), we also consider using the difference of the local classification sufficient statistics in the hypothesis test. The results are particularly useful for the "same class" test involving only two targets.

As a preliminary, consider the difference

$$z = x - y \tag{29}$$

of two discrete-valued random variables

$$x \in \{0,\ldots,n_x\} \qquad y \in \{0,\ldots,n_y\} \tag{30}$$

which yields

$$z \in \{-n_y,\ldots,n_x\}. \tag{31}$$

Then

$$P\{z = n\} = \sum_{k=n}^{n_x} P\{x = k\}P\{y = k-n\}. \tag{32}$$

Let the difference of the local classification sufficient statistics given in (20), now superscripted by the sensor index, be

$$\delta^{ij} = [\delta_1^{ij},\ldots,\delta_{N_a}^{ij}] = [\nu_1^i - \nu_1^j,\ldots,\nu_{N_a}^i - \nu_{N_a}^j] \tag{33}$$

which yields

$$\delta_k^{ij} \in \{-N^j,\ldots,N^i\}, \qquad k = 1,\ldots,N_a. \tag{34}$$

The pmf[5] of the above vector is, based on (32), given by

$$P[\delta^{ij} \mid \kappa = n] = \sum_{k_1=\delta_1^{ij}}^{N^i} \cdots \sum_{k_{N_a}=\delta_{N_a}^{ij}}^{N^i} P\{\nu_1^i = k_1,\ldots,\nu_{N_a}^i = k_{N_a} \mid \kappa = n\}$$
$$\times P\{\nu_1^j = k_1 - \delta_1^{ij},\ldots,\nu_{N_a}^i = k_{N_a} - \delta_{N_a}^{ij} \mid \kappa = n\} \tag{35}$$

where the probabilities in the N_a-fold summation above are given in (24).

5.3. Test using the Likelihood Ratio of H_1 vs. H_0

There is an alternative approach to use the classification information: rely on the likelihood ratio of the two hypotheses, rather than only the likelihood function of H_1. This is done as follows.

The hypothesis H_0: $\mathcal{C}^{i\neq j}$ that the two targets are different, i.e., *belong to different classes*, is composite and can be written with the total probability theorem as follows

$$\Lambda_{\text{class}}(\mathcal{C}^{i\neq j}) \triangleq P[\nu^i,\nu^j \mid \mathcal{C}^{i\neq j}]$$
$$= \sum_{n=1}^{N_c} \sum_{l=1,l\neq n}^{N_c} P[\nu^i,\nu^j \mid \kappa^i = n \neq \kappa^j = l]\mu_n^0\mu_l^0$$
$$= \sum_{n=1}^{N_c} \sum_{l=1,l\neq n}^{N_c} N^i!N^j! \left[\prod_{m=1}^{N_a} \frac{c_{nm}^{\nu_m^i} c_{lm}^{\nu_m^j}}{\nu_m^i!\nu_m^j!}\right] \mu_n^0\mu_l^0. \tag{36}$$

The test statistic (based on the classification information only) is then the ratio of (26) to (36) i.e.,

$$\lambda_{\text{class}}(\mathcal{C}^{ij} : \mathcal{C}^{i\neq j}) = \frac{\Lambda_{\text{class}}(\mathcal{C}^{ij})}{\Lambda_{\text{class}}(\mathcal{C}^{i\neq j})} \tag{37}$$

[5]Denoted as $P[\cdot]$, while $P\{\cdot\}$ denotes the probability of an event.

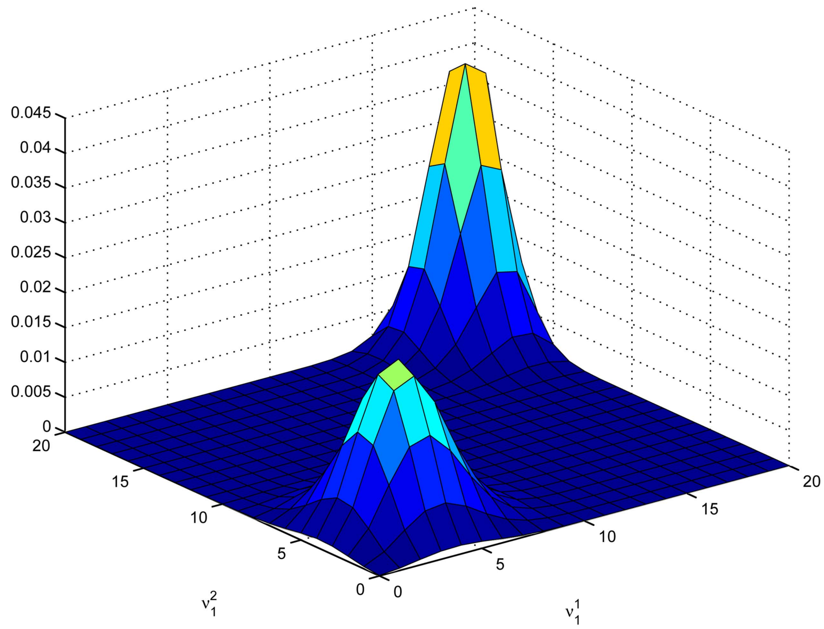


Fig. 1. The probability mass function of the local classifier sufficient statistics (ν_1^1, ν_1^2) under the same class hypothesis (likelihood function of H_1).

and the acceptance region for H_1 is obtained by finding the $1-\alpha$ probability concentration region for this ratio under H_1.

For the kinematic and feature variables, typically, the alternate hypothesis is usually modeled by a diffuse pdf [3], i.e., it amounts to a constant that rescales the likelihood function into the likelihood ratio. Thus they are effectively the same test. Consequently, we shall focus on the tests based on classification variables.

5.4. Performance Evaluation of the Tests

To find the probability of false alarm (acceptance of H_1 when H_0 is true) for the likelihood function based test, one has to evaluate numerically the probability mass of (26) under H_0 in the acceptance region for H_1.

For the likelihood ratio based test, one has to evaluate the probability mass of the likelihood ratio test statistic (37) under H_0 in the acceptance region for H_1. These are illustrated in the next section.

6. NUMERICAL EXAMPLES

Assume that each target belongs to one of $N_c = 2$ classes and each classifier output provides target attribute (taken here as its class, i.e., $N_a = 2$) with the accuracy given by the following time invariant confusion matrix

$$C = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}. \tag{38}$$

Based on the sufficient statistics from two local classifiers, we want to test whether the two targets belong to the same class (hypothesis H_1) vs. different classes (hypothesis H_0). Note that the common origin hypothesis will be rejected if according to the hypothesis test it is very implausible. Assume that the target has equal prior probability of belonging to each of the two classes and the total number of outputs for each local classifier is $N^i = 20$, the pmf of the sufficient statistic under the "same class" hypothesis is shown in Figure 1 for different ν_1^i, $i = 1,2$.[6]

We can see the two peaks at $(18,18)$ and $(4,4)$ corresponding to the two possible classes of the truth. In general the pmf will have N_a (number of attributes) peaks due to the uncertainty about the truth. The decision region based on the *likelihood function* (26) to allow 5% missed detection of H_1 is shown in Figure 2. This follows from the $\alpha = 5\%$ tail probability mass of the pmf plotted in Figure 1. Unlike in the Gaussian case (for continuous valued states), the "same class" test (which is for discrete states), in general, yields an irregular decision region. This is the *attribute gate*, defined in the ν_1^i, $i = 1,2$ space. In this example we have only two possible combinations for the evaluation of (36) when the two targets belong to two different classes, so one can evaluate the likelihood function of H_0 relatively easily.

The pmf of $\nu_1^2 - \nu_1^1$ is shown (as a bar plot) in the top part of Figure 3. It is symmetric around zero, as expected. The (moment matched) normal probability plot (cdf—cumulative distribution function) in the bottom part of Figure 3 indicates that the difference of the local classification statistics (whose cdf is the shown staircase function) can be well approximated by a Gaussian distribution. By numerical calculation we found that the decision region (attribute gate for the above difference)

[6] In view of (25) and the fact that there are only two classes, ν_1^i is a sufficient statistic for classifier i.

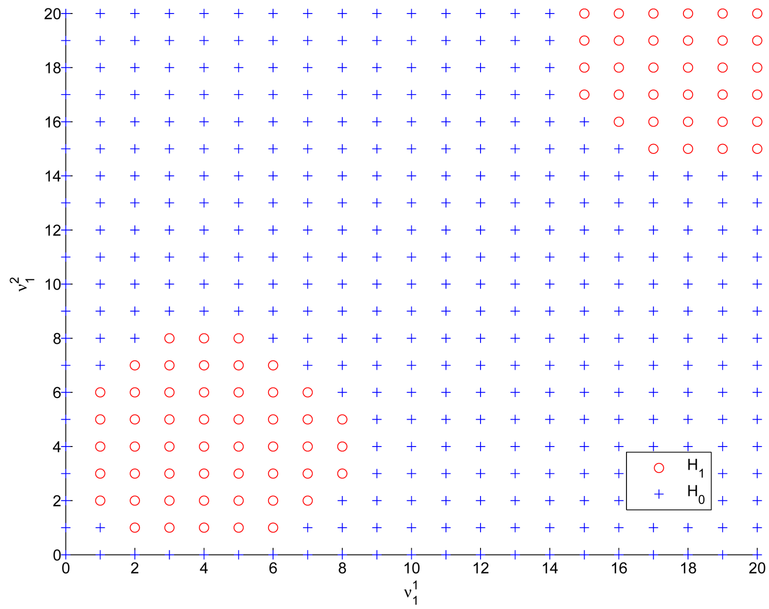


Fig. 2. Decision regions (attribute gate in the (ν_1^1, ν_1^2) space) based on the likelihood function of H_1 with 5% probability of incorrectly rejecting it.

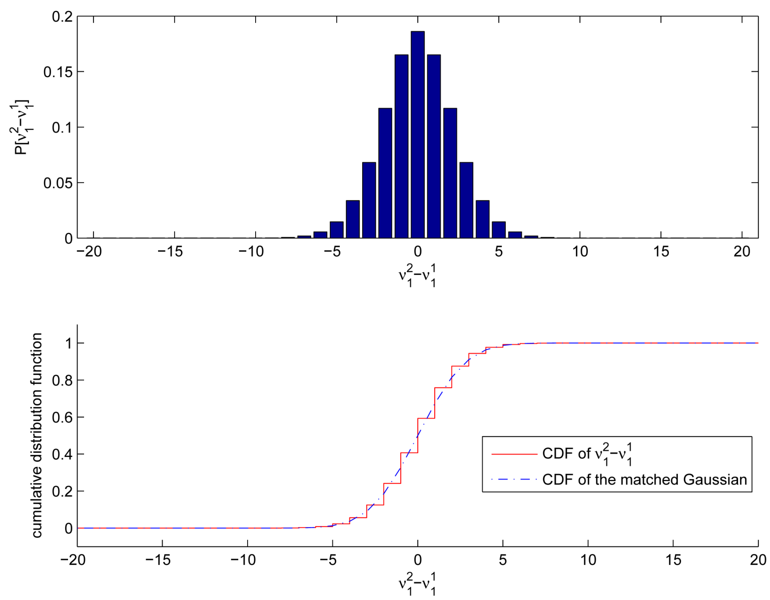


Fig. 3. The probability mass function of the approximate local classifier sufficient statistic $(\nu_1^2 - \nu_1^1)$ and the corresponding moment-matched normal probability distribution under the "same class" hypothesis (likelihood function of H_1). The 95% attribute gate for this difference is $[-4, 4]$.

to allow 5% missed detection of H_1 is $[-4, 4]$. This is much simpler than the class gate given in Figure 2.

The log-likelihood ratio surface of the same class (H_1) vs. the different classes hypothesis (H_0)—the ratio (28)—is shown in Figure 4. The prior for the classes is taken here as uniform. The decision region based on the *likelihood ratio* (28) to allow $\alpha = 5\%$ missed detection of H_1 is shown in Figure 5.

Comparison of the various tests

By numerically evaluating the pmf of the likelihood ratio statistic (37) under H_0 in the acceptance region for H_1, we find the false alarm probability (acceptance of H_1 when H_0 is true) as 8.1×10^{-8}. Compared with the decision region shown in Figure 2, which yields a false alarm probability of 1.9×10^{-7}, the likelihood ratio test has only a marginal performance gain in this case. Interestingly, the decision region based on the difference of the local classification statistics yields a false alarm probability of 9.1×10^{-5}, which is higher than both of the above, but still quite small.

7. SUMMARY AND CONCLUSIONS

The likelihood function and likelihood ratio based tests for track-to-track association using classification

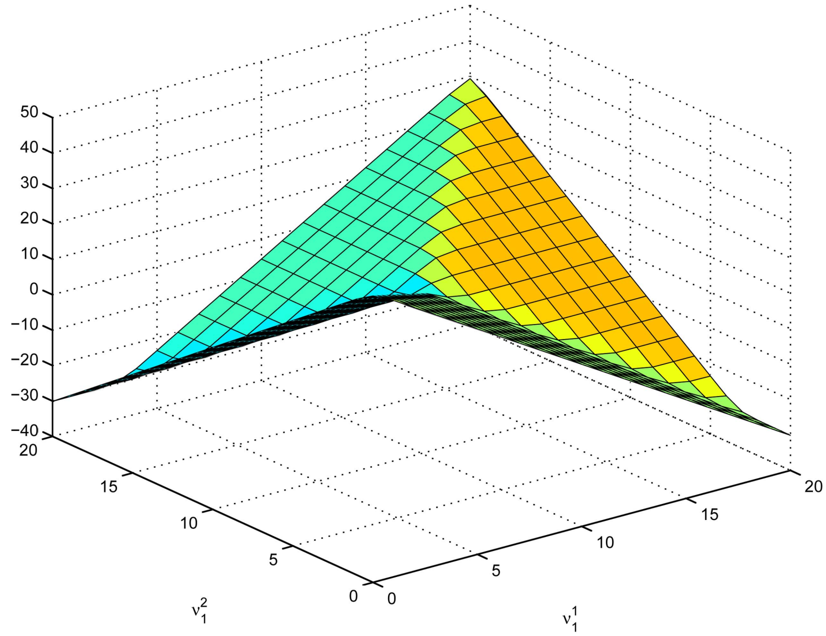


Fig. 4. The surface of the log-likelihood ratio between the same class (H_1) and different classes (H_0) vs. the local classifier sufficient statistics (ν_1^1, ν_1^2). Note the two peaks, one for (ν_1^1, ν_1^2) both large, one for both small.

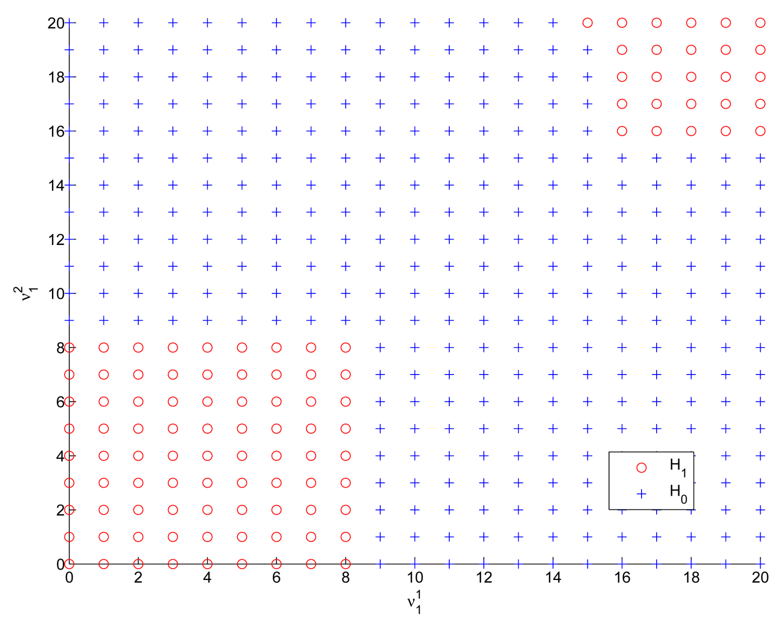


Fig. 5. Decision region for the likelihood ratio test with 5% probability of incorrectly rejecting H_1. This is the 95% attribute gate in the (ν_1^1, ν_1^2) space.

information have been derived and the means for evaluating their performance have been presented and illustrated. The sufficient statistic for the optimal association test in the Neyman-Pearson sense was obtained and its relationship with the class probability vector was discussed. The likelihood ratio test does not appear to have significant advantage over the test based on the likelihood function. The simplest test, based on the difference of the local sufficient statistics—similar to the test for continuous valued state estimates—yields a false alarm probability which is higher than for both of the above tests, but still quite small. The generalization to different and possibly time varying confusion matrices is a topic for future investigation.

Acknowledgments

Stimulating discussions with Drs. John Edwards, Oliver E. Drummond and J. Kent Haspert from the SIAP TCAIET task force are gratefully acknowledged. Thanks to an anonymous reviewer who pointed out that Eq. (27) was incorrectly written in the original version.

REFERENCES

[1] Y. Bar-Shalom and X. R. Li
Multitarget-Multisensor Tracking: Principles and Techniques.
YBS Publishing, 1995.

[2] Y. Bar-Shalom and W. D. Blair (Eds.)
Multitarget-Multisensor Tracking: Applications and Advances, vol. III.
Artech House, 2000.

[3] Y. Bar-Shalom, X. R. Li and T. Kirubarajan
Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction.
Wiley, 2001.

[4] Y. Bar-Shalom, T. Kirubarajan and C. Gokberk
Tracking with classification-aided multiframe data association.
IEEE Transactions on Aerospace and Electronic Systems, **41**, 3 (July 2005), 868–878.

[5] Y. Bar-Shalom and H. Chen
Multisensor track-to-track association for tracks with dependent errors.
Journal of Advances in Information Fusion, **1**, 1 (July 2006), 3–15. Also in *Proceedings of 43rd IEEE Conference on Decision and Control*, Nassau, Bahamas, Dec. 2004.

[6] S. Blackman and R. Popoli
Design and Analysis of Modern Tracking Systems.
Artech House, 1999.

[7] H. Chen, T. Kirubarajan and Y. Bar-Shalom
Performance limits of track-to-track fusion vs. centralized estimation.
IEEE Transactions on Aerospace and Electronic Systems, **39**, 2 (Apr. 2003), 386–400.

[8] C.-Y. Chong and S. Mori
Metrics for feature-aided track association.
In *Proceedings of 9th International Conference on Information Fusion*, Florence Italy, July 2006.

[9] O. E. Drummond
On features and attributes in multisensor multitarget tracking.
In *Proceedings of 2nd International Conference on Information Fusion*, Sunnyvale, CA, July 1999.

[10] O. E. Drummond
Integration of features and attributes into target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 4048, Orlando, FL, Apr. 2000.

[11] O. E. Drummond
Feature, attribute and classification aided target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 4473, San Diego, CA, Aug. 2001.

[12] O. E. Drummond
On categorical feature aided target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5204, San Diego, CA, Aug. 2003.

[13] O. E. Drummond
Tracking and classification with attribute data from legacy sensors.
In *Proceedings of ONR/GTRI Workshop on Tracking*, Key West, FL, June 2004.

[14] J. Ferry and L. D. Stone
Track-to-track association with missing features.
In *Proceedings of MSS Sensor and Data Fusion*, Monterey, CA, May 2005.

[15] J. Ferry
XMAP: Track to track association with metric, feature, and target type data.
In *Proceedings of 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[16] P. R. Halmos
Finite-Dimensional Vector Spaces.
Van Nostrand, 1958.

[17] J. K. Haspert, J. F. Heagy and R. J. Sullivan
General approach to template based target recognition.
IEEE Systems Magazine, **22**, 9, Part 2, (Sept. 2007), 53–67.

[18] X. D. Lin, Y. Bar-Shalom and T. Kirubarajan
Multisensor bias estimation using local tracks without a priori association.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5204, paper #70, San Diego, CA, Aug. 2003.

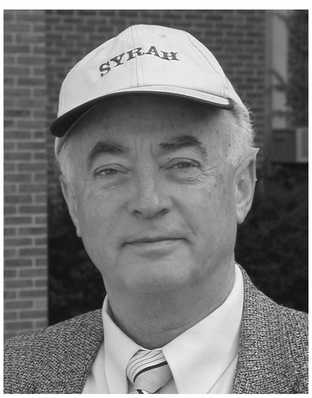
[19] S. Mori and C.-Y. Chong
Effects of unpaired objects and sensor biases on track-to-track association: Problems and solutions.
In *Proceedings of MSS National Symposium on Sensor and Data Fusion*, vol. 1, 2000, 137–151.

[20] S. Mori and C.-Y. Chong
Evaluation of data association hypotheses: Non-Poisson IID cases.
In *Proceedings of 7th International Conference on Information Fusion*, Stockholm, Sweden, July 2004.

[21] E. Parzen
Modern Probability and Its Applications.
Wiley, 1960.

[22] B. J. Slocumb and M. E. Klusman
A multiple model SNR/RCS likelihood ratio score for radar-based feature-aided tracking.
In *Proceedings of SPIE, Conference on Signal and Data Processing of Small Targets*, vol. 5913, San Diego, CA, Aug. 2005.

[23] L. D. Stone, M. L. Williams and T. M. Tran
Track to track association and bias removal.
In *Proceedings of SPIE Conference Signal and Data Processing of Small Targets*, vol. 4728, Orlando, FL, Apr. 2002.



Yaakov Bar-Shalom was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1970, all in electrical engineering.

From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, CA. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and Marianne E. Klewin Professor in Engineering. He is also director of the ESP Lab (Estimation and Signal Processing) at the University of Connecticut. His research interests are in estimation theory and stochastic adaptive control and he has published over 360 papers and book chapters in these areas. In view of the causality principle between the given name of a person (in this case, "(he) will track," in the modern version of the original language of the Bible) and the profession of this person, his interests have focused on tracking.

He coauthored the monograph *Tracking and Data Association* (Academic Press, 1988), the graduate text *Estimation with Applications to Tracking and Navigation* (Wiley, 2001), the text *Multitarget-Multisensor Tracking: Principles and Techniques* (YBS Publishing, 1995), and edited the books *Multitarget-Multisensor Tracking: Applications and Advances* (Artech House, Vol. I 1990; Vol. II 1992, Vol. III 2000). He has been elected Fellow of IEEE for "contributions to the theory of stochastic systems and of multitarget tracking." He has been consulting to numerous companies, and originated the series of Multitarget Tracking and Multisensor Data Fusion short courses offered at Government Laboratories, private companies, and overseas.

During 1976 and 1977 he served as associate editor of the *IEEE Transactions on Automatic Control* and from 1978 to 1981 as associate editor of *Automatica*. He was program chairman of the 1982 American Control Conference, general chairman of the 1985 ACC, and cochairman of the 1989 IEEE International Conference on Control and Applications. During 1983–1987 he served as chairman of the Conference Activities Board of the IEEE Control Systems Society and during 1987–1989 was a member of the Board of Governors of the IEEE CSS. Currently he is a member of the Board of Directors of the International Society of Information Fusion and served as its Y2K and Y2K2 President. In 1987 he received the IEEE CSS distinguished Member Award. Since 1995 he is a distinguished lecturer of the IEEE AESS. He is corecipient of the M. Barry Carlton Awards for the best paper in the *IEEE Transactions on Aerospace and Electronic Systems* in 1995 and 2000, and received the 1998 University of Connecticut AAUP Excellence Award for Research, the 2002 J. Mignona Data Fusion Award from the DoD JDL Data Fusion Group, and the 2008 IEEE D. J. Picard Medal for Radar Technologies and Applications.

Huimin Chen received the B.E. and M.E. degrees from Department of Automation, Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, in 2002, all in electrical engineering.

He was a post doctorate research associate at the Physics and Astronomy Department, University of California, Los Angeles, and a visiting researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University in 2002 where his research focus was on weak signal detection for single electron spin microscopy. He joined the Department of Electrical Engineering, University of New Orleans in January 2003 as an assistant professor. His research interests are in general areas of signal processing, estimation theory, and information theory with applications to target detection and target tracking.