18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

On Track-to-Track Data Association for Automotive Sensor Fusion

Bharanidhar Duraisamy
Group Research and Development
Daimler AG
Ulm, Germany
Email: bharanidhar.duraisamy@daimler.com

Tilo Schwarz
Group Research and Development
Daimler AG
Ulm, Germany
Email: tilo.schwarz@daimler.com

Christian Wöhler
Image Analysis Group
TU Dortmund University
Dortmund, Germany
Email: christian.woehler@tu-dortmund.de

***Abstract*—When fusing data from more than one information source, it is important to associate the correct pair of the data available from the information sources to achieve an optimal fused result. The responsibility of the task of proper association lies on the data association method used in the system. The design of the data association method has to be done considering the requirements of the application, quality and quantity of the information provided by the sensors. Parameters such as: whether the fusion is carried out in a centralized or decentralized architecture, whether the data available from the sensors is raw and unprocessed data or already processed by the built-in signal processing system of the sensor, plays a role in the finalization of the data association problem. The data association problem in a decentralized sensor fusion setting, also known as track-to-track association, is discussed in detail in this paper. The information sources used in this paper are environment perception sensors based on different measurement principles used for automotive safety and autonomous driving functions. The sensors deliver kinematic and as well as non-kinematic information on the tracked targets. To make use of this non-kinematic target information, attribute based association methods in-addition to the traditional data association methods and results based on the real world data are presented in this paper. A video recorded under real world test conditions that include sensor data and results will be made available for the community.**

I. Introduction

The sensor fusion system considered in this paper is designed for advanced driver assistance and safety systems (ADAS) such as adaptive cruise control, automatic emergency braking, crossing assistant and so forth. The ADAS in addition to passive safety systems such as airbags and electronic stability systems has played a great role in reducing the number of road accidents and fatalities. The United Nations report mentioned in [1] estimates that the loss of life and material by 2020 is to be around 1.9 million road deaths and $100 billion respectively. The ADAS has evolved through several generations and have now reached a maturity level. In [2], one can find a detailed study on the evolution of the ADAS, their impact on the society and the future technological focus in this area. As the technology progresses, consumer grade passenger and load-carrying vehicles are destined to become autonomous in the near future [3]. A detailed system level overview of a modern autonomous vehicle can be found in [4]. An autonomous vehicle has to reach its destination without involving in any kind of road accident and without causing any difficulty to any vulnerable road users (VRU) like pedestrians and bicyclists. Such an autonomous vehicle requires input from several information sources such as exteroceptive sensors, proprioceptive sensors, high precision maps and information received from the available vehicular networks.

The input from these information sources has to be combined to form a meaningful and accurate information to enable an autonomous vehicle to achieve its goal of accident free and comfortable driving. The sensor fusion module carries a huge impact on this pivotal role. The design of a sensor fusion module is based on the granularity of the individual sensor data available to the fusion module. A centralized sensor fusion module is beneficial and possible only when low-level signals from the sensors are available at the sensor fusion module. However, availability of such signals from the sensors is dependent on the application area. A sensor fusion architecture in which sensors have their in-built signal and data processing units that have capacity to deliver processed, time tracked signals to the fusion module is called distributed sensor fusion architecture.

This architecture is favoured for application areas where communication bandwidth and computation capabilities are a big constraint. The computation and networking elements available for automotive applications are many orders slower and smaller in capability than the consumer grade counterparts. Furthermore networks and computing elements used for automotive safety applications have to be safety level compliant, certified and cost effective. Due to the increasing usage of many high resolution environment perception sensors for automotive safety applications, the automotive computing and network elements are insufficient to handle the requirements of a low level fusion architecture. A parallel computing architecture or an application specific computing element such as FPGA has to be designed to meet the computational requirements. A detailed review on automotive networks can be found in [5]. An analysis of different distributed sensor fusion architectures can be found in [6] and a study of different distributed sensor fusion algorithms in the field of automotive applications can be found in [7]. This paper focuses on the subject of track-to-track association, which is a problem of forming and combining suitable pairs of tracks received from different sensors at the

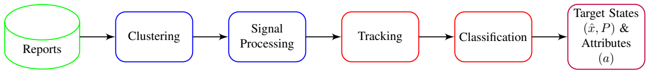

Fig. 1. **Sensor signal flow:** The signal flow chain of a typical automotive intelligent sensor is depicted in this figure. The raw reports received by the sensor are clustered and processed using appropriate thresholds and signal processing algorithms. The outcome of this process is then passed on to the tracking module, which is usually an unimodal belief tracking algorithm with an appropriate algorithm for relating the measurements to tracks and track management rules. The tracking module then pass on its results to the classification algorithm, which classifies the type of the target and other associated classification tasks.

sensor fusion module. This paper focuses on an extension to this track-to-track association (T2TA) problem that uses additional non-kinematic information of the targets delivered by the sensors and this algorithm's application in a sensor fusion layer that receives reports from several environment perception sensors used in the automotive area.

This paper is organized into following sections. The track-to-track data association problem is explained in section (II). Section (III) presents the state of the art solutions for this problem. Derivation of T2TA based on track kinematics is presented in (IV). The challenges and the factors influencing the T2TA is detailed in section (V). T2TA based on track kinematics and track attributes delivered by the sensors is described in (VI). The evaluation of the new method using real world data is carried out in section(VIII). The discussion on the concept presented in this paper and the closing remarks is presented in section (IX).

II. Problem Formulation

Let us consider that m sensors are present in the system. Each of these m sensors can deliver up to n tracked targets. It is not necessary that the number of targets delivered by all the sensors must be the same. The sensors have an overlapping field of view (FOV) or surveillance region as depicted in the Fig. 2. The sensors are equipped with in-built signal processing and target tracking systems. They process the raw sensor reports with-in the system and deliver processed tracks with their respective kinematic information and also some additional information regarding the targets. This additional information is known as attributes. Attributes usually represent some characteristics of the target based on the measurements observed by a particular sensor. They are not always directly measurable and they can be from either a continuous or discrete space. A very detailed study on the types of attributes and their difference with target features is done in [8]. In automotive setting attributes are based on the following target and sensor characteristics:

- **Target physical properties:** Features such as Radar Cross Section (RCS), Length, Width, Target Category.
- **Sensor track management:** Track Age, Id, Score, Quality.

The sensors can be based on different physical measurement principles but they deliver the tracks in a homogenous state space. The tracking module in the sensors assume that targets

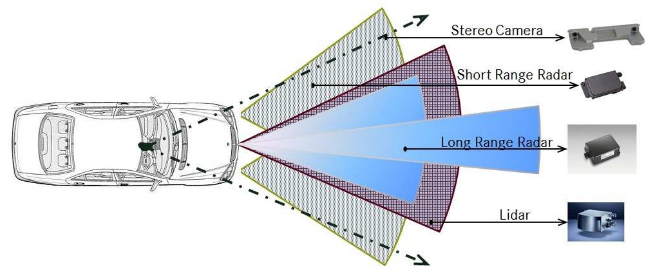

Fig. 2. **Sensor Platform:** The ADAS sensor platform considered here is a Mercedes Benz E class car equipped with different kinds of radar, laser and camera based environment perception sensors. The sensors have an overlapping Field of View (FOV) of the environment to achieve required information refinement and redundancies to ensure a safe operation. The sensor platform is usually equipped with a combination of sensors mentioned in the figure, in all the sides of the car to collect a surround 360 view of the environment. The exact configuration and FOV parameters used for a Mercedes Benz autonomous vehicle can be found in [4].

[9] follow a particular motion and time update model (1) and a suitable observation model as in equation (2).

$$\hat{x}(k+1) = f\left(\hat{x}\left(k\right), \Delta t\right) + v(k) \tag{1}$$

$$z(k) = h\left(\hat{x}(k)\right) + w(k) \tag{2}$$

The k in the equations indicates the index, v and w is process and measurement noise respectively. Different sensors are represented using the sub indexes i and j. The sensors can operate at different frequencies. At every frame a set of tracks $X = \{\hat{x}_i, P_i\}$ and a group of attributes $A = \{a_{y,i}\}$ with $i = 1 \cdots n$ and $y = 1 \cdots d$, where d represents number of different kinds of attributes associated with the track is delivered.

When two sensors deliver their set of tracks, it is necessary to form the correct pair of tracks. The objective of the track association task is to find the correct pair of tracks and to ascertain that given pair of tracks belong to the same true target. The *common origin* hypothesis test known as Normalized Distance Squared (NDS) is used to make this decision. This test is carried out using only the kinematic information of the tracks. This problem of associating tracks from different sensors with each other can also be posed as a bijective mapping [10] between tracks originating from sensor i to sensor j as described in equation (3).

$$H^* = \overset{H \in \mathcal{H}}{arg\ max} \left\{ \frac{Pr(H_1|X)}{Pr(H_0|X} \right\} \tag{3}$$

The H_1 and H_0 in equation (3) represent a set of tracks originated from the same true target and group of tracks originated from different other true targets respectively.

$$\Lambda^{ij}_{combined} = \Lambda^{ij}_{kin} . \Lambda^{ij}_{A} \tag{4}$$

Given two kinematic states $\hat{x}_i, \hat{x}_j$ and their respective non kinematic attributes A_i, A_j, the objective of this paper is to calculate Λ^{ij}_{A} using the attributes taking into account the involved uncertainties, Λ^{ij}_{kin} using the kinematic states and to calculate a joint cost function [11] as in equation (4). The requirement is to derive the combined cost function that could improve the T2TA performance by taking into account one or more attribute based gates in addition to the well established kinematic gate. In this paper, combination of non-kinematic information such as target category, target width and a complete kinematic state information is considered.

$$LRT_{i_{cat}=j_{cat}} = \frac{\Lambda^{cat}_{i=j}}{\Lambda^{cat}_{i\neq j}} \tag{5}$$

Provided N target categories and a set of objects i and j from two sensors I and J the target category attribute gate has to provide a decision whether the objects from the two sensors belong to the same category or not. This is handled through a likelihood ratio based test [9] as in equation (5). The same category likelihood functions $\Lambda^{cat}_{i=j}$ and $\Lambda^{cat}_{i\neq j}$ have to take into account the uncertainty in the decision of the sensor's local classifier using appropriate confusion matrices.

III. State of the Art / Present Work

An overview of track-to-track fusion in the context of automotive safety applications is presented in [7] and the publication presents that ignoring estimation error cross co-variances among the sensors is fatal to the fusion performance. The estimation error cross covariance also has an influence on T2TA performance and this phenomenon in case of single and multi-frame setting is discussed in [12]. Traditionally data association such as measurement to track association (M2TA) and track-to-track association (T2TA) is carried out using a χ^2 test based Mahalanobis metric. This paper presents only a selected set of papers that are relevant to T2TA with and without non-kinematic attribute information. In practice to carry out T2TA a cost look up table is built for $track_i$-to-$track_j$ pairings. An important function is the selection of a pair with the least cost. Algorithms such as Murthy's algorithm are applied to the 2-D assignment problem and also several variants of auction algorithms can be used in a nearest neighbour setting to solve this minimum cost search function. This cost function selection becomes more complex for a $m-2D$ assignment problem. A good survey on several assignment techniques can be found in [13]. A very detailed likelihood calculation for a multi sensor T2TA is published in [14]. In [14] we can find the derivation of likelihood functions based on kinematic distance and the respective track error cross covariance taken into consideration and approximations for track error cross covariance compensation is also presented. A detailed study on the trade-off between performance and computation requirements to calculate T2TA probabilities for a two sensor setting using hypotheses and Monte Carlo methods is presented in [10].

In the area of attribute augmented T2TA the foremost work is [11], which discusses in detail the systematic procedure to integrate features and attributes into target tracking. A detailed study on integrating target categorical information and different conditions of feature aided tracking based on the information and apriori availability is published in [8]. In [15] a detailed procedure to derive metrics for non-kinematic information aided track association is published. The metric is based on Maxium A Posteriori (MAP) probability approach. An extended MAP approach, which extends the method by including the detection probability and joint probability of a feature and the target type for a metric calculation by using a non-informative prior for the feature measurement model can be found in [16]. In [17] a derivation for association based on non-kinematic information and the Neyman-Pearson test function for this case is presented. An interesting publication of using state augmentation with continuous features to perform T2TA is published in [18] and it is found that they significantly improve the association performance but the performance of this state augmented T2TA is dependent on the accurate estimation of the augmented feature. Several analytical expressions are presented to predict performance of T2TA in combination with different non-kinematic discrete or continuous feature states is presented in [19].

IV. Kinematic T2TA

This section describes the derivation of T2TA based on kinematic information of sensor tracks. The association of the track pair is carried out using only the kinematic state components. This association is carried out using a hypothesis test as formulated in the section (II). The basic necessary assumption to carry out this test is that the sensor track estimation errors are Gaussian distributed and the sensors have a homogenous state space (under real-world conditions this assumption may become less strictly valid). The test is carried out by comparing the track states delivered by the sensors. It is necessary to time-synchronize the sensor tracks. Usually this is handled through a procedure called *union of sampling times* as explained in [20]. Following the derivation in [9], when two tracks or one track from two sensors i, j originate from a *same target* at a given time k then the true states are equal and this leads to the following equation

$$\Delta_{ij}(k) = \hat{x}_i(k) - \hat{x}_j(k) = 0 \tag{6}$$

The difference in state estimation error is given by the equation (7)

$$\tilde{\Delta}_{ij}(k) = \tilde{x}_i(k) - \tilde{x}_j(k) \tag{7}$$

The terms $\tilde{x}_i$ and $\tilde{x}_j$ denote sensor track estimation errors. The corresponding covariance T_{ij} corresponds to

$$T_{ij}(k) = E\left\{[\tilde{x}_i(k) - \tilde{x}_j(k)]\,[\tilde{x}_i(k) - \tilde{x}_j(k)]'\right\} \tag{8}$$

which can be reformulated as

$$T_{ij}(k) = \{P_i(k) + P_j(k) - P_{ij}(k) - P_{ji}(k)\} \tag{9}$$

The terms P_i, P_j and P_{ij} are the track covariance of sensor i, track covariance of sensor j and the crosscovariance of the tracks i and j respectively [9]. A Lyapunov type recursion for the P_{ij} crosscovariance is derived in [9], where the crosscovariance term is included in [9] to correct the effect of common process noise and to reduce the error covariance. In our experiments, we have assumed that the track estimation error in equation (7) is zero - mean Gaussian distributed and this assumption also holds for the corresponding estimation error covariance. It is possible to calculate the assignment cost using any one of the following methods.

A. Normalized Distance Squared

Given a random number and a distribution, their statistical closeness can be found using a metric called *Mahalanobis Distance* [21]. A similar metric can be used for T2TA under assumption that the tracks from the two sensors i and j are Gaussian distributed and so their error distribution is based on χ^2 distribution. A χ^2 test [21] is carried out whether the tracks fulfill the following same target hypothesis:

$$D_{ij}(k) = \Delta'_{ij}(k)\,[T_{ij}(k)]^{-1}\,\Delta_{ij}(k) \tag{10}$$

If the squared norm in the equation (10) lies with in the $1-\alpha$ region of the χ^2 distribution then the tracks belong to the *same true kinematic state* or *common origin*. This condition is described in the following equation:

$$D_{ij} \leq \chi^2(1-\alpha) \tag{11}$$

The known dimension of the state space n_x is used to define the degrees of freedom of the χ^2 test. The alpha is mostly chosen as 95%. The interpretation of the equations (10) and (11) is quite straightforward. The assignment meteric D_{ij} obtained in (11) is rejected, if the difference between the two sensors tracks is large.

B. Negative Log Likelihood Ratio

A dimensionless negative log likelihood ratio (NLLR) test can be carried out using equations similar to (6) — (11). According to [12], the kinematic state likelihood function for the Gaussian distributed state estimation error is

$$\Lambda^{ij}_{kin}(k) = \frac{1}{V}\cdot\frac{1}{\sqrt{|2\pi T_{ij}(k)|}}\cdot exp\left[\Delta'_{ij}(k)\,[T_{ij}(k)]^{-1}\,\Delta_{ij}(k)\right] \tag{12}$$

The equation (12) forms the numerator ($H1$) of the NLLR, the denominator ($H0$) that represents that tracks belong to different origin is given by

$$\Lambda^{H0}(k) = \frac{\mu_{ex}}{V} \tag{13}$$

where μ_{ex} represents the spatial density of the extraneous tracks under the assumption that they are Poisson distributed in the state space and the true tracks are not homogeneously distributed. V represents the volume of state space, to denote that false tracks can be uniformly distributed. The NLLR test carried out using the equations (12) and (13) is

$$\mathcal{L}^{ij}_{kin}(k) = -\ln\frac{\Lambda^{ij}_{kin}(k)}{\Lambda^{H0}(k)} \tag{14}$$

$$= \frac{1}{2}\cdot\left[\Delta'_{ij}(k) + [T_{ij}(k)]^{-1} + \Delta_{ij}(k)\right] + \ln\left(\mu_{ex}\sqrt{|2\pi T_{ij}(k)|}\right) \tag{15}$$

For cases of complete assignment of tracks with different lengths [18], the costs can be designed based on the respective sensors detection probability P_d for the dummy or new tracks according to

$$C_{ij}(k) = \begin{cases} \mathcal{L}^{ij}_{kin}(k) & \text{if } i,j \neq 0 \\ -\ln(1-\mathrm{P}_{d1}) & \text{if } i=0,\ j\neq 0 \\ -\ln(1-\mathrm{P}_{d2}) & \text{if } i\neq 0,\ j=0 \end{cases} \tag{16}$$

The calculated cost C_{ij} is then used to form an appropriate cost matrix. The JVC, Munkers or Murthy's algorithm is used to make a decision that corresponds to the most probable association between tracks.

C. Maximum A Posteriori

This maximum a posteriori (MAP) based approach is an another approach to resolve T2TA. The performance of this method depends on the correctness of the prior models and the likelihood functions used in the procedure [15], [16]. According to [16], the following assumption on the probabilities are made to solve the MAP based T2TA:

- The prior distribution of the state corresponds to $p_0(x) = 1/Vol(\mathcal{S})$, where $\mathcal{S}$ denotes the target's region of space.
- The measurement likelihood function $\mathcal{L}^s_i = \mathcal{N}(z_i;\hat{x},P_i)$ of a sensor s is assumed to be Gaussian where x denotes the mean and P the covariance matrix.

The prior region $\mathcal{S}$ can be constructed using the reports from the sensors. It is shown in [16] that the center of the region is

$$\hat{x} = \frac{1}{N}\sum_{i=1}^{N} z_i \tag{17}$$

According to [16], in case that the sensor reports have only kinematic information then a suitable shape to prescribe the region $\mathcal{S}$ is an ellipsoid determined by the covariance P of the sensor reports z_i. It is also possible to calculate the covariance matrix P of the $\mathcal{S}$ by integrating all the points in the region. Once the covariance is calculated, then the volume of the region is given by

$$Vol(\mathcal{S}) = (2\pi\theta)^{n/2}\sqrt{det(P)} \tag{18}$$

with n as the dimensions of the data used. In our case the dimension of the state and the sensor report is equal. The θ is given by

$$\theta = \left(1+\frac{m}{2}\right)\left(\frac{m}{2}!\right)^{\frac{-2}{m}} \tag{19}$$

where m is chosen as divisor of n, leading to $\mathcal{S}$ being a Cartesian product of m-dimensional ellipsoids. As a rule of thumb, m equal to the dimensionality of the space of measurements [16]. As an example for a $2-D$ Cartesian problem with position and velocity components it is $m=2$ and $n=4$. A robust probability ratio that takes into account the the proper prior is

$$R^{ij}_{kin} = \sqrt{\frac{(|I+M_i|)\,(|I+M_j|)}{(|I+M_i+M_j|)}}\ . \tag{20}$$

$$exp\left(-\frac{1}{2}(\hat{x}_i-\hat{x}_j)^T(P_i+P_j)^{-1}(\hat{x}_i-\hat{x}_j)\right) \tag{21}$$

where M in equation (20) is given by $M_i = \theta(P_i)^{-1}P$. In [16] this equation is found to be robust against large values of the sensory track variances. The effectiveness of this ratio is dependent upon the adherence of the assumption that the data to be associated is uniformly distributed over a region and as an alternative to this assumption, fitting an uniform prior over the clustered data is recommended.

V. Factors in Track-to-Track Association

Many factors have to be taken in to account during the design phase of T2TA for any particular application. The availability of data from the sensor plays a vital role. The cross -correlation between the estimation error of different sensor tracks and the possibility of calculating them correctly at the fusion center has an impact in the calculation of association threshold. If the non-kinematic components are added to this mix then the following questions have to be defined:

- **Dependence:** The dependency between the track states and its attributes. It is advisable to know the impact of track kinematic state information used in the feature estimation or in the target classification algorithm of the sensor. Most classifiers usually depend upon the observed target features to do their feature estimation and classification. This is dependent on sensor and its feature estimation algorithms.
- **Models:** The models used for the likelihood estimation of the continuous and discrete features. It is important to know whether all non-kinematic components such as categories uses a same model for all true target categories or does each category require a separate model.
- **Number of models:** The number of kinematic and non-kinematic models used in the sensor local system and the availability of the model switching information at the fusion center.
- **Cross-Correlation:** The cross-correlation between the kinematic and non-kinematic state components. If a cross-correlation exists, then the enumeration of all possible combinations of kinematic and non-kinematic state components has to be made. This cross-correlation might also lead to creation of multiple tracks instead of resulting in an unique track set.
- **Time variant:** This is an important factor to consider for the non-kinematic state components. The information such as target categorical information on a target can change over time or it remains constant over time once it is declared. This factor has an influence in deciding the parameters to fuse the categorical information.

A more detailed taxonomy and classification of the operating conditions can be found in [8]. In practical applications, the availability of the following items have to be taken into consideration:

- Likelihoods: The innovated likelihood from the classifier and feature estimation algorithms for the non-kinematic information.

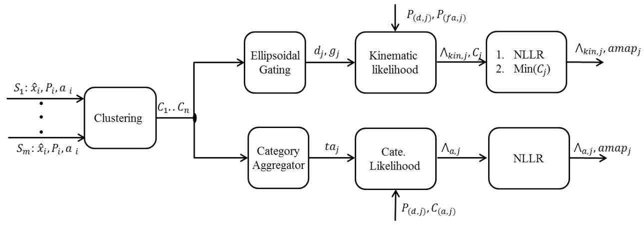

Fig. 3. **T2TA Architecture:** This association method calculates a NLLR based hypothesis test, which is calculated using the kinematic and non-kinematic information of the targets from different sensors to verify the common origin or same target hypothesis.

- Confusion matrices: The exact confusion matrix for the categorical information. This matrix has to be determined based on the classifier performance.
- Covariance matrices: In some cases the availability of the sensor track covariance matrices is itself restricted or not available due to some practical limitations such as communication bandwidth constraints. Even if the covariance is available, the consistency of the covariance delivered by the sensor often is questionable due to various systematic and stochastic errors present in the sensor local processing system.

These limitations can be worked around in practical applications but irrespective of these workarounds, there is a performance degradation in the T2TA. A solution for dealing with tracks with no covariance information can be found in [22] and some solutions for tracking and classification of attributes with incomplete information can be found in [23]. It has to be considered that a workaround proposed for a domain might not be suitable for all other domains however some hints can be taken from them.

VI. Attributes in Track-to-Track association

The attributes considered in this paper are categorical information only and they are discrete variables. The attributes and the kinematic state are independent. In case of continuous features such as radar cross section and discrete attributes generated using a known function, integration of this information in T2TA is straightforward [18]. There is a good solution published in [16], which presents the procedure for a joint feature and categorical component based T2TA method. This paper deals with the case when the fusion center has just the target kinematic information and target-type, which is also known as categorical non-kinematic information available.

The T2TA architecture depicted in the Fig. 3 is proposed in this paper. T2TA based on kinematic information and non-kinematic information requires at least two stages of gating procedure: 1. kinematic gate, 2. attribute gate. The tracks delivered by the different sensors $\mathcal{S}_k$, $k = 1 \dots m$ are segmented into several clusters C_j, $j = 1 \dots n$. The tracks in a cluster are fed-in separately to the ellipsoidal gating module and to the category aggregator. The kinematic gating results in the formation of a table of statistical distances d_j and gated pairs g_j. The gated pairs then pass through the kinematic likelihood function, which takes into account the detection probability P_d and false alarm probability P_{fa} of the respective sensors and calculates the association cost C_j and the association negative log likelihood $\Delta_{kin,j}$ for the *same origin* and different origin hypothesis. The cost function and the likelihood values are then used to calculate the $minimum$ cost for a particular cluster and to carry out the likelihood ratio test using kinematic components. The category aggregator aggregates the category identifiers of different sensor tracks present in the cluster, the aggregated table ta_j is then passed on to calculate the category likelihood function, which takes into account the detection probability P_d respective to the category and the sensor confusion matrix $\mathcal{C}_{(a,j)}$. The calculated likelihood function then can be used to carry out the likelihood ratio test for the *same category*. A unified cost function can be formed using the negative log likelihood function for kinematic and non-kinematic components and a suitable cost selection function can be used to select the appropriate set of tracks. The proposed architecture is a hard decision approach and it does not provide any association weights β_{ij} to derive a soft decision but this architecture can be extended to include a soft decision approach. The existence of statistical independence and sufficient statistics has to be investigated to formulate a soft decision approach.

It is better to understand the following condition before declaring that a track set is from the same origin. It is very likely that if the kinematic state component of two tracks are very *close* or in other words the statistical distance between them is very small then they are probably from the same true target whereas if the categories of a track set are the same then it does not imply that they are definitely from the same true target but there is a possibility based on the employed clustering algorithm. A hard decision based only on the target-type can be made only if it is an unique entity for the track set.

The following assumptions are made to solve this problem: The sensors in the system deliver m number of classes and their respective confusion matrices are time invariant. The target category is a finite set, which is represented by

$$tc \in TC = \{1, \dots, N_f\} \tag{22}$$

where N_f in equation (22) represents the total number of possible target categories. The target category information is the result of the local classifier in the sensor system. Refer to Fig. 1 for a commonly used signal flow in an automotive environment perception sensor system. The next step is to find out a procedure to properly interpret the sensor classifier's output. The classifier usually provides us a set of classified target categories as in the following equation,

$$\kappa \in TC_s = \{1, \dots, N_s\} \tag{23}$$

The associated confusion matrix is given by

$$c_{nm} = P\{tc = n \mid \kappa = m\} \tag{24}$$

where $n = 1, \dots, N_f$ and $m = 1, \dots, N_s$. The category likelihood [17] is represented by c_{nm}, which means the probability of the true target category being n when the sensor local classifier has tagged the observed target category as $\kappa = m$. The particular m category likelihood is the column m of the confusion matrix C. The errors of the local classifiers of the sensors are assumed to be white. But as mentioned in [17], the classifier outputs from different sensors will be correlated since they observe the same true target. The classifier correlation might not be a problem in case the local classifier present in the different sensors are based on entirely different physical observations. For instance, the classifier present in a camera system is based on image and depth features where as the local classifier of a radar would be most likely based on radar cross section and other similar features to carry out their classification. However the interdependence of the features of the classifiers and between the sensors would be present. According to [17], the updated probability of a target of category n, when the sensor local classifier has tagged it as category m is given by

$$\mu_n = P\{tc = n \mid \kappa = m\} = \frac{c_{nm}\mu_n^0}{\sum_{r=1}^{N_f} c_{rm}\mu_r^0} \tag{25}$$

and the category probability vector updated at any given time k can be written as

$$\mu(k) = P\left\{tc = n \mid \kappa(k) = m, \kappa^{k-1}\right\} = \frac{c_m \otimes \mu(k-1)}{c_m' \mu(k-1)} \tag{26}$$

In equation (26) the term c_m represents the m^{th} column of the confusion matrix C, κ^{k-1} denotes the classification information accumulated till time $k-1$ and μ^0 is the prior before updating with any new information concerning the target category. The symbol $\otimes$ represents term-by-term product. As described in detail in [17], this is a recursive calculation and the initialization at time step $k = 0$ is

$$\mu(k = 0) = \mu^0 \tag{27}$$

where the prior is a non-informative prior. The observations are innovated through the likelihood function c_m. The sufficient statistics of the classifier output is based on the number of times kn the classifier has tagged the target category m and the equation (26) can be expanded as

$$\mu(k) = \frac{1}{\alpha}\left[c_1^{kn_1} \otimes c_2^{kn_2} \otimes \dots \otimes c_{na}^{kn_{na}} \otimes \mu_0\right] \tag{28}$$

$$c_m^{[kn]} = \left[c_{lm}^{kn}\right]' \tag{29}$$

where $l = 1, \dots, N_s$ and the term c_m is calculated using the term by term product and it is raised to the power of number of times kn the classifier's decision is category m. The term α represents the normalizing constant. Under the assumption that confusion matrix is constant, the sufficient statistic for the sensor local classifier is given by the vector of number of times each category class was classified or tagged by the classifier.

It is shown in [17] that the probability mass function required to carry out the hypothesis test based on the same category information for track i from sensor I is given by

$$P[kn^i|\kappa^i = n] = P[kn^i, \ldots, kn^{N_s}] \tag{30}$$

$$= N^i! \prod_{m=1}^{N_s} \frac{c_{nm}^{kn_m^i}}{kn_m^i!} \tag{31}$$

The equation (30) is a multinomial distribution and this equation is the cumulative classifier information. The term N^i represents the total number of classifier outputs

$$\sum_{m=1}^{N_s} kn_m^i = N^i \tag{32}$$

Using the above equations, the likelihood function for the same class hypothesis $H1(C_{ij})$ for tracks i and j from two different sensors is

$$\begin{aligned}
\Lambda_{cat} C_{ij} &= P\left[kn^i, kn^j | C_{ij}\right] \\
&= \sum_{n=1}^{N_f} P\left[kn^i, kn^j | \kappa_i = \kappa_j = n\right] \mu_n^0 \\
&= \sum_{n=1}^{N_f} P\left[kn^i | \kappa_i = n\right] P\left[kn^j | \kappa_j = n\right] \mu_n^0 \\
&= \sum_{n=1}^{N_f} N^i! N^j! \left[\prod_{m=1}^{N_s} \frac{ci_{nm}^{kn^i} cj_{nm}^{kn^j}}{kn^i! kn^j!}\right] \mu^0 \\
&= \sum_{n=1}^{N_f} N^i! N^j! \left[\prod_{m=1}^{N_s} \frac{ci_{nm}^{kn^i + kn^j}}{kn^i! kn^j!}\right] \mu^0
\end{aligned} \tag{33}$$

In [17], the equation (33) is derived under the assumption of temporal independence of the classification errors and the sensors I and J local classifiers are not dependent on each other. It has to be noted that proper prior μ^0 is used. An non-informative uniform prior is preferred. The confusion matrix c_{nm} can be same for the two sensors. If required the equation can be adopted to have different confusion matrices ci and cj for the sensors respectively. This is a bit different to kinematic based T2TA, which uses the total probability theorem to diffuse the improper prior [9].

If the track set i, j lies in the tail of the distribution of the category pmf $H1$ then they are not from the same origin. But such a gating procedure is not possible due to the fact that the difference of two different discrete attribute vectors is not the exact sufficient statistic as calculated in the NDS for kinematic state components. There is no straight forward solution available to find out the $H1$ rejection tail region as in continuous valued kinematic states. A tail of the likelihood pmf can be found using exhaustive calculation on all its point masses. This procedure is initially derived in [9].

An alternative approach described in [24] is to derive the $H0$ to carry out the likelihood ratio test using the $H1$ and $H0$ hypotheses instead of finding out the tail region of the $H1$ pmf. The next step is to derive the likelihood function for the *not same category* hypothesis $H0(C_{i \neq j})$, which means the sensor tracks i and j belong to *different true target category*. This equation is derived using total probability theorem and it is given by

$$\begin{aligned}
\Lambda_{cat} C_{i \neq j} &= P\left[kn^i, kn^j | C_{i \neq j}\right] \\
&= \sum_{n=1}^{N_f} \sum_{r=1, r \neq n}^{N_s} P\left[kn^i, kn^j | \kappa^i = n \neq k^j = r\right] \mu_n^0 \mu_0^r \\
&= \sum_{n=1}^{N_f} \sum_{r=1, r \neq n}^{N_s} N^i! N^j! \left[\prod_{m=1}^{N_s} \frac{c_{nm}^{kn_m^i} c_{rm}^{kn_m^j}}{kn_m^i! kn_m^j!}\right] \mu_n^0 \mu_r^0
\end{aligned} \tag{34}$$

$$LRT_{cat}\left(\frac{C_{ij}}{C_{i \neq j}}\right) = \frac{\Lambda_{cat} C_{ij}}{\Lambda_{cat} C_{i \neq j}} \tag{35}$$

The likelihood ratio test, which describes the test statistic based only on the target category information can be carried out using the equations (33) and (34). The acceptance region lies in the $1 - \alpha$ probablity region of the hypothesis $H1$.

VII. T2TA using Combined Kinematic and Category Components

In sections IV and VI T2TA likelihood calculation based on kinematic state components using *kinematic gate* and non-kinematic information such as target category information based *attribute (category) gate* is presented. A combined likelihood ratio: kinematics + non-kinematic (category) is presented here. They are not integrated as a joint function but as independent factors, however a joint pdf is possible similar to the XMAP method [16] provided a proper model for the considered features can be formulated.

$$\Lambda_C^{ij}(k) = \Lambda_{kin}^{ij}(k) \Lambda_{cat}^{ij}(k) \tag{36}$$

The equation (36) represents the joint likelihood function calculated using kinematic and non-kinematic (target category) likelihood functions under the *independence* assumption and this equation can be simplified using NLLR

$$C^{ij}(k) = \begin{cases} \mathcal{L}_{kin}^{ij}(k) + \mathcal{L}_{cat}^{ij}(k) & \text{if } \ i, j \neq 0 \\ -\ln(1 - \mathrm{P}_{d1}) & \text{if } \ i = 0, \ j \neq 0 \\ -\ln(1 - \mathrm{P}_{d2}) & \text{if } \ i \neq 0, \ j = 0 \end{cases} \tag{37}$$

The cost function has to be in negative log likelihood and it also reduces the arithmetic complexity of the likelihood functions. The terms $\mathcal{L}_{kin}^{ij}(k)$ and $\mathcal{L}_{cat}^{ij}(k)$ represents NLLR for kinematic and target category at time k respectively. Once the cost $C^{ij}(k)$ is calculated, then an algorithm based on 2-D assignment cost optimization such as Auction or Murthy's, can be used according to the application requirements. This is a hard decision based approach. An association map $aMap_{ij}$ function is defined, which retains the decision on the association of tracks from sensor I and tracks from sensor J.

$$aMap_{ij}(k) = \begin{cases} 1 & \text{if track } i \text{ (sensor} I\text{), track } j \text{ (sensor } J\text{)} \\ & \text{correspond to the same true target and} \\ & \text{allocated for fusion} \\ 0 & \text{Otherwise} \end{cases} \tag{38}$$

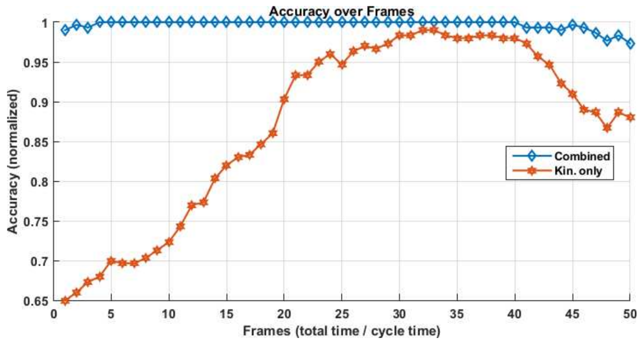

Fig. 4. **Accuracy:** The accuracy of the different T2TA methods are presented here. The blue line with diamond markers represent the combined T2TA method that considers kinematic and non-kinematic information for the calculation of association metric, the red line with the hexagonal markers represent the T2TA method that considers only the kinematic information for the calculation of association metric.

$$AD(k) = min \sum_{d=0}^{N_I} \sum_{l=0}^{N_J} aMap_{dl}(k) C_{dl}(k) \qquad (39)$$

Now the equations (37) and (38) can be used to form the min. 2-D assignment cost optimization function (39) to calculate the final association decision AD at time k, the terms N_I and N_J represents the number of tracks from sensor I and sensor J respectively. A dummy track $d, l = 0$ is added to handle the incomplete assignments. It is possible to include other continuous target features and discrete track id's in this T2TA combination. It has to be noted that proper models, templates and confusion matrices have to be designed to integrate them in the existing T2TA architecture. A 2-D assignment problem is presented here and it is scalable for $S > 2$ sensors with different number of tracks but with increasing combinatorial complexity.

VIII. DISCUSSION

The performance of the algorithm is analyzed using Monte Carlo simulation and data from real world environment perception sensors such as radar and lidar. The parameters for simulation is chosen to reflect the near real world conditions. The settings used in the simulation are explained as follows,

- Three true trajectories are generated based on the constant velocity model with a known acceleration power spectral density. The trajectories are designed in such a way that they resemble a trajectory of an automobile in a highway.
- Two sensors that adheres to the uncertainties of real world sensors are simulated. The number of targets and measurements generated by the sensors are identical and each sensor generates z measurements, which are then processed in a built-in Kalman filter of the sensor to generate n targets.
- A unity P_d is assumed. The sensors have a false alarm rate of 1%, the number of false alarms are Poisson distributed and they are uniformly distributed over the state space. The cycle rate and the state space of the sensors are designed to be homogenous. The tracking is carried out in a Cartesian coordinate system.
- A matched process noise is chosen to avoid model mismatch errors. The tracking module of the sensors and the fusion center uses a DCWNA process model [9]. The measurement to track association at the sensor's tracking module is assumed to be perfect without any mis-association errors.
- The sensors at each cycle deliver kinematic information in adherence to the state space, the non-kinematic information such as estimated width of the target and the target category such as car, truck and pedestrian.
- The target category information of the simulated true targets is designed to be ambiguous. Two of the three simulated true targets have the same target category and the remaining other target has a different target category. The non-kinematic target category information is uniformly initialized with an non-informative prior. The likelihood of the continuous feature like target width information can be handled using the standard χ^2 test.

One unique confusion matrix for the target category can be used in the fusion center, however two different confusion matrices are chosen for this simulation and also for the real world data. The confusion matrices of the sensor 1 is M_{s1} and sensor 2 is M_{s2}, they are as given below:

$$M_{s1} = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.10 & 0.85 & 0.05 \\ 0.10 & 0.05 & 0.85 \end{bmatrix} M_{s2} = \begin{bmatrix} 0.75 & 0.10 & 0.15 \\ 0.10 & 0.75 & 0.15 \\ 0.10 & 0.15 & 0.75 \end{bmatrix}$$

The performance of the T2TA methods can be evaluated similar to the performance evaluation of pattern classification algorithms. Information on the following five cases are collected:

- True Positive Correct Association (TP_c): Number of targets that really exists and are correctly associated.
- True Positive False Association (TP_f): Number of targets that really exists but their association is incorrect.
- False Positive (FP): Number of targets that does not exist in reality but they are detected and associated in the fusion center.
- False Negative (FN): Number of targets that are flagged as non existing even though they exist in reality. These are the targets that were not detected by both the sensors or they got mis-associated and terminated by the track manager of the sensor.
- True Negative (TN): Number of targets that in reality does not exist. They are correctly not reported by the sensors and no association is carried out based on them in the fusion center.

The above collected data are used to calculate the miss rate (MR) and accuracy (ACC) of the different T2TA methods defined in [25] as

$$MR = \frac{TP_f + FN}{TP_f + FN + TP_c} \qquad (40)$$

An accuracy of 1 represents 100% correct association and likewise, a miss rate of 0 means no false associations are made.

$$ACC = \frac{TP_c + TN}{N} \tag{41}$$

The ACC values observed for a 500 Monte Carlo simulation runs are presented in the Fig. 4. The blue curve labeled as "combined" with the diamond markers represents the association method that combines all the kinematic and non-kinematic information to make the association decision is *more accurate* than the association method that uses only the kinematic information to derive the association decision, which is the red curve labeled as "kin. only" with the hexagonal markers. The reason for the lower ACC of the kinematic only method is due to the larger uncertainty of the sensor local tracks until these tracks converge to the steady state. The ACC of the kinematic only T2TA method increases as the convergence increases. The kinematic only metric leads to mis-association with the neighboring tracks, whereas this problem is the avoided by using non-kinematic (target category and target width) information to calculate the final association cost. This additional information provides sufficient target discrimination information to carry out the correct association on most of the frames. The presence of clutter provides an additional challenge to the kinematic only T2TA method. Clutters are easily eliminated by the combined T2TA method due to its target discrimination capability. This aspect is difficult to solve in case the clutters are set with a uniform likelihood over all target categories and the target width is set similar to the true target. The ACC of the kinematic only T2TA method can be improved by using either one of the extensions:

- Unique target id number matching. The target ids are generated by the track manager of the local sensor. Once a the track id set from different sensors is known, then this information could be used to carry out the association in the future fusion cycles.
- Track history can be used to increase the degree of correctness of association. However the length of the track history, which is also known as frame length is a design parameter and it has to be chosen according to the application's performance requirements.

These extensions are not the focus of this paper and the impact of these factors on the T2TA methods are reserved for the future investigation. The data from the real world automotive sensors are presented in the figure (6) and the corresponding camera image is reproduced in the figure (5) for a better understanding of the scene. Data from different sensors are colour coded. Objects belonging only to the target category of automobiles are considered for this paper. The decision of the image classifier present in the stereo camera system can be used as an enabler for this category selection at the fusion center. The green line with a width w in the plot represents the set of objects that are associated and fused. The estimated parameter width w of an object can be obtained from the lidar and the stereo camera. Due to space constraints, only the sensor objects present in the ego vehicle's direction of travel are presented here. In addition to the dynamic moving objects such as automobiles, the sensors also provide a set of objects

Fig. 5. This figure is the corresponding camera image of the scene presented in Fig. 6

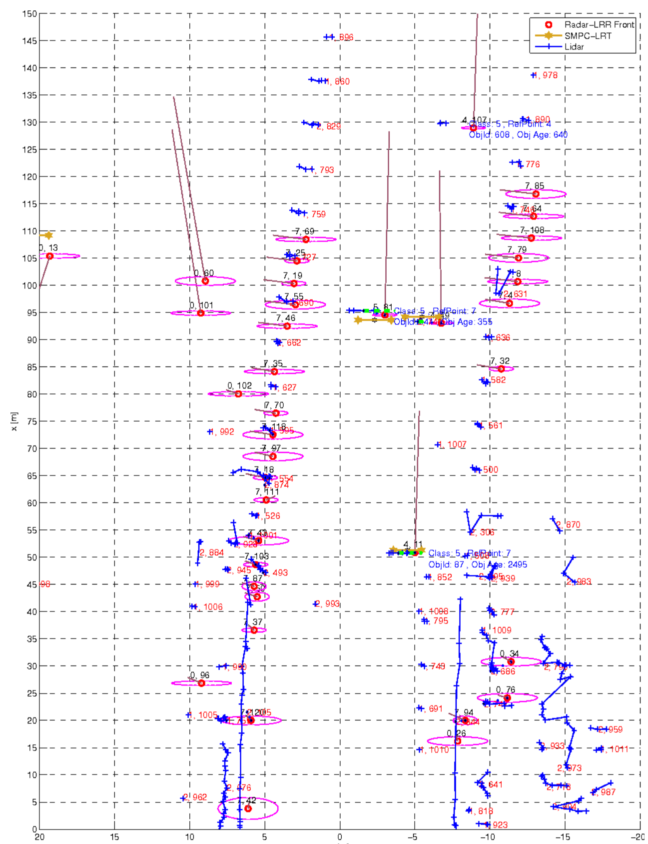

Fig. 6. **Real world data - sensor objects view**: The objects tracked by the radar, lidar and stereo camera are plotted in this figure. The red coloured dots are radar objects, blue coloured objects are from the lidar and the objects from stereo camera are represented by the gold colour. The green line represents the associated and fused objects that belongs to the target category of *automobile*. The data presented here is limited to the travel direction of the ego vehicle due to space constraints. The camera image of the scene in this plot is the Fig. 5.

that belong to the static environment like guard rails present in an usual automotive environment. The target category information is useful in cases when the set of dynamic targets cannot be resolved from the static objects by using same origin hypothesis. It is quite intuitive to map the automobiles in the figure (5) to the sensor and fused objects in the Fig. 6. For e.g.: The white hatchback car in the Fig. 5 is the associated and fused object at the vehicle coordinates $(50, -5)m$. At present the exact performance of different T2TA methods cannot be evaluated for the real world data for the following reasons:

- The ground truth data of the automobiles present in the scene are not available. Usually they are obtained either by equipping a test true target with a highly accurate DGPS system or as an alternative, a target labeling technique in the image space can be employed to extract the ground truth.
- The stereo camera classifier decision can be used as a ground truth but the classifier used in the stereo camera has limitations in the range and accuracy of the detection. Outage of stereo camera data also has to be taken into account. Due to these limitations, it is not advisable to use stereo camera data as a ground truth for the analysis of T2TA methods.

The preliminary results presented in this paper for the real world data, has to be further investigated for several use case scenarios and conditions. A more detailed association performance analysis with ground truth will be carried out.

IX. Conclusion

The objective of this paper is to study and verify the improvement of T2TA performance by integrating the non-kinematic information in addition to kinematic information in the T2TA procedure. A detailed state of the art of the existing algorithms that integrates non-kinematic target information in the T2TA decision making procedure has been presented. Integration of non-kinematic target category information in T2TA decision making procedure has been detailed. The classification uncertainty is handled using the sensor local classifier specific confusion matrix. A T2TA decision making architecture based independent kinematic and non-kinematic information has been proposed and explained in this paper. Results based on Monte Carlo simulation with 3 true targets affected by clutter have been presented. The performance of the T2TA algorithm based on the proposed architecture is of good accuracy when compared to the T2TA method that uses only the kinematic information. Preliminary results using the data obtained from the real world automotive environment perception sensors have been presented and a snapshot of the association view of automotive real world sensors has been depicted.

References

[1] International Road Federation, "The irf vienna manifesto on its: Smart transport policies for sustinable mobility," Geneva/Vienna, 23.10.2012. [Online]. Available: http://www.irfnet.ch/files-upload/pdf-files/irf_itsvienna_final_web.pdf

[2] K. Bengler, K. C. Dietmayer, B. Färber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[3] Daimler AG, "Mercedes-benz future truck 2025: World premiere of the spectacular study of tomorrow's trucks - autonomous driving into an exciting future," Stuttgart, 22.09.2014.

[4] Ziegler, J. and et al., "Making bertha drive—an autonomous journey on a historic route," *Intelligent Transportation Systems Magazine, IEEE*, vol. 6, no. 2, pp. 8–20, 2014.

[5] S. Tuohy, M. Glavin, C. Hughes, E. Jones, M. Trivedi, and L. Kilmartin, "Intra-vehicle networks: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 534–545, 2015.

[6] Y. C. Chee, S. Mori, W. H. Barker, and C. C. Kuo, "Architectures and algorithms for track association and fusion," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 15, no. 1, pp. 5–13, 2000.

[7] B. Duraisamy, T. Schwarz, and C. Wohler, "Track level fusion algorithms for automotive safety applications," in *Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on*, 2013, pp. 179–184.

[8] O. E. Drummond, "On categorical feature-aided target tracking," *Proc. SPIE*, vol. 5204, 2003.

[9] Y. Bar-Shalom, X. Tian, and P. K. Willett, *Tracking and data fusion: A handbook of algorithms*. Storrs, Conn.: YBS Publishing, 2011.

[10] B. Kragel, S. Herman, and N. Roseveare, "A comparison of methods for estimating track-to-track assignment probabilities," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 3, pp. 1870–1888, 2012.

[11] O. E. Drummond, "Integration of features and attributes into target tracking," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4048, pp. 610–622, 2000.

[12] Y. Bar Shalom and C. Huimin, "Multisensor track-to-track association for tracks with dependent errors," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 3, 2004, pp. 2674–2679. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1428864

[13] Y. Bar-Shalom and W. D. Blair, Eds., *Multitarget-Multisensor Tracking: Applications and Advances*, ser. Artech House radar library. Boston and London: Artech House, op. 2000.

[14] L. M. Kaplan, W. D. Blair, and Y. Bar Shalom, "Simulations studies of multisensor track association and fusion methods," in *Aerospace Conference, 2006 IEEE*, 2006, p. 16 pp.

[15] Y. C. Chee and S. Mori, "Metrics for feature-aided track association," in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1–8.

[16] J. Ferry, "Xmap: Track-to-track association with metric, feature, and target-type data," in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1–8.

[17] Y. Bar-Shalom and H. Chen, "Track-to-track association for tracks with features and attributes," in *Optics & Photonics 2005*, ser. SPIE Proceedings, O. E. Drummond, Ed. SPIE, 2005, pp. 59 131A–59 131A–12.

[18] R. W. Osborne, Y. Bar Shalom, and P. Willett, "Track-to-track association with augmented state," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, 2011, pp. 1–8.

[19] S. Mori, C. C. Kuo, and Y. C. Chee, "Performance prediction of feature-aided track-to-track association," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 4, pp. 2593–2603, 2014.

[20] Xin Tian and Y. Bar-Shalom, "On algorithms for asynchronous track-to-track fusion," in *2010 13th International Conference on Information Fusion (FUSION 2010)*, pp. 1–8.

[21] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proceedings National Institute of Science, India*, vol. 2, 1936, pp. 49–55.

[22] C. Huimin and Y. Bar Shalom, "Track fusion with legacy track sources," in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1–8.

[23] O. E. Drummond, "Attributes in tracking and classification with incomplete data," in *Defense and Security*, ser. SPIE Proceedings, O. E. Drummond, Ed. SPIE, 2004, pp. 476–496.

[24] Yaakov Bar-Shalom and Huimin Chen, "Track-to-track association using attributes," *J. Adv. Inf. Fusion*, vol. 2, no. 1, pp. 49–59, 2007.

[25] Powers, David Martin Ward, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *International Journal of Machine Learning Technology*, vol. 2, no. 1, pp. 37–63, 2011.