Data Association with Ambiguous Measurements

Matthew Travers, Todd Murphey, and Lucy Pao
Electrical and Computer Engineering
University of Colorado at Boulder
Boulder, Colorado 80309

***Abstract*— The problem addressed in this paper is that of tracking a single object in the neighborhood of several other closely spaced, similar objects where there is measurement mixing between objects observed at the sensor used to do the tracking. This problem is unique in that it does not consider there to be any clutter in the environment, producing erroneous measurements. The objects are all moving together at every time a measurement is taken. Another unique aspect of the problem considered here is the assumption that only one of the objects in the area of interest returns a measurement to the sensor, whether or not it be from the object being tracked. This situation of having unknown mixed-measurements actually occurs in quite a few real world systems. We will consider the mixed-measurement problem in a very particular situation in the problem section, but the ideas, and in particular the algorithms, can be applied to any number of associated systems with little alteration.**

I. Introduction

Tracking a single object in a unknown environment is a relatively well studied and understood process. Tracking a group of objects in formation is also well studied. Tracking a single object in a group is not.

The problem of tracking a single object in a group can be equated to tracking a single automobile in a convoy, the example considered later in this paper. In the example, the cars all drive down a known road together in formation. This makes the problem difficult because there is no maneuvering, a priori, or a posteriori knowledge of individual trajectories. Another aspect of the problem we will consider that adds difficulty is how the sensor makes measurements. The sensor only returns one measurement per time step. The sensor itself does not know if the measurement comes from the car we are tracking or simply one of the other vehicles in the convoy. As will be shown, certain data association algorithms can be used to effectively sort out measurements and provide reliable tracking.

The data association algorithms presented in this paper are very powerful tracking techniques. Not only are these ideas applicable to systems with distributions in space, as we will consider here, they are applicable to systems with distributions in time. An example which we have done some previous work on uses Geiger counter's as a distributed network of sensors to track a moving nuclear source. This situation is similar to the convoy of cars, except that instead of a second nearby car producing erroneous measurements, background radiation acts to interfere with the measurements coming from the "object of interest."

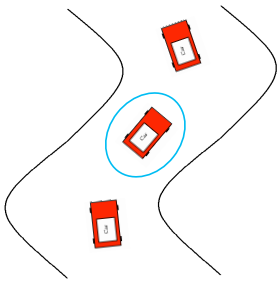

Fig. 1. Convoy of cars example..

The format of this paper will be to start out with a specific problem definition, followed by a short description of algorithms, then an explanation of how to apply these algorithms, followed by a presentation of results. The final section will be a short conclusion followed by a discussion of future work.

II. Problem Definition

To illustrate the effectiveness of these data association concepts applied to the mixed measurement sensor case, consider the simple case of a convoy of cars driving down a known road. The convoy can consist of any number of automobiles. One of the cars is somehow deemed to be of interest, i.e. there is a specific car that we would like to track. This situation is illustrated in Figure 1, the car of interest is circled.

As the cars drive down the road we will assume that there is no maneuvering. The vehicles stay in a line and the individual separation between cars remains basically constant for the entire time we are taking measurements. The road is assumed to be perfectly flat and a model for it defined. The cars are also assumed to be constrained to stay on the road.

The sensor is fixed to a tower with an unobstructed view of the entire road where the vehicles are traveling. The sensor only makes measurements on cartesian position, in kilometers (km), in a frame fixed to the base of the sensor tower and is of the type discussed previously, i.e. a single measurement is taken at each time step and the origin of that measurement is not known. Based on prior knowledge of the road, the sensor is tasked to always be pointing where it thinks the object of interest should be. The entire road is within the sensors range, thus there will be no missed measurements.

III. ALGORITHMS

A. The Extended Kalman Filter

The Extended Kalman Filter, EKF is a well known data association method for filtering nonlinear systems. The EKF algorithm produces an estimate of the state based on the current measurement. It will be used later in gating as well as the Kolmogorov-Smirnov tests, both discussed in this section.

The EKF is an estimation algorithm that maps the current estimate of the state, $\hat{x}(k|k)$, forward one time step. The current estimate is conditioned on the current state given the current measurement [B]

$$\hat{x}(k|k) \triangleq E[x(k)|Z^k]$$

where

$$Z^k \triangleq \{z(j), j = 1, ..., k\}$$

is the cumulative set of measurements up to time k. The associated state error covariance matrix is

$$P(k|k) \triangleq E\{[x(k) - \hat{x}(k|k)][x(k) - \hat{x}(k|k)]'|Z^k\}.$$

We wish to apply this one step estimation process to a nonlinear system. Consider a system with the following dynamics

$$x(k+1) = f[k, x(k)] + v(k). \tag{1}$$

Assume that the process noise, $v(k)$, is additive, zero-mean, and white

$$E[v(k)] = 0$$
$$E[v(k)v'(j) = Q(k)\delta_{kj}$$

where E is the variance. The measurement is

$$z(k) = h[k, x(k)] + w(k). \tag{2}$$

The measurement noise, $w(k)$, is also assumed to be additive, zero-mean, and white

$$E[w(k)] = 0$$
$$E[w(k)w'(j)] = R(k)\delta_{kj}.$$

It is assumed that the process noise, measurement noise, and initial state $\hat{x}(0|0)$ with the associated covariance $P(0|0)$ are all uncorrelated.

We proceed with finding $\hat{x}(k+1|k)$, the prediction, by expanding the nonlinear function (1). The expansion is accomplished by evaluating the Taylor series of (1) about $\hat{x}(k|k)$, the current estimate

$$\begin{aligned} x(k+1) = {} & f[k, \hat{x}(k|k)] + f_x(k)[x(k) - \hat{x}(k|k)] \\ & + 1/2 \sum_{i=1}^{n_x} e_i[x(k) - \hat{x}(k|k)]' f^i_{xx}(k)[x(k) - \hat{x}(k|k)] \\ & + (higher - orderterms) + v(k) \end{aligned} \tag{3}$$

where e_i is the i^{th} Cartesian basis vector and $f_x(k)$ is the Jacobian of the vector f

$$f_x(k) \triangleq [\bigtriangledown_x f'(k, x)]'_{x=\hat{x}(k|k)}.$$

Similarly, the Hessian of the i^{th} component of f is

$$f^i_{xx}(k) \triangleq [\bigtriangledown_x \bigtriangledown'_x f^i(k, x)]_{x=\hat{x}(k|k)}.$$

The prediction of the state at time $k+1$ from time k is then obtained by taking the expectation of the taylor expansion, (3), conditioned on Z^k (neglecting higher-order terms):

$$\hat{x}(k+1|k) = f[k, \hat{x}(k|k)] + 1/2 \sum_{i=1}^{n_x} e_i tr[f^i_{xx}(k)P(k|k)] \tag{4}$$

The covariance associated with this state prediction is found by first using (3) and (4) to form the estimation error

$$\begin{aligned} \tilde{x}(k+1|k) = {} & f_x(k)\tilde{x}(k|k) \\ & + 1/2 \sum_{i=1}^{n_x} e_i[\tilde{x}'(k|k) f^i_{xx}(k)\tilde{x}(k|k) \\ & - tr[f^i_{xx}(k)P(k|k)]] + v(k) \end{aligned} \tag{5}$$

Then multiplying (4) by its transpose and taking the expectation conditioned on Z^k

$$\begin{aligned} P(k+1|k) \triangleq {} & E[\tilde{x}(k+1|k)\tilde{x}'(k+1|k)|Z^k] \\ = {} & f_x(k)P(k|k)f'_x \\ & + 1/2 \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} e_i e'_j tr[f^i_{xx}(k)P(k|k)f^j_{xx}(k)P(k|k)] \\ & + Q(k). \end{aligned}$$

The measurement prediction is

$$\begin{aligned} \hat{z}(k+1|k) = {} & h[k+1, \hat{x}(k+1|k)] \\ & + 1/2 \sum_{i=1}^{n_z} e_i tr[h^i_{xx}(k+1)P(k+1|k)]. \end{aligned}$$

Subtracting this prediction from (2) yields

$$\begin{aligned} \tilde{z}(k+1|k) = {} & z(k+1) - \hat{z}(k+1|k) \\ = {} & h[k+1, \tilde{x}(k+1|k)] + w(k+1). \end{aligned}$$

The associated covariance, the *measurement prediction covariance* is

$$\begin{aligned} S(k+1) \triangleq {} & h_x(k+1)P(k+1|k)h'_x(k+1) \\ & + 1/2 \sum_{i=1}^{n_z} \sum_{j=1}^{n_z} e_i e'_j [h^i_{xx}(k+1)P(k+1|k) \\ & \cdot h^j_{xx}(k+1)P(k+1|k)] + R(k+1) \end{aligned} \tag{6}$$

where

$$h_x(k+1) \triangleq [\bigtriangledown_x h'(k+1, x)]'_{x=\hat{x}(k+1|k)}$$

and

$$h^i_{xx}(k+1) \triangleq [\bigtriangledown_x \bigtriangledown'_x h^i(k+1, x)]'_{x=\hat{x}(k+1|k)}.$$

The covariance between the state and measurement is

$$\begin{aligned} E[\tilde{x}(k+1|k)\tilde{z}'(k+1|k)|Z^k] & \\ = {} & P(k+1|k)h'_x(k+1) \\ = {} & E\{\tilde{x}(k+1|k)[h_x(k+1)\tilde{x}(k+1|k) \\ & + w(k+1)]'|Z^k\} \end{aligned} \tag{7}$$

Using (6) and (7), the filter gain is

$$W(k+1) \triangleq P(k+1|k)h'_x(k+1)S^{-1}(k+1).$$

The next estimate is then

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + W(k+1)v(k+1) \quad (8)$$

where,

$$\begin{aligned} v(k+1) &\triangleq z(k+1) - \hat{z}(k+1|k) \\ &= z(k+1) - h_x(k+1)\hat{x}(k+1|k) \end{aligned}$$

This vector, $v(k+1)$, is known as the innovation or measurement residual.

Lastly, the covariance associated with the the next estimate of the state is

$$\begin{aligned} P(k+1|k+1) &= P(k+1|k) - P(k+1|k)h'_x(k+1) \\ &\quad \cdot S^{-1}(k+1)h_x(k+1)P(k+1|k) \\ &= P(k+1|k) - W(k+1)S(k+1)W'(k+1) \\ &= [I - W(k+1)h_x(k+1)]P(k+1|k) \end{aligned} \quad (9)$$

B. Gating

Gating is a method of determining whether or not it is likely that a particular measurement came from the target we are tracking. We use gating to throw out measurements that do not make sense. Take for example an airplane flying close to the earth's surface. We are tracking this airplane with a sensor fixed to the ground and the sensor returns one measurement each time a sample is taken. If that airplane suddenly descends into a canyon, out of the sensors range, something (rock, tree, other airplane, etc...) else somewhere in the landscape covered by the sensors range will eventually register as a measurement. Gating says, if we know something about where the plane we are tracking was, we can make a prediction as to where we think it should be, any measurement not around there should be thrown out. Thus, in our example, the measurement returned by the tree 10 km away from where we think the next measurement from the plane should be, will be tossed out.

Since we are assuming that the measurement at each time step is normally distributed around the truth trajectory, it is possible to define a region in the measurement space where there is a high probability of finding the measurement

$$\begin{aligned} \tilde{V}_{k+1}(\gamma) &\triangleq \{z : [z - \hat{z}(k+1|k)]' \\ &\quad \cdot S^{-1}(k+1)[z - \hat{z}(k+1|k)] \leq \gamma\} \\ &= \{z : v'(k+1)S^{-1}(k+1)v(k+1) \leq \gamma\}. \end{aligned} \quad (10)$$

The weighted norm of the innovation, (10), is known as the validation region or the gate. The parameter γ can be thought of as the number of standard deviations squared. This parameter is set prior to gating; it defines the size of the gate in measurement space. The weighted norm of the innovation is chi-square distributed with the dimension of the measurement defining the number of degrees of freedom. This norm is chi-square distributed because the measurement is assumed Gaussian. If the measurement is not assumed to be Gaussian, (such as the case of accounting for the underlying geometry of the system in the measurement) gating can still be carried out, but with some modification [3].

To actually accomplish Gating, we make use of the EKF. Assume that the filter is initialized correctly, i.e. that we have a valid first estimate and sample. The size of the gate is determined by setting the value of γ. We then run through the first iteration of the filter up to the calculation of the innovation, $v(k+1)$, and its associated covariance, $S(k+1)$. These values are used to form the weighted norm in (10), which is compared to the chosen value of γ. If the value of the weighted norm is less than γ, the measurement falls within the gate and will be kept in the list of validated measurements, the next estimate and its associated covariance are calculated, and the EKF continues on to the next measurement. If the value of the weighted norm is greater than γ, the measurement falls outside of the gate and will be dropped from the list of validated measurements and no new estimate is calculated. When a measurement is dropped, the EKF continues running, but it must reuse the estimate associated with the last validated measurement to predict forward to the next estimate. If a measurement is dropped, the time step between various measurements becomes variable and must be accounted for in the algorithm. This process is carried out for the whole list of measurements.

C. Probabilistic Data Association

The $probabilistic\ data\ association\ filter$ (PDAF), also known as the Suboptimal Bayesian Algorithm, incorporates ideas from both the EKF as well as Gating to yield an estimate of state in the presence of an unspecified number of environmental objects. This is a main difference between the $PDAF$ and standard EKF, the EKF assumes that each measurement it receives is from the correct distribution where the $PDAF$ accounts for the fact that some of the measurements may not have not have originated from the target of interest.

Much like gating, the PDAF also defines a validation region in the measurement space. Instead of using the γ parameter that showed up in Gating directly, the PDAF uses γ (number of sigma's squared) along with the dimension of the measurement to define P_G (Table-1), the probability mass

$$P_G \triangleq P\{z(k+1) \in \tilde{V}_{k+1}(\gamma). \quad (11)$$

P_G is literally the probability that the true measurement will fall within the gate. A very high P_G, i.e. around 100%, means that we are allowing our validation region to incorporate nearly the whole measurement space, or in terms of Gating, measurements many standard deviations away from where we think the measurement should be. An issue that needs to be addressed here is that of the true measurement being detected at all. Looking back to the gating example where the airplane descends into a canyon: If we know that 20% of the total time we spend tracking the plane it is actually

in the canyon, out of range of the sensor, the filter will be incorrectly affected by these incorrect measurements (20%) if we do not account for them. This is done through a parameter called the *detection probability*, P_D

$$P_D \triangleq P\{The\ true\ measurement\ is\ detected\}. \quad (12)$$

The PDAF assumes that we are only tracking one object at a time. It is assumed that the object is modeled by (1) and that this objects track has been properly initialized. The set of validated measurements at time k is

$$Z(k) \triangleq \{z_i(k)\}_{i=1}^{k} \quad (13)$$

where m_k is the number of measurements in the validation region. The cumulative set of measurements is

$$Z^k \triangleq \{Z(j)\}_{j=1}^{k}. \quad (14)$$

What makes the PDAF suboptimal (i.e. suboptimal Bayesian filter) is that it decomposes the estimation with respect to the origin of each element of (13) as opposed to (14). Another way to say this is: The Suboptimal Bayesian Algorithm decomposes the estimate with respect to each element of the latest set of measurements, while the optimal algorithm decomposes with respect to each element of the sequence of measurements.

In order to carry out the decomposition of the estimate in the PDAF, the following basic assumption must be made

$$p[x(k)|Z^{k-1}] = N[x(k); \hat{x}(k|k-1), P(k|k-1)],$$

which says that the state is normally distributed according to the latest estimate along with its covariance matrix.

From this assumption and with $Z(k)$, define the events

$$\begin{aligned}\theta_i(k) \triangleq \{&z_i(k) is\ the\ target-originated\\ &measurement\} \qquad ,i=1,...,m_k\\ \theta_0(k) \triangleq \{&none\ of\ the\ measurements\ are\ target\\ &-originated\ at\ time\ k\}\end{aligned} \quad (15)$$

with probabilities

$$\beta_i(k) \triangleq P\{[\theta_i(k)|Z^k\}, \quad i=0,1,...,m_k \quad (16)$$

The process which yields these probabilities is known as *probabilistic data association*. Note:

$$\sum_{i=0}^{m_k} \beta_i(k) = 1$$

i.e. the events (15) are mutually exclusive as well as exhaustive.

We then use the total probability theorem with respect to these two events to obtain the conditional mean, or next update, of state at time k

$$\begin{aligned}\hat{x}(k|k) &= E[x(k)|Z^k]\\ &= \sum_{i=0}^{m_k} E[x(k)|\theta_i(k), Z^k]P\{\theta_i(k)|Z^k\}\\ &= \sum_{i=0}^{m_k} \hat{x}_i(k|k)\beta_i(k)\end{aligned} \quad (17)$$

where,

$$\hat{x}_i(k|k) = \hat{x}(k|k-1) + W(k)v_i(k), \quad i=1,...,m_k \quad (18)$$

$\hat{x}_i(k|k)$ is conditioned on $\theta_i(k)$, and

$$v_i(k) \triangleq z_i(k) - \hat{z}(k|k-1).$$

Since $\hat{x}_i(k|k)$ is conditioned on $\theta_i(k)$, there is no measurement origin uncertainty and the filter gain, $W(k)$, is thus the same as it was in the EKF. This also leads to the fact that if none of the measurements at time k are correct, the updated state estimate will be

$$\hat{x}_k(k+1|k+1) = \hat{x}(k+1|k). \quad (19)$$

Plugging (18) and (19) into (17) yields the overall state update equation of the PDAF

$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)v(k) \quad (20)$$

where

$$v(k) \triangleq \sum_{i=1}^{m_k} \beta_i(k)v_i(k).$$

The error covariance associated with (20) is

$$P(k|k) = \beta_0(k)P(k|k-1) + [1-\beta_0(k)]P^c(k|k) + \tilde{P}(k) \quad (21)$$

where $P(k|k-1)$ is the same as is in the Kalman Filter,

$$\tilde{P}(k) = W(k)[\sum_{i=1}^{m_k} \beta_i(k)v_i(k)v_i'(k) - v(k)v'(k)]W'(k)$$

and

$$P^c(k|k) = [I - W(k)H(k)]P(k|k-1).$$

The last thing needed to run the PDAF is to define the association probabilities (16). We start by writing out the probabilities explicitly

$$\begin{aligned}\beta_i(k) &\triangleq P\{\theta_i(k)|Z^k\}\\ &= P\{\theta_i(k)|Z(k), m_k, Z^{k-1}\},\ i=0,1,...,m_k\end{aligned}$$

Then, using $Bayles' rule$,

$$\beta_i(k) = \frac{1}{c}p[Z(k)|\theta_i(k), m_k, Z^{k-1}]P\{\theta_i(k)|m_k, Z^{k-1}\} \quad i=0,1,...,m_k \quad (22)$$

where c is the normalization constant.

To find the β_i's, the relative joint density of validated measurements as well as several other parameters must be found. The product of the probability density function, PDF, associated with the correct measurement and PDF associated with the incorrect measurement is the joint density of validated measurements conditioned on the events θ_i for $i \neq 0$. [1]

The PDF associated with the correct measurement is

$$\begin{aligned}p[z_i(k)|\theta_i(k), m_k, Z^k] &= P_G^{-1}N[v_i(k); 0, S(k)]\\ &= P_G^{-1}|2\pi S(k)|^{-1/2}\\ &\quad \cdot exp[-1/2v_i'(k)S^{-1}(k)v_i(k)].\end{aligned}$$

Th PDF of the incorrect measurement is assumed to be uniformly and independently distributed in the validation region.

The PDF in (22) is then just

$$p[Z(k)|\theta_i(k), m_k, Z^{k-1}] = \begin{cases} V_k^{-m_k}, & i = 0 \\ V_k^{-m_k+1} P_G^{-1} N[v_i(kK); 0, S(k)] & i = 1, ..., m_k. \end{cases} \quad (23)$$

The probabilities of the events θ_i conditioned on the number of validated measurements are

$$\gamma_i(m_k) = \begin{cases} 1 - P_D P_G. & i = 0 \\ P_D P_G / m_k, & i = 1, ...m_k \end{cases}$$

The sample spatial density is defined as

$$\lambda = \frac{m_k}{V_k}.$$

The volume of the validation region is

$$V_k = c_{n_z} |\gamma S(k)|^{1/2} = c_{n_z} \gamma^{n_z/2} |S(k)|^{1/2} \quad (24)$$

where n_z is the dimension of the measurement and c_{n_z} is the volume of the n_z-dimensional unit hypersphere.

Finally, we combine (23) and (24) into (22) to yield the equations for the β_i's

$$\beta_i(k) = \frac{e_i}{b + \sum_{j=1}^{m_k} e_j} \quad (25)$$

$$\beta_o(k) = \frac{b}{b + \sum_{j=1}^{m_k} e_j} \quad (26)$$

where

$$e_i \triangleq exp[-1/2 v_i'(k) S^{-1}(k) v_i(k)] \quad (27)$$

and

$$b \triangleq \lambda |2\pi S(k)|^{1/2} (1 - P_D P_G)/P_D \quad (28)$$

$$= (2\pi/\gamma)^{n_z/2} \lambda V_k c_{n_z} (1 - P_D P_G)/P_D. \quad (29)$$

D. Kolmogorov-Smirnov Tests

Another method which is very similar to gating in terms of what it does is known as the Kolmogorov-Smirnov tests, $KS\ tests$. The KS tests are similar to gating because they can be used to isolate and get rid of a particular measurement based on some comparison. They are different in the metrics used to do the comparison.

Consider $x_1, x_2, ..., x_n$, independent observations of a random variable with unknown cumulative distribution function F(x). The null hypothesis is

$$H_0 : F(x) = F_0(x), \quad (30)$$

then, any test of this hypothesis is a goodness-of-fit test [A]. The KS tests and many simple variants are just goodness of fit tests. Empirical cumulative distribution functions, CDF's, are formed for a $window$ of n measurements and an empirical CDF if formed. Some metric is then applied to measure the distance between theoretical and empirical CDF's.

What we will actually present here is a variant of the "normal" KS tests. Again consider n samples ordered such that

$$x_1 \leq x_2 \leq .. \leq x_n.$$

The empirical cumulative distribution function is defined to be [2]

$$S_n(x) = \begin{cases} 0, & x < x_1 \\ r/n, & x_r \leq x < x_{r+1} \\ 1, & x_n \leq x \end{cases} \quad (31)$$

Assuming that $F_0(x)$ is the true, fully specified cumulative distribution function that the measurements $x_1 \rightarrow x_n$ have actually come from, the following result follows from the strong law of large numbers

$$\lim_{n \rightarrow \infty} P\{S_n(x) = F_0(x)\} = 1$$

This is a very powerful capability of this type of test. As long as it is certain that all of the observations are from a single distribution, the theoretical cumulative distribution function for that distribution can de defined in terms of the empirical data.

Define the metric used for measuring the separation between theoretical and empirical CDF's to be

$$D_n = \int_0^\infty (F_0(x) - S_n(x)) dx. \quad (32)$$

The main difference between "normal" Kolmogorov-Smirnov tests and this variant is how the metric is defined. The metric for the actual KS tests is defined as [Source]

$$D_n = \sup_x |S_n(x) - F_0(x)| = max\{D_n^+, D_n^-\} \quad (33)$$

This normal metric proved to not be sensitive enough for the particular case here being considered. The metric (32) is extremely sensitive to incorrect measurements. This results from the fact that (32) largely takes advantage of how a single incorrect measurement affects the shape of the empirical CDF.

There is currently no theoretical basis for specifying a critical value of D_n for this version of the Kolmogorov-Smirnov tests. The method used for the problem considered in this paper will be further discussed in the example section.

IV. Example

Consider now the specific case were we have only two cars traveling in a small convoy. We choose two cars out of simplicity, but in reality there can be many cars and it will not change anything. The effect separation distance between the two vehicles will be explored.

The dynamics of this system are modeled by using the constrained Euler-Lagrange equations. To obtain the Euler-Lagrange equations, start by defining the $Lagrangian$ for the generalized state q, defined as the kinetic energy, T, minus the potential, V

$$L(q, \dot{q}) = T(q, \dot{q}) - V(q).$$

Then define the constrained $Euler-Lagrange\ equations$ in vector form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \lambda \omega^T(q) \tag{34}$$

where λ is the $Lagrange\ multiplier$ and $\omega(q)$ is the constraint matrix. We make use of the $Lagrange - d'Alembert\ formulation$, i.e. the forces of constarints do no virtual work, to write down the equation

$$\omega(q)\dot{q} = 0. \tag{35}$$

In the car example, we will be considering the cars to be fixed to a flat road, thus the only terms that will show up in the Lagrangian are those associated with the kinetic energy. Thus,

$$L(x, y, \dot{x}, \dot{y}) = 1/2M_{car}(\dot{x}^2 + \dot{y}^2),$$

assuming Cartesian coordinates. The left hand side of the Euler-Lagrange Equations then just becomes

$$\ddot{q} = 0.$$

To find the constraint matrix, $\omega(q)$, we write down what the constraint must be. The model of the road is assumed to be known, call it g(x). Then write down the constraint equation

$$y(x) = g(x) \tag{36}$$

We obtain $\omega(q)$ by differentiating (36) with respect to time and writing the result in the form of (35). The Lagrange multiplier is found by simultaneously solving (34) along with

$$\frac{d^2}{dt^2}(y(x) = g(x)) \tag{37}$$

For our example, $g(x) = sin(x)$. From this, ω and λ are

$$\omega(\{x, y\}) = \{-cos(x(t)), 1\}$$
$$\lambda = -\frac{2M_{car}sin(x(t))x'(t)^2}{3 + cos(2x(t)}$$

Thus the complete equations of motion for the system are defined. To these equations we then add in noise to simulate process noise. For this example, the process noise is drawn from the distribution $N[0, 0.0001]$. The truth trajectories are obtained by integrating these equations (with noise added in). In practice it is assumed that we do not know truth, but for the purposes of "creating" samples and for analysis, we will here make use of it.

Sampling is accomplished by assuming that the samples are drawn from a normal distribution centered around zero with standard deviation one half meter, i.e $N[0, 0.0005]$. The sampling period is $dt = 0.0001s$. This implies that for our example of the two cars, we find two truth trajectories by integrating, sample truth at the corresponding times for each car, and add gaussian distributed noise ($N[0, 0.0005]$) to each truth sample. For our particular example, what I am left with after sampling are two lists of measurements from two separate truth trajectories with gaussian noise added at each time step for each vehicle individually.

From the samples, we then wish to add the effect of mixing. This is done by first determining the percentage of incorrect measurements to consider. For the purposes of this paper, we will assume that this percentage is very low, 10%. but this percentage is itself a parameter still being explored. The mixing is carried out by simply determining the overall number of incorrect samples to add, then randomly replacing measurements from the from the list of the car being tracked (obtained after sampling) with measurements from the car driving close to the first, but not being tracked.

Once the mixed list of measurements is obtained, gating and the PDAF are relatively straight forward processes. The actual implementation for our specific case will be further addressed in the results section.

Before directly implementing the KS tests on the measurements, the way in which the noise is actually added into the system needs to be accounted for. All the noise in the system enters with zero mean in an additive way. This means that each measurement is actually the state plus the noise, thus to compare measurements to each other directly, we must fist subtract off the state, or because we do not know the state exactly, our best estimate of it. This means that the measurements referred to in the KS tests Algorithm description are actually the innovations associated with the measurements at each time step.

In attempting to implement the "normal" Kolmogorov-Smirnov tests for this example it was determined that the tests were not reliable enough to throw out a single incorrect measurement occurring at some random point in the list of otherwise true measurements of the object being tracked. The problem that occurred: Our expectation for adding a single incorrect measurement to the empirical CDF was that doing so would increase the D_n defined in (33) past some preset threshold, thus indicating a false measurement. In implementation, this was not the case. Adding a single incorrect measurement actually had no discernible difference on the value of D_n as defined in (33). By geometrically looking at the empirical CDF, it was observed that a significant difference in the areas between theoretical and empirical CDF's occurred when a single incorrect measurement was added. This is what ultimately lead to defining the metric for our particular situation as it is in (32).

There is a wide body of well developed methods for setting the threshold value, D_n in (33), for the regular Kolmogorov-Smirnov tests [3,4]. For the variation on these normal tests presented in the Algorithms section, there is no theoretical basis for calculating this value. Fo the present example, the threshold value will be set largely by inspection of D_n values in areas where there are known to be only true measurements. Using the well developed methods for calculating thresholds of the standard KS tests, we are currently developing a sound theoretical basis from which a more objectively set threshold can be obtained for the presented variation to these tests.

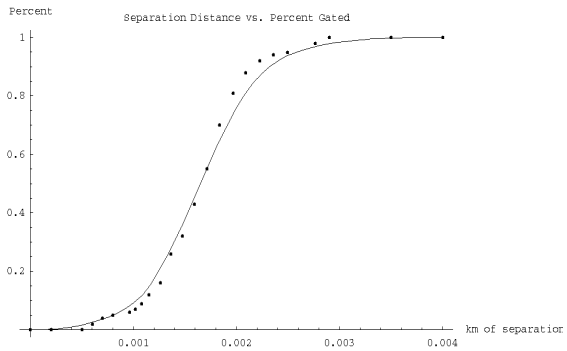

Fig. 2. Percent of incorrect measurements correctly gated out..

V. Results

All measurements for the two car system are just cartesian x and y, as stated in the problem definition. All the results in this section will be assumed to be stated in a "radial" format. This means that each distance presented in this section is actually the Euclidean norm of the measurement $\{x, y\}$.

A. Gating

The only parameter to be set in gating is the γ from (10), which effectively sets the gate size. Since our goal is to exclude the false measurements while keeping as many of the true measurements as possible, we set this value to $\gamma = 9$ ($\sigma = 3$, $\gamma = \sigma^2$), thus allowing about 97% of the true measurements into the gate [1]. In this example the value of γ is being chosen somewhat randomly, optimum methods for calculating this threshold do exist [3].

Figure 2 shows gating performance as a function of the separation distance between the two vehicles, for the given gate size. The points are actual data points obtained by varying separation distance; the solid line is a curve fit to the data. The performance is measured by the percentage of the incorrect measurements that get correctly gated out. For example, looking at Figure 2, under about 0.0007 km of separation between vehicles, not one of the incorrect measurements will get gated out; all 10% of the incorrect measurements will be left in the measurement list. In the same way, above about 0.0035 km all of the incorrect measurements get gated out. Thus, with a standard deviation for the measurements of $\sigma = 0.0005$, at around seven σ's separation, gating works perfectly.

B. PDAF

The PDAF results will be presented in both the gating and non-gating situations. It will be shown that gating is not only an effective method for cropping out incorrect measurements, but in so doing, better overall tracking performance can be achieved.

As mentioned previously, the PDAF will need to be initialized in much the same way as the EKF. The difference between the two is that P_D and P_G will need to be set for the PDAF. First consider the case where the two vehicle are relatively far away, i.e. about 10m. For the case where there is no gating, $P_G = 1$ (because nothing has been gated out)

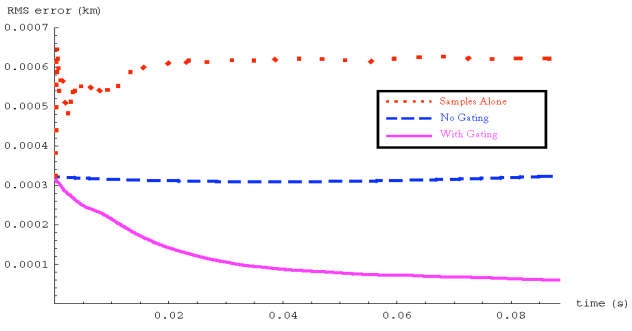

Fig. 3. RMS errors for PDAF when cars are driving 10m apart.

and $P_D = 0.90$ (because 10% of the true measurements have been replaced with false measurements). When there is gating, $P_G = 0.971$ (because $\gamma = 9$ in the gating, see [1]) and $P_D = 1$.

While the cars are driving far apart from each other, the PDAF will have lower RMS error when there is gating as opposed to when there is no gating. This can be seen in Figure 3. The dotted line at the top of the figure is the RMS error of the raw measurements compared to truth at each time step. The dashed line is the PDAF estimates compared to truth when there is no gating; the solid line is the PDAF estimates compared to truth when there is gating.

Then consider the case where the two automobiles are driving very close to each other i.e. about 2m. apart. The results for this situation are pictured in Figure 4.

Looking back at Figure 2, at 2m of separation, gating will not "gate out" all of the incorrect measurements. In fact, for this separation distance, about 20% of the incorrect samples will still be left in the list of measurements This means that when we run the PDAF with gating, P_D must be lowered appropriately to account for this. Since the number of incorrect measurements left after gating is still relatively low for this case (about 20 in 1000), P_D is only dropped to $P_D = 0.98$. Everything else stays the same. In this case, the gating versus non-gating performances are much closer than in the case where the cars were far apart, although they both still do better than the samples alone, as can be seen in figure 4. This decrease in separation distance can be explained by the fact that some of the incorrect measurements in the gating case are not going to get gated, but also by the fact that the error incurred by incorrect measurements in the non-gating

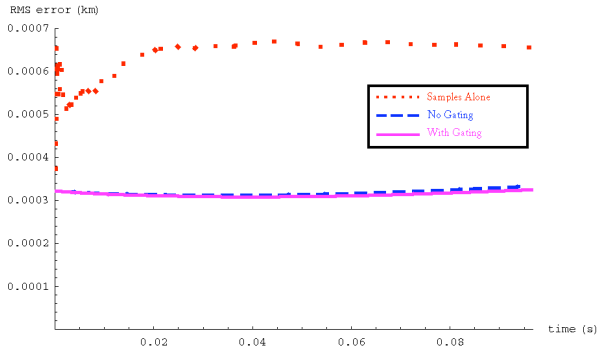

Fig. 4. RMS errors for PDAF when cars are driving 2m apart.

case will be lower when the vehicles are closer together. This implies that as the separation distance is decreased, the RMS error plots for the gating and non-gating cases will move toward each other until they eventually become the same (occurs at a separation distance where gating fails completely, around 0.005 km).

C. Kolmogorov-Smirnov tests

The ultimate goal of the variation of the Kolmogorov-Smirnov tests described in the Algorithms section is to isolate measurements that come from distributions not associated with the one of the car that we are tracking. As described, this is accomplished by calculating the integral difference between theoretical and empirical cumulative distribution functions, then using this value as a metric, there is a comparison against a preset threshold.

We will assume for this test that we have limited knowledge of the sensor characteristics. It will also be assumed that we have a large (n=20,000, where n is the number of measurements) stream of measurements that we have 100% confidence came from the car that we are tracking, perhaps from a previous trip down the road, or perhaps from a similar car's previous trip down the road. This stream of measurements is used to form (using (31)) the theoretical CDF, used as a comparison for the empirical CDF's described shortly (Again, this is possible due to the strong law of large numbers).

The window size for the two car example was set to ten, i.e. n=10. Thus, an empirical CDF function is formed according to (31) for the first ten measurements hitting the sensor. This function is used to obtain a value of D_n according to (32). This value is then compared against a threshold. This process iterated over the first 210 measurements is pictured in figure 5. The horizontal axis is the window number, where 1 corresponds to the window covering measurements 1 through 10, 2 corresponds to 2 through 11, and so on. The vertical axis is the value of D_n for that window. The solid line is the threshold, set by inspection. The dashed line represents the situation where the two cars are far away from each other, 7m. The dotted line represents the situation in which the cars are very close, about 2m. The incorrect measurements are at positions 28, 59, 89, 117, 146, and 178. In figure 5 we can see "humps" in the values of D_n centered around these positions. This happens because there are actually 10 windows that will span the incorrect measurement. When the cars are far from each other, 7m, these humps are very pronounced and are way above the threshold. When the cars are close, the incorrect measurements barley produce values of D_n that get over the threshold.

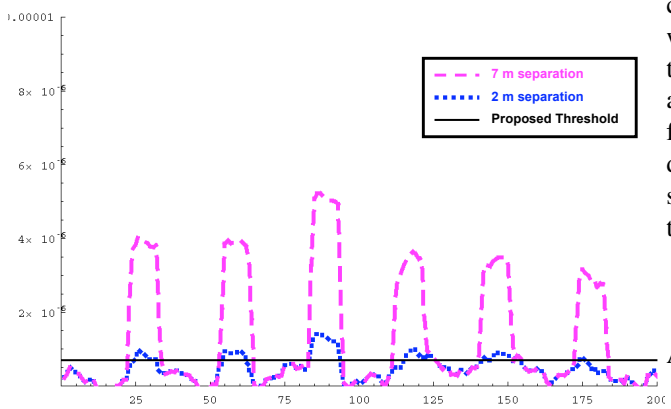

Fig. 5. Integral difference between empirical and theoretical CDF's as a function of window number

The future hopes for this test are to apply more complicated filtering techniques such that the "humps" are more pronounced at lower separation distances. Thus the threshold can be lowered without affecting the truth data and better performance can be achieved.

VI. Conclusions & Future Works

We have shown that the data association ideas of gating, the probabilistic data association filter, and a Kolmogorov-Smirnov type test provide an effective manner in which to deal with the problem of ambiguous measurements. It has been shown that in the situation of mixed measurements, these methods can be applied to provide reliable tracking, as defined by relatively much better than the samples alone.

Although reliable tracking was the actual stated result of this paper, another important result is how highly adaptable the overall algorithm is. This is in part due to the essentially black box nature of some of the data association algorithms themselves, but is largely due to using the Euler-Lagrange method of obtaining the dynamics for the system. The entire algorithm presented can be run on any system for which we can define the kinetic energy, potential energy, and constraints. After defining these quantities, little or no modification needs to be made to the rest of the algorithm to accommodate an enormous variety of systems.

Due to the effectiveness of these data association methods applied to this problem, there are several avenues of future work being pursued. The first direction in which this work is heading is to more formally address the problem from a more geometrical interpretation. These data association techniques have previously been applied in a somewhat naive way. Although we have been considering a nonlinear system, we have been assuming that the various noise distributions vary nicely along the trajectories. In the two car example, we basically assumed that the road was actually $\Re^2$ instead of some constrained sub-manifold. In the future, we would like to show that improved tracking performance is possible when the distributions are dealt with in a more geometrically correct way. Another immediate direction that we see this work heading in was discussed in the introduction, namely the nuclear detector system. Recall that this system utilizes a distributed network of sensors that produce time stamps for events. The nuclear is a particular example that we have done some previous work on, but there are many distributed sensing situations imaginable where the ideas presented in this paper are readily applicable.

VII. Acknowledgments

The authors gratefully acknowledge the contribution of AFOSR and reviewers' comments.

REFERENCES

[1] Y. Bar-Shalom and T.E. Fortmann, *Tracking and Data Association*, Academic Press, INC., San Diego, CA; 1988.

[2] T. Bailey, B. Upcroft, H. Durrant-Whyte, "Validation Gating for Non-Linear Non-Gaussian Target Tracking", *in Ninth International Conference on Information Fusion*, Florence, Italy, 2006, pp. 1-6.

[3] Y. Kosuge and T. Matsuzaki, "The Optimum Gate Shape and Threshold for Target Tracking," *SICE 2003 Annual Conference* , Fukui, Japan, 2003, pp 2152-2157.

[4] A. Stuart, K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, Oxford University Press, 1999.

[5] L.H. Miller, "Table of Percentage Points of Kolmogorov Statistics," *J. American Stati. Assoc.*,51(273):111-121, Mar. 1950.