A Multiple Hypothesis Tracker with Interacting Feature Extraction

A MULTIPLE HYPOTHESIS TRACKER WITH INTERACTING FEATURE EXTRACTION

BY

JAMES MCANANAMA, B.Eng., P.Eng.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2010) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE:	A Multiple Hypothesis Tracker with Interacting Feature Extraction
AUTHOR:	James McAnanama B.Eng., (Materials Engineering and Science) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. T. Kirubarajan
NUMBER OF PAGES:	xi, 60

To My Life's Greatest Gifts:

Leslie, Matthew, and Graham.

Abstract

The multiple hypotheses tracker (MHT) is an optimal tracking method due to the enumeration of all possible measurement-to-track associations. However, its practical implementation is limited by the NP-hard nature of this enumeration. To bound the computational complexity, some means of limiting the number of possible associations is required. Typical solutions include the interposition of rules to guide the pruning and merging of tracks. Other proposals have shown that the performance of a tracker, MHT or not, can be improved using feature information (e.g., signal strength, size, type) in addition to kinematic data. The inclusion of feature information allows for the discrimination to further gate the data associations. However, in most tracking systems, the schemes to manage the data association problem are extraneous to the Bayesian framework of the MHT. Further, the extraction of features from the raw sensor data is typically independent of the subsequent association and filtering stages. The features are then used in either an ad hoc way or are they are fused with the MHT tracker; they are not used intrinsically within MHT framework. In this thesis, a new approach whereby there is an intrinsic interaction between feature extraction and the MHT is presented. The measure of the quality of feature extraction is input into measurement-to-track association while the prediction step feeds back information to be used in the next round of feature extraction to increase the information available

a priori. The motivation for this forward and backward interaction between feature extraction and tracking is to improve the performance in both steps. This approach allows for a more rational partitioning of the feature space, removing unlikely features from the assignment problem. In addition, a track-specific detection probability becomes available to the prior. This probability significantly improves the coasting behavior when measurements are not available for track continuation. Simulation results demonstrate the benefits of the proposed approach.

Acknowledgements

I have come to learn that, when you work fulltime and enter into graduate school as a part-time student coincident with your wife being delivered of your second child, by the end of your studies you have some acknowledging to do... Therefore, I sincerely acknowledge the love and support of my wife, best friend, and most accommodating critic Leslie. This certainly would not have happened without your help. To my biggest fans, Matthew and Graham: if you happen to find yourselves as adults in Thode Library with a level of déjà-vu, it will be because you used to come with me to return books.

The first year of my studies was a trial by fire. For this I have to thank and blame my supervisor, Kiruba. Your calm manner and reassuring humour combined with your intellect and mastery of this field can fool your student into believing that it is all so easy - until they jump in! Now that the panic has subsided, I can look back and appreciate how much I learned in a very short period of time. Of course, one cannot find their way out of an ECE degree without acknowledging the constant direction available from Cheryl Gies.

Thanks to my colleagues and superiors at Wescam for their interest and encouragement. Special thanks to the late Lawrence Sinclair - you were absolutely right.

It may seem odd, but this thesis started eleven years ago under the title of "Liquid-Liquid Phase Separation During Polymerization". If estimation can teach us anything it is that where you end up may be hard to predict but makes a lot of sense in retrodiction. Thanks then to my initial thesis supervisor, Professor Gyan Johari, for showing me the universe in a grain of sand.

We are the product of our parents efforts, although I don't think we appreciate it fully until we are ourselves parents. Thanks to my mother, Judith, for your example and encouragement; I suspect you are the source of my own feelings of McMaster déjà-vu. Thanks to my father, Owen, for the validation a young boy seeks. Thanks also to my second set of parents, Nettie and Mal - sometimes a grown boy benefits from such validation too!

Contents

List of Figures

Chapter 1

Introduction and Problem Statement

1.1 Motivation Statement

The use of airborne vision systems in intelligence, security, and reconnaissance (ISR) is a relatively new and growing field (Best, 2010). One aspect of this technology is the acquisition and tracking of multiple targets. As discussed below, the field of target tracking is itself a developing area of study with the majority of the effort applied to radar and sonar applications. The focus of this study is on the advancement of algorithms used in tracking multiple targets with infrared (IR) and electro optical (EO) imaging sensors used onboard an airborne platform without any prior target information.

It is generally accepted that the current state of the art for multiple target tracking is given by the multiple hypothesis tracker (MHT) combined with the interacting

multiple model (IMM) kinematic filter (Blackman, 2004). The multiple hypothesis tracker is an optimal tracking method due to the enumeration of all possible measurement-to-track associations. However, its practical implementation is limited by the NP-hard nature of this enumeration. To bound the computational complexity, some means of limiting the number of possible associations is required. The problem for this thesis is the prosect of aiding the track to measurement associations by exploiting feature information within the native MHT framework. Specifically, can additional information be garnered and can performance be improved with more interaction between the different steps of the MHT algorithm? An approach to this issue is presented in section 1.2 following a literature review.

1.2 Literature Review

The state of the art of multiple target tracking MTT has evolved over the last half century with milestones realized in both the underlying algorithms and in the available computational speed of computing hardware. The result is the existence of multiple classes of solutions to the MTT problem. This evolution continues, driven in part by the development of new sensors with which one can track as well as increasing computational capabilities that allow for more complex algorithms. The following section describes the abstract steps of multiple target tracking (MTT) followed by a review of the common algorithms used to instantiate MTT.

Semantically the problem has been described in its simplest form as a process of measurement formation, association of these measurements to tracked targets, followed by an estimation of the target states (Bar-Shalom and Li, 1995). In the context of video tracking, measurements are formed by processing a frame of video to extract

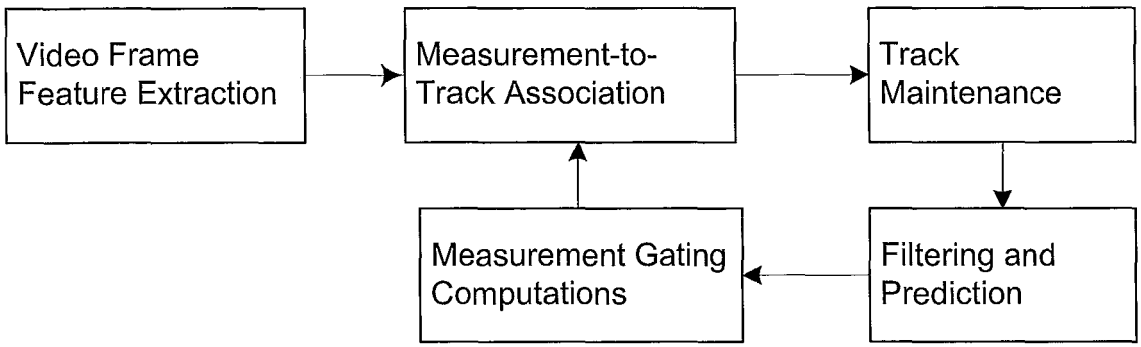


Figure 1.1: MTT Flow Diagram

trackable features. Each frame of video is a source of a new set of measurements. A track maintenance step is added after the data association to manage track initiation, propagation, and termination. In addition, a gating step is required in practical implementations to bound the data association step (Blackman, 2004). The resulting video MTT architecture is shown in figure 1.1 and each process is described below.

1.2.1 Feature Extraction

A video image is created when a sensor on a focal plane converts incoming electromagnetic energy into a 2-D array of intensity values proportional to the incident intensity and responsivity of the sensor to the incident spectrum. The end result is a spatial domain representation of the sampled energy intensity. Information in the spatial domain is represented by edges, areas of high spatial frequency. Colour images are made by combining triplets of picture elements that are responsive to the red, green, and blue wavelengths. In this case, the combined image can be thought to have both an average intensity content (grayscale content), and a chromatic content (Gonzalez and Woods, 2002).

There are many methods that can be used to process an image into trackable

measurements based on user defined windows, shape segmentation, detection of deformable contours found in the previous frame, visual learning methods, and feature extraction (Trucco and Plakas, 2006). Window tracking employs correlation to match a given pattern in the current video frame. The window definition requires some prior knowledge of the target appearance, typically cued by the operator. As the name implies, shape segmentation techniques dichotomize foreground and background based on some prior notion of target shape. Similarly deformable contours and visual learning methods require a prior notion of target appearance and are thus not suited to the focus of this study.

Feature extraction is the process whereby corners, lines, contours, regions, etc are filtered from the video frame (Sonka *et al.*, 2007). This technique has been shown to perform well when no prior target information is known (Cox and Hingorani, 1994; Papanikolopoulos *et al.*, 2006; Zetterlind and Matechik, 2006). Further, an beneficial interaction between feature extraction and tracking has been presented for another well known tracking algorithm, the Probabilistic Data Association Filter (Kirubarajan *et al.*, 1997). As a result, feature extraction techniques were selected for use in this study. Specifically, the video images are converted to a binary representation based on a threshold that dichotomizes the image to maximize the gray level variance between targets and background(Otsu, 1979; Sonka *et al.*, 2007). Next the targets are defined as connected pixels in the foreground. The principal components of these target regions are invariant to rotation (Gonzalez and Woods, 2002) and for this reason they were chosen as the trackable feature used herein.

1.2.2 Data Association

Once a set of measurements has been obtained, a difficult problem remains in determining the relationship between the measurements and the targets to be tracked. For example, a measurement may be from a target under track, from a new target, or it could be the result of a false measurement from clutter or noise. The data association step is computationally the most difficult task in multiple target tracking. In one study, the data association was found to take over 95% of the processing time (Pattipati *et al.*, 2000).

Early approaches to data association include the nearest neighbor and strongest neighbor filters which validate measurements based on their relative location or strength(Blackman, 1986; Li, 1998; Leung *et al.*, 1999). These filters look for the 'best' measurements and discard the others. Further, they allow for multiple tracks to share the same measurement. These techniques do not perform well in scenes of high clutter, and fail to account for the chance that all measurements may be false. The nearest neighbor approached was refined to find the best association hypothesis across all possible track to measurement associations. This approach, called the Global Nearest Neighbor (GNN) algorithm, only works well in the case of widely spaced targets with accurate measurements and low clutter (Blackman, 2004).

Another class of algorithms was developed to consider all of the measurements in a probabilistic sense - Probabilistic Data Association (PDA). These algorithms include probabilistic extensions to the nearest and strongest neighbor filters (Li, 1993; Li and Zhi, 1996) whereby the event that the nearest/strongest measurement does not originate from the tracked target is considered. In addition, another approach is

to use a weighted combination of all the measurements. This was first done in the single target, multiple measurement scenario as the probabilistic data association filter (PDAF) (Bar-Shalom and Tse, 1975). The multiple target problem adds a combinatorial problem of associating many measurements with many tracks. To this end, the PDA algorithm was extended to evaluate the joint(JPDAF) measurement-to-track association probabilities (Fortmann *et al.*, 1980). Under JPDAF, multiple association hypotheses are formed after each measurement report and the tracks are then updated by the sum of measurements weighted by their probabilities (measurements beyond a set Mahalanobis distance are ignored through a gating process) (Bar-Shalom and Li, 1995). The JPDAF approach has some undesirable qualities. First, measurements can be shared between targets that can result in track biases and coalescence (Fitzgerald, 1985). Secondly, these PDA filters do not innately initiate or terminate tracks - it is assumed that the tracked targets exist. To address both of these issues, an interacting multiple model filter (IMM, discussed further in the sequel) can be used to improve performance in one of two ways. In the first method, the IMM more accurately predicts target dynamics to prevent track coalescence (Bar-Shalom, 1986; Blom and Bloem, 2002). In the second method, two PDAF algorithms are used in an IMM framework to model the case where the target exists and is observable and the case where the target exists but is unobservable (this second approach is called IMMPDAF) (Bar-Shalom *et al.*, 1989). In an alternate approach, the track existence assumption is not used. Rather, the probability of track existence is integrated with the probability of data association. With this approach the integrated probabilistic data association (IPDA) and joint integrated probabilistic data association (JIPDA) analogs to PDA JPDA were developed(Musicki *et al.*, 1994; Musicki and Evans, 2004).

A more powerful but computationally expensive approach is the Multiple-Hypothesis Tracking (MHT) algorithm. The MHT algorithm enumerates all possible assignment possibilities and evaluates the target likelihood given the measurements (Reid, 1979). This approach is optimal in the sense that it offers a complete evaluation of the association problem. However, as the number of measurements and targets increase, the cost of the algorithm is exponential. As a result, practical implementations of the MHT algorithm currently require some suboptimal concessions (Blackman and Popoli, 1999). Despite the complexity, the MHT remains of prime interest to the author because it is an optimal approach, while the suboptimal concessions will diminish with the evolving speed of computing hardware. The MHT method is regarded as the best available approach to tracking multiple targets (Blackman, 2004). Therefore, the MHT algorithm was selected as the base on which this thesis offers improvement.

The Multiple Hypothesis Tracker

Given the virtues of MHT,there has been significant effort to bound the association process through approximations to the algorithm (Blackman, 1986; Bar-Shalom and Li, 1995; Pattipati *et al.*, 2000; Blackman, 2004; Bar-Shalom *et al.*, 2007). To this end, the association process in MHT is typically divided in two steps. First, unlikely assignments are removed through a gating process where a validation region is defined around the predicated measurements. Observations outside of this region are rejected as candidates for track continuation. The validation gate is typically defined as an ellipsoidal region based on some maximum Mahalanobis distance using the innovation covariance (Blackman and Popoli, 1999). To account for the event where the true target measurement falls outside of the validation gate resulting in a missed detection,

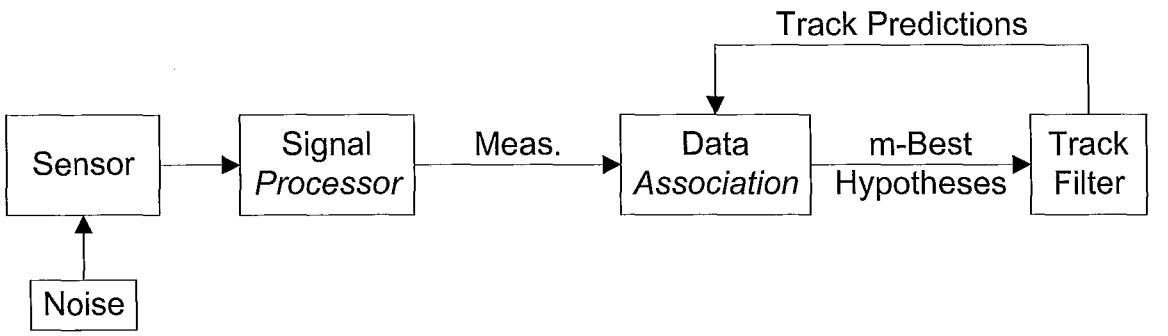


Figure 1.2: MHT Flow Diagram

Li has proposed that the predicted state covariance should be increased (resulting in a larger validation gate in the next round of measurements)(Li, 1998). After gating, the remaining space is mapped as a linear assignment problem using the negative log likelihood as the cost of assigning a given measurement to a specific track (new, existing, or false tracks) (Miller *et al.*, 1995; Cox and Hingorani, 1994; Cox *et al.*, 1997; Miller *et al.*, 1997). Common methods used to bound this space include clustering, pruning of hypotheses and tracks, and track merging (Blackman and Popoli, 1999). In addition, an efficient search algorithm for finding the N-best assignments based on Murty's method has been developed (Cox and Hingorani, 1994) that results in a MHT flow diagram like that shown in figure 1.2. A lucid description of this algorithm is given in 6.7.2 of (Blackman and Popoli, 1999).

In implementing a N-best scheme, it is important to note that the likelihood of a measurement conditioned on any of the three possible assignments is not a dimensionless value. Therefore, one must be careful in comparing global likelihoods between hypotheses containing differing number of measurement associations (Bar-Shalom and Li, 1995). An eloquent description of MHT with guidance on this issue is provided by Bar-Shalom, Blackman, and Fitzgerald (Bar-Shalom *et al.*, 2007).

To make such comparisons possible, these authors propose the use of dimensionless likelihood ratios.

The measurements used in MHT are typically given as a position within the surveillance volume (area in an imaging sensor). However, in the case of an imaging sensor, additional, feature-based target information is available (e.g., shape, intensity, color). The inclusion of features, attributes, and classification in Bayesian tracking is fundamentally described in a series of papers by Drummond(Drummond, 2000, 2001, 2002, 2004, 2005). Therein, features are defined as measurements other than kinematic which come from a continuous sample space. This is in contrast with parameters which come from a discrete sample space. If the features are used strictly to describe the target state, then the process is called Feature-Aided Tracking(Drummond, 2000). In this case, the features augment the state vector and are filtered similarly to the kinematic data. However, if there exists a database that can cross-reference target types based on features, then it is possible to use these features in Classification-Aided Tracking (Drummond, 2000, 2001; Lancaster and Blackman, 2006). In both approaches, the above mentioned references indicate that the feature information can improve the gating and assignment of the data association process. Heuristic rules for the use of features such as supplemental correlation matching to aid the MHT data association have also been proposed (Tissainayagam and Suter, 2001). Further, the use of feature-dependent motion models within MHT was proposed (Cox and Leonard, 1994). These rules, while useful in some cases, sit as external add-on's to the MHT algorithm.

1.2.3 Filtering and Prediction

The optimal estimation of a target's state and the prediction of this state at a future time is typically pursued using a Bayesian approach (Blackman and Popoli, 1999; Bar-Shalom *et al.*, 2001; Crassidis and Junkins, 2004). For the time invariant case, the invariant state is specifically referred to as a parameter. In the Bayesian approach, this parameter is treated as a random variable described by a prior probability density function (pdf). A measurement is then a realization of the parameter at some point in its pdf (Bar-Shalom *et al.*, 2001); the measurement is of the parameter corrupted by noise. It is well known that if the measurement noise has a Gaussian distribution, the optimal estimate is given by the linear least squares approach. A recursive form of this estimator is available to filter the estimate over time as more measurements become available (Bar-Shalom *et al.*, 2001). Given the standoff ranges and slow illumination changes typical in many airborne surveillance scenarios, the principal component features described in 1.2.1 are assumed to be time invariant target parameters in this thesis and are filtered over time using a recursive least squares estimate.

For the case of a time varying state, a process noise pdf is modeled in addition to the measurement noise. For example, a target with near constant velocity can be modeled by having position and velocity states that are random variables with pdf's defined by the process noise on the acceleration. If the measurements are a linear combination of the target state and if the target acceleration is zero mean, then the the target state at any time is a function of the previous state and the change in velocity. Further, if the measurement and process noises are independent and Gaussian, the optimal estimate of such a target is given by the Kalman filter (Bar-Shalom *et al.*, 2001). For nonlinear measurements or target dynamics, the extended Kalman filter

(EKF) is available. In the EKF, the measurements and target dynamics are linearized over the observation interval (Bar-Shalom *et al.*, 2001). In the case where the target has multiple modes of motion, such as an aircraft on take off, cruise, and landing, it is possible to use multiple Kalman filters. Each filter models the process noise of each motion mode. The interacting multiple model filter (IMM) offers an efficient means to combine these Kalman filters according to a Markov model describing the motion mode transitions (Bar-Shalom *et al.*, 2001).

Bayesian filtering takes the prior pdf and updates with information of the measurement. That is, the probability of the prior target state is combined with the likelihood of the measurement to give the posterior distribution of the target state. Given that this posterior state is true, it is possible to predict the target state and prior pdf at the next measurement interval. By predicting the prior pdf, it is also possible to calculate the likelihood of the next round of measurements. Gating, by placing a limit on the measurement likelihood, allows unlikely measurements to be removed from the data association problem described in section 1.2.2 (Blackman and Popoli, 1999; Bar-Shalom *et al.*, 2001).

1.3 Problem Statement

This thesis presents a probabilistic framework to include feature information in the MHT data association problem. In addition, it is shown that including the information available from the MHT predictions improves the feature extraction from the next round of measurements. Using this approach, the signal processor adds information to the prior probability of detection of a given track as indicated by Figure 1.3. The addition of feature information increases the effectiveness of the gating operation.

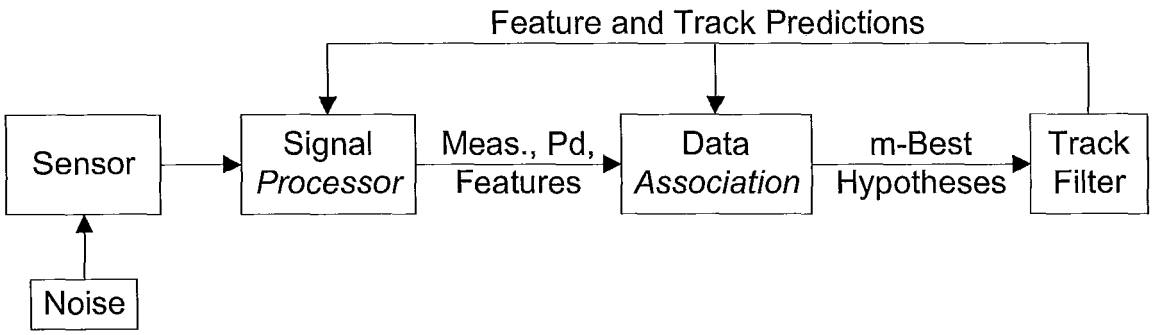


Figure 1.3: JMHT Flow Diagram

Moreover, by calculating the prior detection probability, the performance of tracking coasting targets is significantly improved. The interactions between the signal processor and the target tracker is denoted as a judicious behavior and as such the altered MHT is referred to as the JMHT. This thesis builds on the concepts presented earlier(McAnanama and Kirubarajan, 2009) by inclusion of the improvements from a calculated prior detection probability.

The format of this thesis is as follows. In section 2 the standard Multiple Hypothesis Tracker is reviewed. Section 3 describes contrast based feature extraction with the requirements of a feature for use in a judicious multiple hypothesis tracker and the use of principal component analysis (PCA) in particular. Further, a method to determine the likelihood that the principal components of a given measurement are from a tracked target is presented. Finally, a method is developed on how to gauge the prior probability of detection based on the segmentation method.

Section 4.1 presents the JMHT algorithm with a method on how to extend the MHT to include PCA feature information. Direction is also provided on how to make the prior detection probability an adaptive parameter so as to significantly improve the performance of the tracker. Finally the dimensionless likelihoods required for the

data association problem are derived.

In section 5 we present the results of JMHT through a simulation. This is followed by a final summary and closing remarks in the conclusion, section 6.

Chapter 2

Review of the Multiple Hypothesis Tracker

The Multiple Hypothesis Tracking algorithm developed by Reid (Reid, 1979; Bar-Shalom and Li, 1995; Blackman and Popoli, 1999) is a hypothesis-orientated assignment problem where an enumeration is made of all possible origins of every measurement. A complete hypothesis assigns each measurement to at most one target. A joint event is hypothesized (a global hypothesis) where each measurement is assigned as a new track, a continuation of a previous track (from a previous global hypothesis), or as a false alarm. That is, a global hypothesis is formed from a parent hypothesis and a current association hypothesis. This joint event is made cumulative, conditioned upon the sequence of measurements up to the current time k. The posterior

probability of this cumulative joint event is given by (Blackman and Popoli, 1999)

$$\begin{aligned} P\left(\Theta^{k,l}|Z^k\right) &= P\left(\theta(k), \Theta^{k-1,s}|Z^k\right) \\ &= \frac{1}{c} p\left(Z(k)|\theta(k), \Theta^{k-1,s}, Z^{k-1}\right) \\ &\cdot P\left(\theta(k)|\Theta^{k-1,s}, Z^{k-1}\right) P\left(\Theta^{k-1,s}|Z^{k-1}\right), \end{aligned} \tag{2.1}$$

where $\Theta^{k,l}$ is the event of the global hypothesis l through to time k, Z^k is the cumulative measurements through to time k, $\theta(k)$ is the current association event, c is the factor from Bayes' rule (the sum of the numerator over all possible parent and association hypotheses) that normalizes the distribution, and $\Theta^{k-1,s}$ is the event of the parent global hypothesis s through to time $k-1$. The three functions on the right hand side of (2.1) represent, from left to right, the likelihood of a current association event, the prior probability of a current association event, and the probability of the parent hypothesis.

The likelihood of a current association event depends on the type of association. For association with an existing track t_i, the likelihood is given by the pdf of the innovation (measurement residual) with the notation $f_{t_i}[z_i(k)]$ for measurement $z_i(k)$. The pdf of the likelihood of a false alarm is considered uniform in the surveillance area A (i.e., A^{-1}). Likewise, the pdf of a new target likelihood is taken as A^{-1}.

The prior probability of a current event is given by the joint pdf of the association event, the detection events and the extraneous measurement events. Where the association event comes from the product of the permutations of track to measurement associations and the combinations of possible new target associations (the permutation is required as each track has an identity) (Bar-Shalom and Li, 1995). While the

target detections can be considered as positive results of a Bernoulli Trial (a binomial distribution). However, if the probability of a new target measurement or new false measurement is low within the surveillance area, the binomial nature of these events will be approximated by Poisson distributions (Papoulis, 2001). It is assumed that the probability of detecting measurements from existing tracks, new measurements, and false measurements are independent.

Substitution of the likelihoods and priors as define above into 2.1 results in the following form of the global hypothesis

$$P\left(\Theta^{k,l}|Z^k\right) = \frac{1}{c}\frac{\phi!\nu!}{m(k)!}\mu_F(\phi)\mu_N(\nu)A^{-\phi-\nu}\prod_{i=1}^{m(k)}\{f_{t_i}[z_i(k)]\}^{\tau_i} \cdot \prod_t (P_D^t)^{\delta_t}(1-P_D^t)^{1-\delta_t}P\left(\Theta^{k-1,s}|Z^{k-1}\right). \quad (2.2)$$

Where all constant terms have been incorporated into c, the number of false measurements is given by ϕ, ν is the number of new targets, while the total number of measurements at time k is given by $m(k)$. The continuation likelihood uses τ_i, an indicator variable, which is equal to 1 if $z_i(k)$ originated from a previous track, otherwise equal to 0. Similarly, the detection probability enters with the indicator variable δ_t (equal to 1 for track detection or 0 otherwise) and the probability of detection of track t given by P_D^t. The pmf for the number of false and new measurements is given by $\mu_F(\phi)$ and $\mu_N(\nu)$ respectively. As noted earlier, these probabilities are assumed Poisson distributed and are assigned spatial densities λ_ϕ and λ_ν. By including the Poisson representation for the extraneous targets, Reid was able to cancel the dependence on the surveillance volume with the following result

$$P\left(\Theta^{k,l}|Z^k\right) = \frac{1}{c'}(\lambda_\phi)^\phi(\lambda_\nu)^\nu \prod_{i=1}^{m(k)}\{f_{t_i}[z_i(k)]\}^{\tau_i} \cdot \prod_t (P_D^t)^{\delta_t}(1-P_D^t)^{1-\delta_t} P\left(\Theta^{k-1,s}|Z^{k-1}\right). \tag{2.3}$$

The global hypotheses are typically compared using a cost function based on the negative log-likelihoods of each assignment. That is, the normalizing constant is dropped and comparisons are made between the association likelihoods in the rhs of the argument

$$P\left(\Theta^{k,l}|Z^k\right) \propto (\lambda_\phi)^\phi(\lambda_\nu)^\nu \prod_{i=1}^{m(k)}\{f_{t_i}[z_i(k)]\}^{\tau_i} \cdot \prod_t (P_D^t)^{\delta_t}(1-P_D^t)^{1-\delta_t} P\left(\Theta^{k-1,s}|Z^{k-1}\right). \tag{2.4}$$

However, the likelihood of a measurement conditioned on any of the three possible assignments is not a dimensionless value. Therefore, it is not valid to directly compare global likelihoods between hypotheses containing differing number of measurement associations (Bar-Shalom and Li, 1995). This issue can be overcome by using dimensionless likelihood ratios(Bar-Shalom *et al.*, 2007). These ratios are obtained by dividing the rhs of (2.4) by the likelihood that all measurements are false. That is, as

$$\phi + \nu + \sum_{i=1}^{m(k)} \tau_i = m(k), \tag{2.5}$$

then

$$(\lambda_\phi)^{m(k)} = \lambda_\phi^{\phi+\nu+\Sigma_{i=1}^{m(k)}\tau_i}, \tag{2.6}$$

resulting in

$$P\left(\Theta^{k,l}|Z^k\right) \propto \left(\frac{\lambda_\nu}{\lambda_\phi}\right)^{\nu} \prod_{i=1}^{m(k)} \left\{\frac{f_{t_i}[z_i(k)]}{\lambda_\phi}\right\}^{\tau_i} \cdot \prod_t (P_D^t)^{\delta_t}(1-P_D^t)^{1-\delta_t} P\left(\Theta^{k-1,s}|Z^{k-1}\right). \tag{2.7}$$

That is, the likelihood ratio for the continuation of a track t through measurement i is

$$\mathcal{L}_{ti} = \left\{\frac{f_{t_i}[z_i(k)]}{\lambda_\phi}\right\} P_{D_t}(k), \tag{2.8}$$

the likelihood ratio for a missed detection of track t at time k is

$$\mathcal{L}_{t0} = 1 - P_{D_t}(k), \tag{2.9}$$

and the likelihood ratio of a new target versus a false measurement is

$$\mathcal{L}_{\nu/\phi} = \frac{\lambda_\nu}{\lambda_\phi}. \tag{2.10}$$

Moreover, the log of (2.7) with the substitution of (2.8) to (2.10) results in a summation of the log-likelihood ratios (LLR) and the log of the parent hypothesis probability. Assignment costs are then created with the negative log-likelihood ratios (NLLR) of these separable, track independent, variables.

The best global hypotheses have the lowest total costs. There are several methods for ranking the best hypotheses in the solution space from exhaustive searching to optimized methods (Pattipati *et al.*, 2000). In this thesis, Murty's Ranked Assignment Method with the JVC algorithm was employed to find the k-best, 2-D assignments (Miller *et al.*, 1997).

Chapter 3

Feature Extraction

The two-way interaction between the sensor processing and the MHT tracking proposed here is via the feed-forward of feature information and track detection probabilities and via the feedback of feature and kinematic state estimations. This interaction is what we term as the judicious nature of JMHT. The feed-forward interaction adds information to the prior and makes the association likelihood critical of the joint event of the feature and location of a measurement being of a given track. Features that lend themselves to our approach should be static during the tracking interval such that the observations can be filtered using parameter estimation. For example, the principal components of a target boundary, region, or multi-spectral content (e.g., color, multi-spectral IR) may be useful features. Principal components are a rotational invariant measure of the maximum covariance within the distribution of a random variable. In this paper the PCA of a target's pixel coordinates are used. It is assumed that, at a reasonable standoff range and illumination level, the frame-to-frame change in the scale of a target will be small such that the pixel coordinate PCA can be estimated as a parameter. This estimation is then accomplished recursively by segmenting each

image and determining the principal components of each connected pixel group (the target measurements). In this section we review a common segmentation method and describe the frame-to-frame parameter estimation of a target PCA.

3.1 Contrast Based Segmentation

Target features are commonly taken from the boarder (e.g., corners, edges) or within the region of the target (e.g., color, intensity)(Trucco and Plakas, 2006). One computationally simple extraction method is to segment foreground from background based on pixel intensity (with a gray-level threshold) where each target is based on the connectivity of foreground pixels. That is, let the N pixels of an image be represented in L gray levels $[1, 2, \ldots, L]$ with n_j pixels in each gray level (i.e., $N = n_1 + n_2 + \cdots + n_L$). Then a threshold, κ, segments the image such that targets pixels are identified as $n_j, j > \kappa$. A well known method of determining the best threshold, κ, between targets and background is Otsu's wherein the image is dichotomized with the objective of maximizing the gray level variance between targets and background(Sonka *et al.*, 2007; Otsu, 1979),

$$\max_{\kappa} \sigma_B^2(\kappa) = \frac{[\mu_T \omega(\kappa) - \mu(\kappa)]^2}{\omega(\kappa)[1 - \omega(\kappa)]}, \tag{3.1}$$

where $\omega(\kappa) = \sum_{j=1}^{L} p_j$ and $\mu(\kappa) = \sum_{j=1}^{L} j p_j$ are the zeroth- and first-order cumulative moments up to the κth level of the normalized histogram (with pmf $pj = n_j/N$) and $\mu_T = \mu(L) = \sum_{j=1}^{L} j p_j$ is the image mean intensity level. In addition to an optimal threshold, Otsu's method provides a "goodness" of the threshold given by

$$\eta = \frac{\sigma_B^2}{\sigma_T^2} \quad (0 \leq \eta \leq 1), \tag{3.2}$$

where $\sigma_T^2 = \sum_{j=1}^{L} (j - \mu_t)^2 p_j$ is the total variance of intensity levels. The lower bound of η occurs only when for an image with a single gray level and the upper bound occurs with two gray levels.

Target detection is the joint event of successfully segmenting targets from background and the intensity of the targets being above the threshold. We assume that the gray level distribution of existing tracks is captured by the histogram of the underlying target (with a normalized pmf given by $p_{jt} = n_{jt}/N_t$). As a result, we can express the detection probability by using Otsu's method as $P_D = \eta \cdot (1 - p_{\kappa t})$. In Section 4.1, we will show how to create an adaptive method for determining the prior probability of detecting a tracked target based on this P_D calculation.

3.2 Regional Description by Principal Component Analysis

Principal component analysis is a rotationally invariant method of describing the the region of a target measurement(Gonzalez and Woods, 2002; Gonzalez *et al.*, 2003; Sonka *et al.*, 2007). The PCA of a region is determined by forming a set of 2-D vectors from the coordinates of each pixel within the region. Each set is treated as a sample of n random vectors, $\boldsymbol{u} = (a, b)^T$, with mean

$$\bar{\boldsymbol{u}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}_i, \tag{3.3}$$

and sample covariance

$$\boldsymbol{C_u} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{u}_i - \bar{\boldsymbol{u}})(\boldsymbol{u}_i - \bar{\boldsymbol{u}})^T. \tag{3.4}$$

Next, let $\boldsymbol{A}$ be a matrix with rows formed from the eigenvectors of $\boldsymbol{C_u}$ in descending order of eigenvalue. The principal components of the sample are then found by applying the *Hotelling transform* $\boldsymbol{v} = \boldsymbol{A}(\boldsymbol{u} - \bar{\boldsymbol{u}})$, with a mean value $\bar{\boldsymbol{v}} = E\{\boldsymbol{v}\} = \boldsymbol{0}$ and transformed covariance $\boldsymbol{C_v} = \boldsymbol{AC_uA^T}$. An example of a segmented target with an ellipse plotted along the target's pixel region PCA is show in figure 3.1.

As the covariance matrix is square and symmetric, the transformation matrix $\boldsymbol{A}$ is an orthogonal basis for the column space of $\boldsymbol{C_u}$. Therefore the transformation is a rotation such that $\boldsymbol{C_v}$ is a diagonal matrix made from the eigenvalues of $\boldsymbol{C_u}$. These eigenvalues, denoted ℓ_i, are the sample estimate of the population's eigenvalues λ_i (i.e., as $\boldsymbol{C_u}$ is an estimator of the population's covariance $\boldsymbol{\Sigma_u}$). That is, there exists a principal component decomposition $\boldsymbol{\Sigma_v} = \boldsymbol{\Gamma\Sigma_u\Gamma}^T$, where $\boldsymbol{\Gamma}$ is a matrix made from the eigenvectors of $\boldsymbol{\Sigma_u}$. Our goal is to recursively estimate λ_i while considering the likelihood that the measured ℓ_i is from λ_i.

Indeed, of primary importance in video tracking is to identify a successful sampling of a target in each frame of video. To this end, we make the assumption that $\boldsymbol{u}(k)$ for a given target is distributed as $\mathcal{N}(\boldsymbol{\mu_u}, \boldsymbol{\Sigma_u})$, where $\boldsymbol{\mu_u}$ is the population mean. The translation through PCA gives $\boldsymbol{v}(k) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_v})$. In addition, we assume that the targets are sufficiently large to make the arguments of the Central Limit Theorem. As shown by Girshick (Morris, 2007; Härdle and Simar, 2007; Girshick, 1939), under

the stated assumptions,

$$\frac{\ell_j - \lambda_j}{\lambda_j\sqrt{\frac{2}{n-1}}} \sim \mathcal{N}(0,1). \tag{3.5}$$

In addition, the *log* transformation of Equation (3.5)results in

$$\log \ell_j - \log \lambda_j \sim \mathcal{N}\left(0, \frac{2}{n-1}\right). \tag{3.6}$$

Using this result, we will consider each sample of the population as a measurement in the form of

$$\log(\boldsymbol{\ell}(k)) = log(\boldsymbol{\lambda}) + \boldsymbol{w}(k), \tag{3.7}$$

where $\boldsymbol{\ell} = (\ell_1, \ell_2)^T$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$ are the eigenvalues of the covariance matrices, k is the frame index and $\boldsymbol{w}(k)$ is the measurement noise distributed as $\mathcal{N}\left(0, \boldsymbol{I}\left(\frac{2}{n(k)-1}\right)\right)$. A recursive, least squares estimate of the parameter $\boldsymbol{\lambda}$ is given byBar-Shalom *et al.* (2001)

$$\begin{aligned}\log(\hat{\boldsymbol{\lambda}}(k+1)) =& \log(\hat{\boldsymbol{\lambda}}(k)) \\ &+ \boldsymbol{W}(k+1)[log(\boldsymbol{\ell}(k+1)) - \log(\hat{\boldsymbol{\lambda}}(k))].\end{aligned} \tag{3.8}$$

The matrix $\boldsymbol{W}(k+1)$ is the update gain given by

$$\boldsymbol{W}(k+1) = \boldsymbol{P}(k)\boldsymbol{S}(k+1)^{-1}, \tag{3.9}$$

where $\boldsymbol{P}(k)$ is the covariance of the estimate given by

$$\boldsymbol{P}(0) = \boldsymbol{I}\left(\frac{2}{n(0)-1}\right) \tag{3.10}$$

$$\boldsymbol{P}(k+1) = \boldsymbol{P}(k) - \boldsymbol{W}(k+1)\boldsymbol{S}(k+1)\boldsymbol{W}(k+1)^T, \tag{3.11}$$

and $\boldsymbol{S}$ is the residual covariance given by

$$\boldsymbol{S}(k+1) = \boldsymbol{P}(k) + \boldsymbol{I}\left(\frac{2}{n(k+1)-1}\right). \tag{3.12}$$

Using the above, the track continuation likelihood based on the principal components of a measurement in frame $k+1$ versus the track estimated principal components is

$$\begin{aligned}\Lambda_t[z(k)] &= |2\pi\boldsymbol{S}|^{-\frac{1}{2}} \\ &\quad \cdot \exp\left(-\frac{1}{2}[\boldsymbol{\nu}(k+1)]^T[\boldsymbol{S}(k+1)]^{-1}[\boldsymbol{\nu}(k+1)]\right),\end{aligned} \tag{3.13}$$

where $\boldsymbol{\nu}(k+1) = log(\boldsymbol{\ell}(k+1)) - \log(\hat{\boldsymbol{\lambda}}(k))$ is the residual of the parameter estimation.

The likelihood of the principal components of the false alarms and new targets is assumed to be uniformly distributed in $[1, \frac{A}{2}]$ where A is the surveillance area. As the principal components are a measure of the maximum variance possible along any rotated axis within a sample, we set the bounds around the uniform distribution as follows. The minimum bound is set at 1 under the assumption that for a target to have a region-based feature it must be at least two pixels large. Further, as the variance is defined as $\sigma^2 = E((u-\bar{u})(u-\bar{u})^T)$, an approximate but reasonable upper bound of the maximum variance occurs with the centroid of the target in the middle

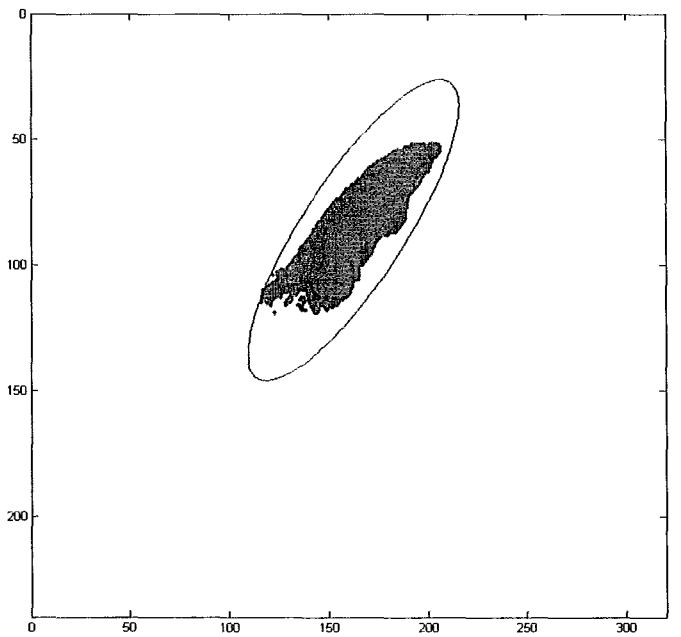
of the surveillance region and the extents of the target at the corners of this region. For a square surveillance area this configuration gives a variance of $\frac{A}{2}$. Further, if $A >> 1$ then the pdf of the likelihood of the PCA features is $\propto 2A^{-1}$.

Therefore, the likelihood of a current association based on the PCA features is given as

$$\begin{aligned} p\left(Z(k)|\theta(k), \Theta^{k-1,s}, Z^{k-1}\right) &= \prod_{i=1}^{m(k)} \{\Lambda_{t_i}[z_i(k)]\}^{\tau_i} \cdot 2A^{(1-\tau_i)} \\ &= 2A^{-\phi-\nu} \prod_{i=1}^{m(k)} \{\Lambda_{t_i}[z_i(k)]\}^{\tau_i}. \end{aligned} \tag{3.14}$$

In Section 4.1 we will use the feature likelihoods and feature-based detection probabilities within the MHT to create the JMHT.

(a) Segmented Image



(b) Principal Components

Figure 3.1: Example of a segmented target and its principal components.

Chapter 4

The Judicious Multiple Hypothesis Tracker

4.1 Adding Judicious Behavior to MHT

Recall from Section 2, the probability of a complete track to measurement association (a global hypothesis) is proportional to the product of the likelihood of a current association event, the prior probability of a current association event, and the probability of the parent hypothesis. The likelihood of a current association event in the judicious case is given as the joint likelihood of the kinematic state and the feature

parameters

$$
\begin{aligned}
& p\left(Z(k)|\theta(k), \Theta^{k-1,s}, Z^{k-1}\right) \\
&= A^{-\phi-\nu} \prod_{i=1}^{m(k)} \{f_{t_i}[z_i(k)]\}^{\tau_i} \cdot 2A^{-\phi-\nu} \prod_{i=1}^{m(k)} \{\Lambda_{t_i}[z_i(k)]\}^{\tau_i} \\
&= 2A^{-2\phi-2\nu} \prod_{i=1}^{m(k)} \{f_{t_i}[z_i(k)] \cdot \Lambda_{t_i}[z_i(k)]\}^{\tau_i}. \qquad (4.1)
\end{aligned}
$$

The prior probability of a current association event is largely unchanged in the judicious scheme. For example, the combinatorics involved in the association are unchanged. Likewise, treating the detections as the result of a Bernoulli Trial remains the same. Further, the Poisson distribution of new and false targets remains valid regardless of feature. For example, the spurious glint off of specular surfaces will give rise to false alarms with apparent features. However, what is available under the judicious scheme is a means to extract an adaptive prior probability of detection. This adaptive P_D is important as the prior probability of detection varies across the surveillance region. For example, the probability of detecting a target with an intensity equal to the search area (i.e. the probability of detecting a target without contrast) is zero. That is, in addition to the PCA information, we can also maintain a record of the tracked target's pixel intensity histogram; this information is available a priori to a measurement. Therefore, an adaptive prior P_D is defined by the application of (3.1) to every pixel location in the surveillance area. Fortunately there is no need to calculate the prior P_D for every pixel. Instead, we only need to know P_D at measurement locations and at the predicted measurement location. The prior P_D at each measurement location is written as $P_D|_{m(k)}$. The prior P_D at the location of the predicted measurement, $\hat{z}$, is written as $P_D|_{\hat{z}_t(k)}$ and is obtained by considering

the segmentation of the region of interest defined by all the pixels n_j that are in

$$\mathcal{R} = \{f_t[z|_{n_j}] \leq \rho\}, \tag{4.2}$$

where ρ is a gating limit that is selected depending on the desired size of the region of interest relative to the measurement accuracy. In our implementation, we have used a bounding box approximation of ρ.

In considering the local segmentation of ρ, the threshold has been set from the global threshold κ. But the goodness of this threshold depends on the pixel values within ρ. If ρ contains some foreground (target) pixels, then the detection probability for a given target depends on the local foreground/background dichotomy of ρ given the global threshold. For example, a given target that is bounded by background pixels is easier to detect than one that is occluded by other foreground pixels. Further, a loss of target contrast is indicated if ρ contains only foreground pixels. The local "goodness" of the threshold is given by equation 3.2 applied only to the pixels within ρ. If, however, there are no foreground pixels in ρ, then the detection probability is set by the global segmentation described in section 3.1. That is, we have no more detection information in the prior other than what is available from the global threshold that lead to ρ containing only background pixels. In this case equation 3.2 is evaluation using the entire image.

An accurate determination of $(1\text{-}P_D|_{\hat{z}_t(k)})$ is desirable as it will allow for a low cost assignment of a missed detection in cases where a target contrast vanishes with the local background or occlusion from crossing targets. Without such an accurate determination, the track may be seduced by new or false measurements. The resulting

prior probability of an association is then,

$$\begin{aligned} & P\left(\theta(k)|\Theta^{k-1,s}, Z^{k-1}\right) \\ & = \frac{\phi!\nu!}{m(k)!}\prod_t (P_D^t|_{m(k)})^{\delta_t}(1-P_D^t|_{\hat{z}_t(k)})^{1-\delta_t} \\ & \cdot e^{-\lambda_\phi A}\frac{(\lambda_\phi A)^\phi}{\phi!}e^{-\lambda_\nu A}\frac{(\lambda_\nu A)^\nu}{\nu!}. \end{aligned} \tag{4.3}$$

By substituting (4.1) and (4.3) into (2.1) and incorporating all of the constant terms into c, we get that the posterior probability of a global hypothesis is

$$\begin{aligned} & P\left(\Theta^{k,l}|Z^k\right) \\ & = \frac{1}{c}\frac{(\lambda_\phi)^\phi}{A^\phi}\frac{(\lambda_\nu)^\nu}{A^\nu}\prod_{i=1}^{m(k)}\{f_{t_i}[z_i(k)]\cdot\Lambda_{t_i}[z_i(k)]\}^{\tau_i} \\ & \cdot\prod_t (P_D^t|_{m(k)})^{\delta_t}(1-P_D^t|_{\hat{z}_t(k)})^{1-\delta_t}P\left(\Theta^{k-1,s}|Z^{k-1}\right). \end{aligned} \tag{4.4}$$

To create a separable cost function we follow the work of Bar-Shalom et al. (Bar-Shalom *et al.*, 2007). and divide by the likelihood that all measurements are false. The false detection likelihood is $\frac{\lambda_\phi}{A}$ and so we divide (4.4) by

$$\frac{(\lambda_\phi)^{m(k)}}{A^{m(k)}} = \frac{\lambda_\phi^{\phi+\nu+\Sigma_{i=1}^{m(k)}\tau_i}}{A^{\phi+\nu+\Sigma_{i=1}^{m(k)}\tau_i}}. \tag{4.5}$$

By incorporating $\frac{A^{m(k)}}{(\lambda_\phi)^{m(k)}}$ into c we get

$$\begin{aligned} P\left(\Theta^{k,l}|Z^k\right) & = \frac{1}{c}\left(\frac{\lambda_\nu}{\lambda_\phi}\right)^\nu\prod_{i=1}^{m(k)}\left\{\frac{f_{t_i}[z_i(k)]\cdot\Lambda_{t_i}[z_i(k)]\cdot A}{\lambda_\phi}\right\}^{\tau_i} \\ & \cdot\prod_t (P_D^t|_{m(k)})^{\delta_t}(1-P_D^t|_{\hat{z}_t(k)})^{1-\delta_t}P\left(\Theta^{k-1,s}|Z^{k-1}\right). \end{aligned} \tag{4.6}$$

From the above, the track continuation likelihood ratio of a track t through measurement i is

$$\mathcal{L}_{ti} = \left\{ \frac{f_{t_i}[z_i(k)] \cdot \Lambda_{t_i}[z_i(k)] \cdot A}{\lambda_\phi} \right\} P_D^t(k)|_i, \tag{4.7}$$

the likelihood ratio for a missed detection of track t at time k is

$$\mathcal{L}_{t0} = 1 - P_{D_t}(k)|_{\hat{z}_t(k)}, \tag{4.8}$$

and the likelihood ratio of a new target versus a false measurement is

$$\mathcal{L}_{\nu/\phi} = \frac{\lambda_\nu}{\lambda_\phi}. \tag{4.9}$$

Note that the residual covariance in the estimation of the principal components in (3.12) is $\propto \frac{1}{n}$. As a result the physical dimension of $\Lambda_{t_i}[z_i(k)]$ is the same as A^{-1}. Thus (4.7) is indeed dimensionless. As we have a separable, dimensionless score function we can use the NLLR values and employ any of the assignment methods discussed in Section 2 such as Murty's Ranked Assignment Method with the JVC algorithm to find the k-best, 2-D assignments(Miller *et al.*, 1997). Each of these hypothesis are filtered to update the PCA parameter estimates as described in Section 3.2. In addition, the kinematic state can be updated with any of the commonly used tracking filters such as the interacting multiple model (IMM) estimator (Bar-Shalom *et al.*, 2001, 2007; Yeddanapudi *et al.*, 1997; Wang *et al.*, 1999). The use of IMM is straightforward but with attention given to the correlation between the feature information and kinematic measurements. For example, the parameter estimation of PCA completed on pixel location is correlated with the position measurement noise and uncorrelated with the process noise. That is, the PCA is the estimate of the population covariance rotated

to align with the axes of maximum variance. Therefore, the PCA estimate should be used to generate the target centroid measurement covariance.

4.2 Kinematic Filter

The tracking filter implemented in our study is an Interacting Multiple Model filter (IMM) for which there are standard discussions on its derivation(Bar-Shalom *et al.*, 2001) and implementation(Blackman and Popoli, 1999). A discussion is presented here only on the details relevant to the handling of the correlation with the feature estimate and for the improvements available from the proposed judicious interaction scheme.

The multiple models in IMM refer to a set of kinematic models that are chosen to best represent the aggregate target dynamics. The estimates from each of these models is mixed to create an initial condition for each model filter based on the probability that the given model was in effect at the previous iteration. Each of these augmented priors is then run through its own Kalman filter to account for state and measurement data. The outputs of these filters are combined into an aggregate state estimate based on the posterior likelihoods. The mixing probabilities for the next round are generated from the model likelihoods and the probability of switching from one model to another. The portion of the IMM that is specifically relevant to our discussion occurs at the level of the Kalman filters.

4.2.1 Kalman Filter Update

The models of the IMM differ in their kinematic models. However, they can be describe at an abstract level as follows. The target state is described by a discrete Markov process as

$$\boldsymbol{x}(k+1) = \boldsymbol{\Phi}\boldsymbol{x}(k) + \boldsymbol{q}(k) + \boldsymbol{f}(k+1|k), \tag{4.10}$$

where $\boldsymbol{x}$ is the n-dimensional state vector, $\boldsymbol{\Phi}$ is an assumed state transition matrix, $\boldsymbol{q}(k)$ is the process noise assumed distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ with an assumed covariance matrix $\boldsymbol{Q}$, and $\boldsymbol{f}(k+1|k)$ is the relative change in position of the sensor.(Blackman and Popoli, 1999)

Measurements are assumed to be observations of M state variables corrupted by noise uncorrelated with the process noise, given as

$$\boldsymbol{z}(k) = \boldsymbol{H}\boldsymbol{x}(k) + \boldsymbol{v}(k), \tag{4.11}$$

where $\boldsymbol{H}$ is M x n measurement matrix and the measurement noise, $\boldsymbol{v}$, is assumed distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$ with known covariance matrix $\boldsymbol{R}$.

Using this state model and measurement equation, the Kalman filter equations

are as follows Blackman and Popoli (1999)

$$\hat{\boldsymbol{x}}(k|k) = \hat{\boldsymbol{x}}(k|k-1) + \boldsymbol{W}(k)[\boldsymbol{z}(k) - \boldsymbol{W}\hat{\boldsymbol{x}}(k|k-1)] \tag{4.12}$$

$$\boldsymbol{W}(k) = \boldsymbol{P}(k|k-1)\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{P}(k|k-1)\boldsymbol{H}' + \boldsymbol{R}]^{-1} \tag{4.13}$$

$$\boldsymbol{P}(k|k) = [\boldsymbol{I} - \boldsymbol{W}(k)\boldsymbol{H}]\boldsymbol{P}(k|k-1)$$

$$\hat{\boldsymbol{x}}(k+1|k) = \Phi\hat{\boldsymbol{x}}(k|k) + \boldsymbol{f}(k+1|k) \tag{4.14}$$

$$\boldsymbol{P}(k+1|k) = \Phi\boldsymbol{P}(k|k)\Phi' + \boldsymbol{Q} \tag{4.15}$$

Under the judicious interaction suggested herein, the feature information is modeled as a target region estimation described in section 3.2. The target region is defined about its centroid. This centroid is also used as the kinematic position measurement Therefore, while the feature parameter noise is assumed uncorrelated with the process noise, it is correlated with the measurements. Therefore, the measurement covariance $\boldsymbol{R}$ is given from the target pixel covariance $\boldsymbol{C}_u$. For example, if the video pixels are related to target state coordinate system by a linear transformation matrix $\boldsymbol{A}$ such that $\boldsymbol{z} = \boldsymbol{A} \cdot \boldsymbol{u}$, then $\boldsymbol{R} = \boldsymbol{A}\boldsymbol{C}_u\boldsymbol{A}'$.

4.2.2 Measurement Prediction and Gating

Given the state prediction of 4.14, a predicted measurement for the next scan is given by $\hat{\boldsymbol{z}}(k+1|k) = \boldsymbol{H}\hat{\boldsymbol{x}}(k+1|k)$. Feeding this predicted position back to the signal processor allows the judicious filter to determine the prior P_D for the tracked target as discussed in section 4.1. Measurements from this scan(now referred to as scan k) are then fed into the data association engine . If a target measurement is associated with track continuation, then a measurement residual is given by $(z) - \boldsymbol{H}\hat{\boldsymbol{x}}(k|k-1)$.

Under the linear Gaussian assumption of the Kalman filter, the residuals (known as innovations), are distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{S})$, where the covariance matrix is $\boldsymbol{S} = \boldsymbol{HPH'} + \boldsymbol{R}$. The track continuation likelihoods, $f_{t_i}[z_i(k)]$, are given by the likelihood of a measurement under the distribution of the innovation. A gate can then be defined by a set Mahalanobis distance about the innovation. As mentioned in the introduction, such gating is useful in managing the assignment problem.

A gating scheme limits the number of measurements for the data association problem based on a minimum likelihood. For the event that the true measurement is contained in the validation region, gating is an effective way of eliminating unlikely measurements. If the target dynamics are known a priori, then it is possible to account for different maneuvers by using appropriate models within the IMM filters. Using this approach, there will be a high likelihood that the true measurement will sit within the gate. However, in the case of unknown targets, the IMM can be setup to cover likely scenarios only. As a result, it is possible to be faced with the event that the true measurement is beyond the gate. That is, if the kinematic estimation error is large, gating can result in the true measurement being excluded from the data association problem. When the true target behavior diverges from the kinematic models, gating can result in track failure. Therefore, as noted by Blackman (Blackman and Popoli, 1999) and developed by Li (Li, 1998), the track covariance should grow when no track continuation measurement is found. Li's conclusion is based on the fact that gating truncates the pdf of the innovation, resulting in a gated innovation covariance that is a scaled reduction of the true covariance, $S_G = c_T S$. Where Li's scale factor c_T for a 2-D measurement is

$$c_T = 1 - \frac{\frac{\gamma}{2} e^{-\frac{\gamma}{2}}}{1 - e^{-\frac{\gamma}{2}}} \tag{4.16}$$

When a measurement is not associated with a track, the state covariance prediction should then grow weighted by the probability that the true measurement was beyond the gate (and not just undetected within the gate)(Li, 1998),

$$P(k|k-1)_{t0} = P(k|k-1) + \frac{P_D P_G (1 - c_T)}{1 - P_D P_G} W(k) S(k) W(k)'. \tag{4.17}$$

In the case of video tracking, sampling intervals can run at the frame rate. As a result the covariance growth of 4.17 is significant when the probability of detection is high. This behavior is desirable when tracking a target that is truly maneuvering beyond the predicted gate. However, in the case where the modeled probability of detection is overstated and the true target is within the gate but not detected, this covariance growth is undesirable as it dilutes the likelihood when the target is finally detected. Therefore, the the judicious evaluation of P_D described in 4.1 is desirable as it regulates the rate of covariance growth. The performance of using a set P_D to an that of an adaptive P_D for track coasting conditions is compared in section 5.

Chapter 5

Results

5.1 Improvements to the MHT

To demonstrate the benefits of the judicious interactions between the signal processor and MHT tracker two synthetic video segments were created to simulate tracking through clutter and coasting through an area of vanishing target contrast. The merits of the adaptive P_D, use of features, and combined use of these were evaluated over a series of 1000 monte carlo simulations.

5.1.1 Tracking Through Clutter

In the first synthetic segment, four targets maneuver through a cluttered area such that the targets are occluded by each other and by the clutter. The clutter area consists of randomly placed point targets that are dilated into disks and blurred in each frame. A sample frame showing the four targets, their routes and the nature of the clutter is shown in figure 5.1. The measurement reports for all frames from one

Table 5.1: Results of 1000 Monte Carlo Simulations.

	Average Over the Final 100 Frames		
Algorithm	RMS Error $[pixels^2]$	True Tracks [%]	False Tracks [%]
MHT	140	47	29.1
MHT_P	79	58	18.7
MHT_F	49	72	0.8
JMHT	31	82	0.4

simulation are plotted in figure 5.2. Each simulation generated a new set of video frames. The clutter was random in each frame of each run. Further, the motion for each target had a specified tangent and axial noise. This noise varied with one standard deviation across all of the simulations. The average RMS pixel error, % true and % false tracks for the simulations are summarized in Table 5.1. Where a true track is one where the tracking algorithm correctly associates a measurement to a track. A false track is one where the tracking algorithm associates a measurement to the incorrect track. The balance of tracks are coasting without measurement association. As the most difficult portion of tracking is through the cluttered region, only the final 100 frames of each simulation are used in the tabled average. The results across all frames are plotted below. In the following discussion, MHT refers to the classical algorithm, MHT_P refers to MHT augmented with an adaptive P_D, MHT_F refers to MHT augmented with feature information, and JMHT refers to the extension of MHT using both features and an adaptive P_D.

RMS Error

The average RMS tracking errors (the pixel error from track state estimate to ground truth) across the monte carlo simulations are shown in Figure 5.3. By comparing these plots and observing the summary data in Table 5.1, it is apparent that the tracking error is the highest with the classical MHT algorithm. The error is reduced by 44% when the adaptive probability is included and by 58% when feature information is used. A symbiotic error reduction of 78% is achieved when both feature and P_D information are used in the JMHT case.

True and False Tracks

The average percent of true and false tracks per frame across the monte carlo simulations are shown in Figures 5.4 to 5.7. Again, by comparing these plots and observing the summary data in Table 5.1, it is apparent that the number of true tracks is lowest and false tracks highest with the classical MHT algorithm. The true tracks are improved by 23% with the adaptive probability, and by 53% when feature information is used. While the corresponding reductions in false tracks is 36% and 97% respectively. The JMHT algorithm improves the true and false tracks over the MHT algorithm by 74% and 99% respectively.

Discussion

The classical MHT method had the lowest tracking performance on all targets. These results are expected due to the mutual occlusion as the targets pass each other in the presence of clutter. However, it is possible to increase the performance of kinematic-only tracking if the true P_D is calculated by the signal processor and included in

the prior. The reason for the improvement is due to the consideration of a loss of P_D when a target is occluded. As shown in Figure 5.8, the calculated P_D can drop significantly depending on the local target background (i.e. the true background and mutual occlusion of targets and clutter). A low P_D allows the tracker to correctly assign a target coasting event as the most probable.

The improvements for tracking with region PCA feature and kinematic information, but with a set $P_D = 0.9$ are most evident in the number of false tracks. When there is enough information in the PCA features, the tracker can correctly discriminate against extraneous measurements. However, an artificially high P_D and availability of somewhat similar clutter measurements can result in false assignments. This effect is seen by comparing the number of false tracks in the middle of the cluttered region as shown near frame 250 in Figures 5.6b and 5.6b. However, the performance can be increased by running the full judicious version of the tracker with both feature and P_D information.

5.1.2 Tracking Through Occlusion

In the second segment, a single target moves through a censored area such that the target is completely obscured for a long period of the track life. A sample frame showing the target, target route and the nature of the obscuration is shown in figure 5.9. The calculated prior P_D under the judicious scheme is shown in figure 5.10. The calculated probability of detection is high away from the censored area and very low within this area. These results are as expected given the contrast of the target and background. The low P_D within the censored area implies that a missed detection is likely due a lack of detection rather than the event that the target outmaneuvered

our gating function. As a result, the state covariance is not forced to grow due this unlikely event. However, a small increase in the covariance naturally occurs from the filter iteration without a measurement. The slow growth in state covariance results in a slow growth in the innovation covariance and resulting gate size. The gate size is proportional to the square root of the innovation covariance eigenvalues. These values are shown in figure 5.11. This slow increase in the gate size is desired as it allows a target with a true low P_D to coast for long periods while maintaining a meaningful gating function. The tracking results are shown in figure 5.12. Using this approach, the tracker can coast for long periods while still remaining critical of targets emerging from the censored area.

However, if the detection probability is set as an algorithm design parameter then the covariance growth through the censored area may not be valid. For example, setting $P_D = 0.9$ for the scenario described here results in a rapid growth of the state covariance. As shown in figure 5.13, the growth of the validation gate reaches an exponential rate within the censored area. The likelihood at any point within the gate quickly approaches the machine epsilon resulting in zero likelihood everywhere within the gate. As a result, even targets that are exactly modeled by the kinematic filter will be spawned as new targets when they reappear like as shown in figure 5.14.

In summary, by calculating the prior P_D the tracker can rapidly grow the validation gate in the case that the P_D appears high (indicating that the target may be maneuvering beyond the gating limit), while it can maintain a low growth in the validation region when P_D is low allowing it to remain critical of future measurements.

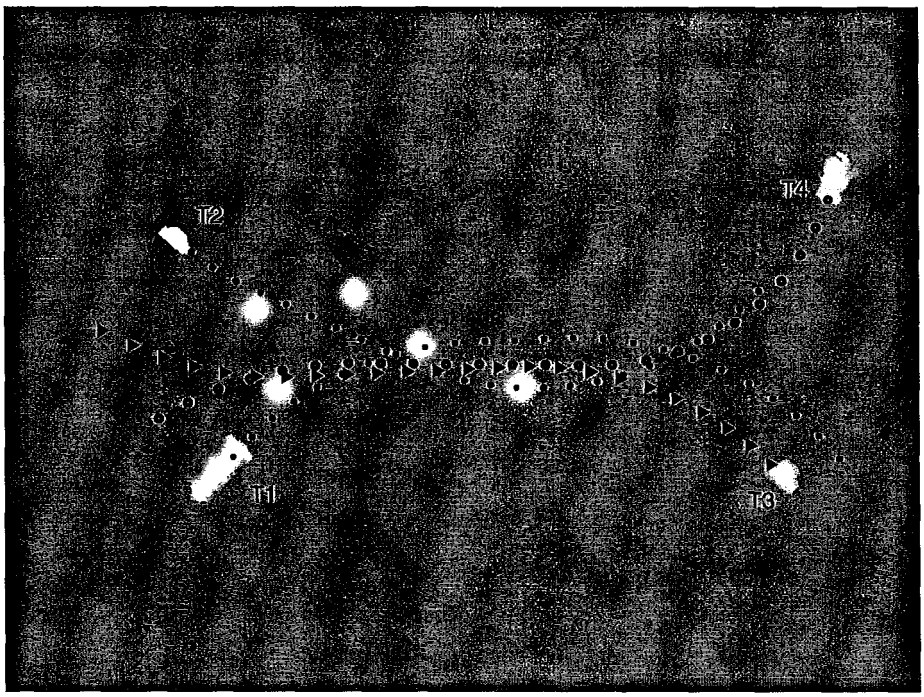


Figure 5.1: Synthetic video frame with target routes. These targets are referenced as T1, T2, T3, and T4 as shown here.

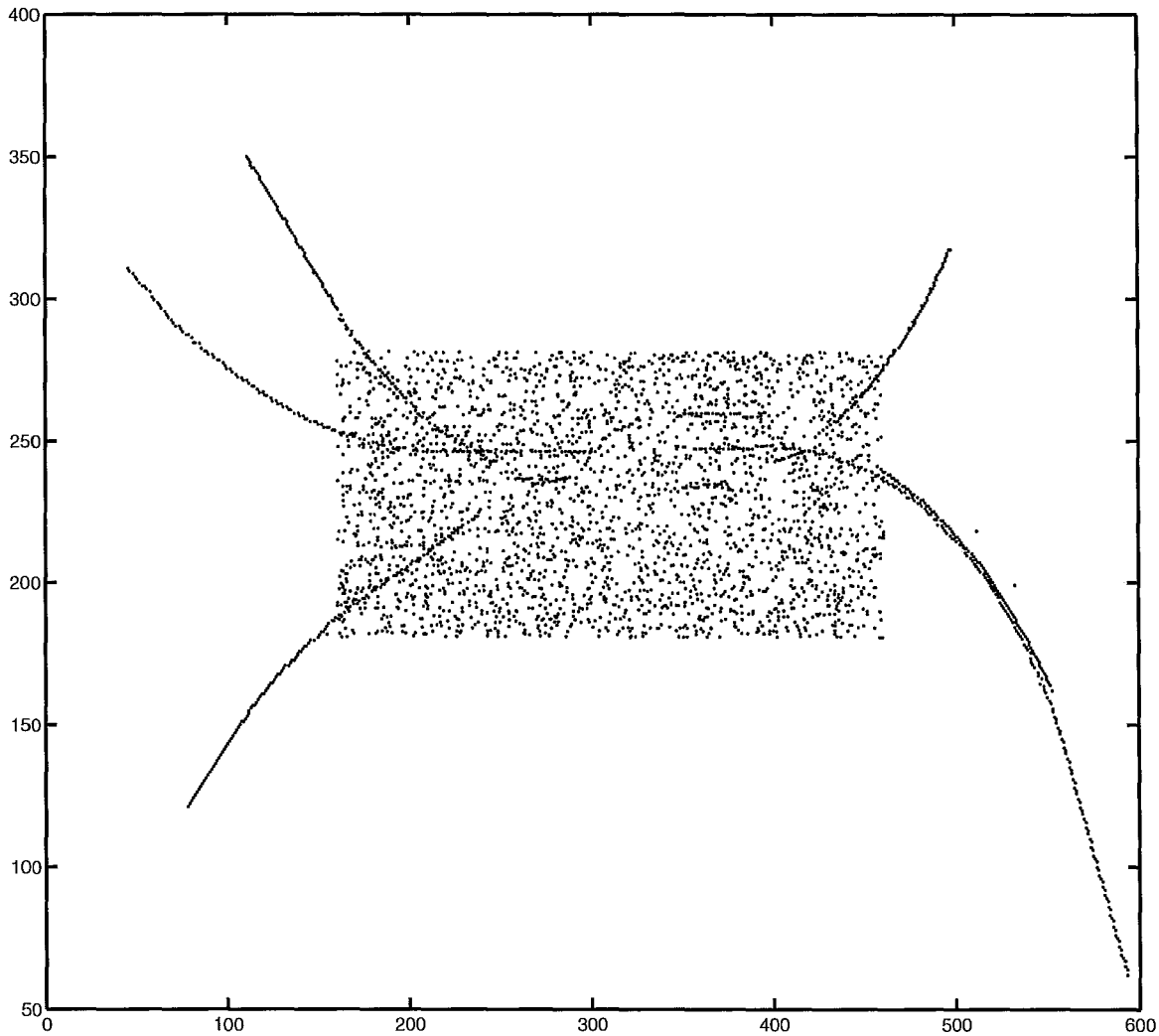
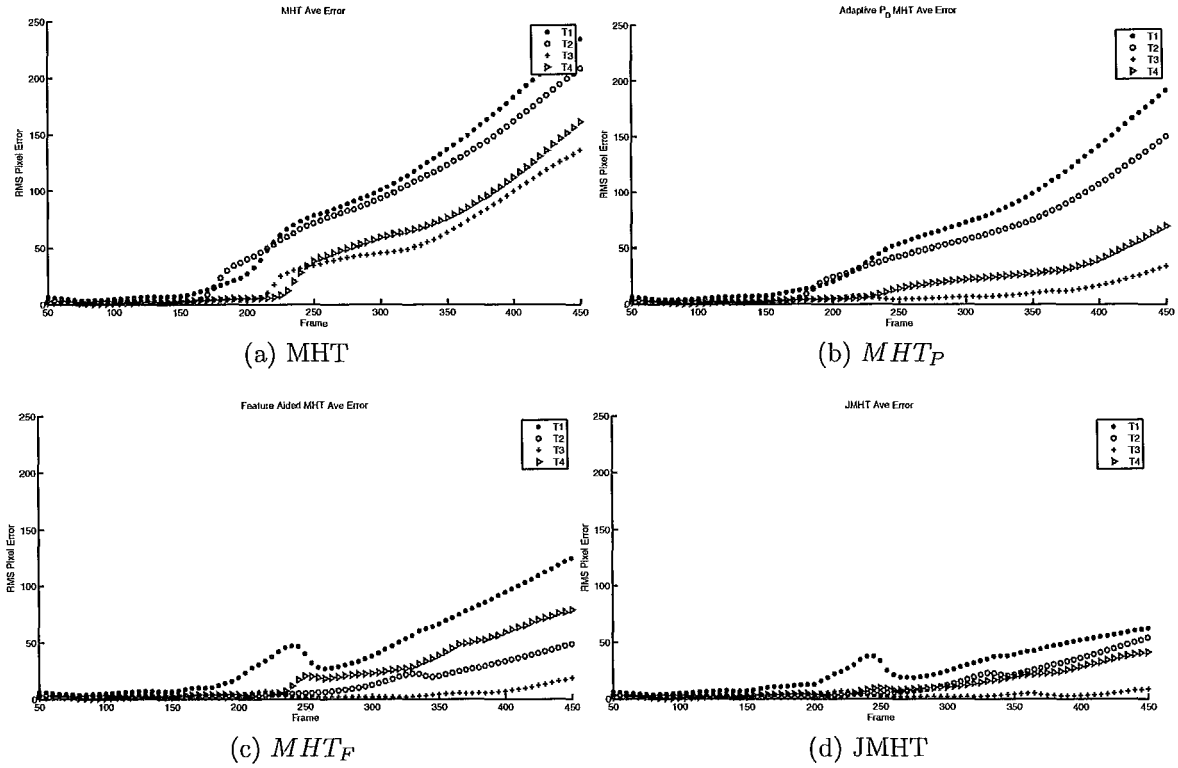


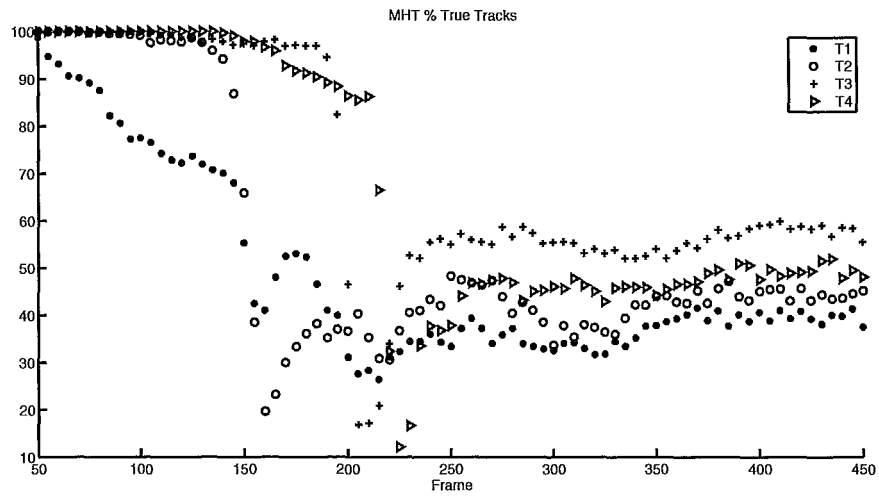
Figure 5.2: Measurement reports from all frames.



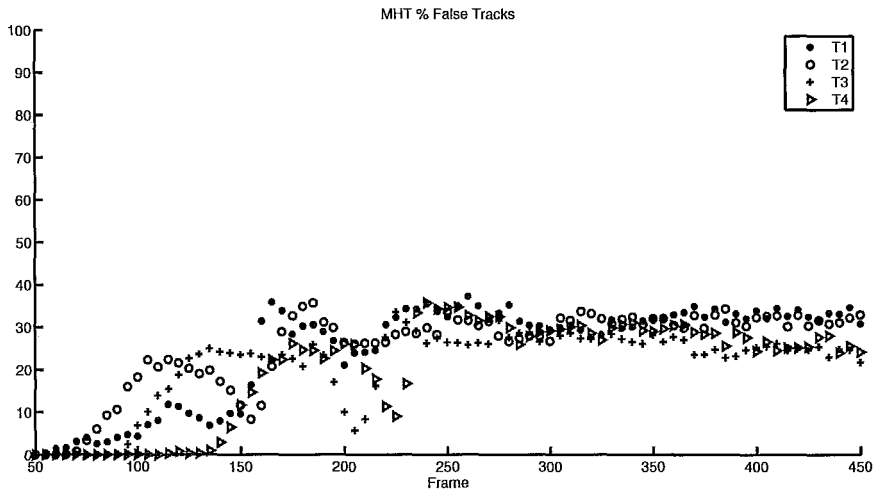
(a) MHT (b) MHT_P

(c) MHT_F (d) JMHT

Figure 5.3: By video frame, the average RMS tracking error across the simulations.

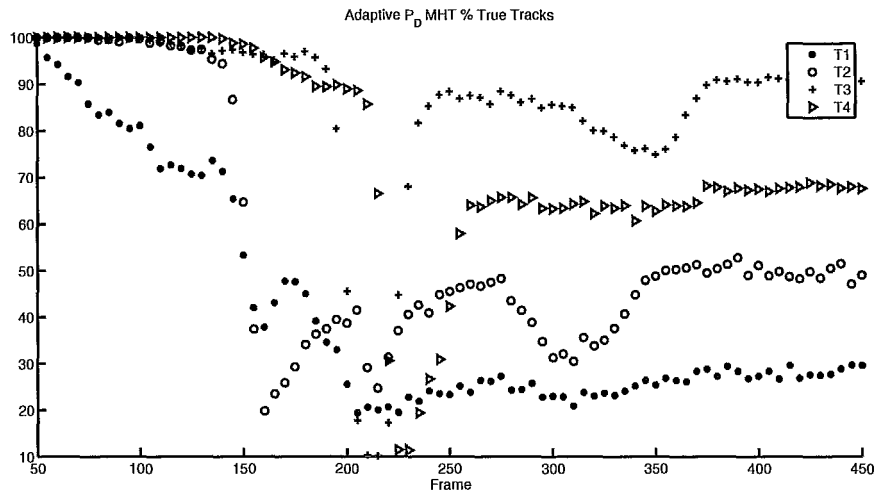


(a) True Tracks

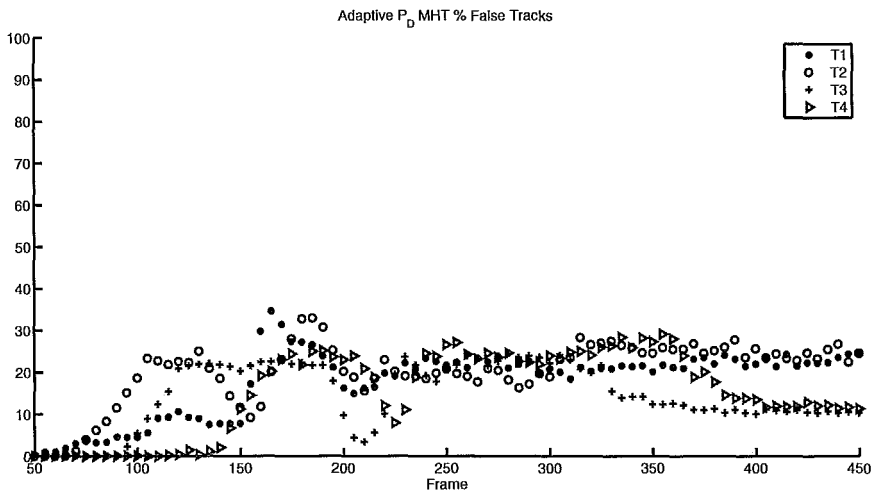


(b) False Tracks

Figure 5.4: Classical MHT, assuming $P_D = 0.9$: By video frame, the percentage of true and false tracks over the simulations.

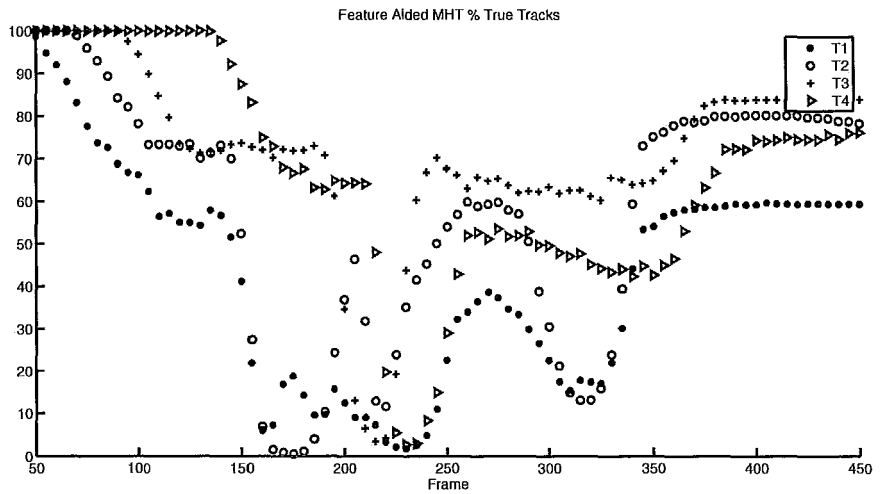


(a) True Tracks

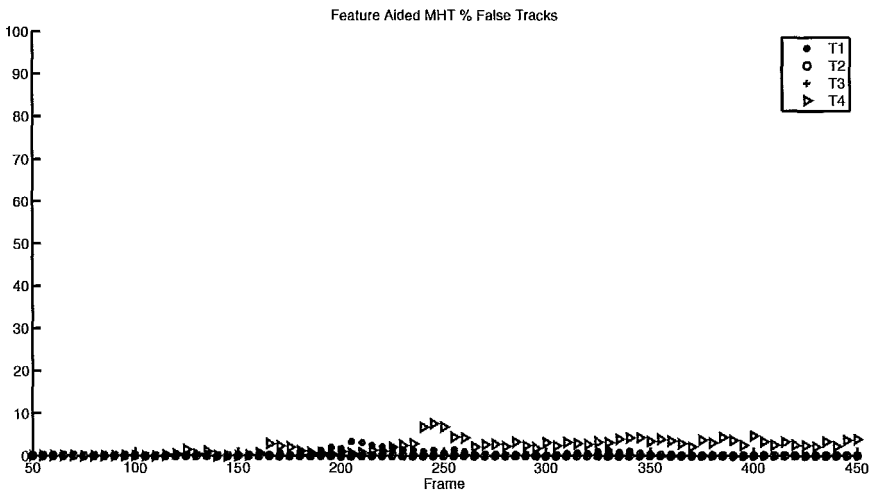


(b) False Tracks

Figure 5.5: Classical MHT augmented by an adaptive P_D: By video frame, the percentage of true and false tracks over the simulations.

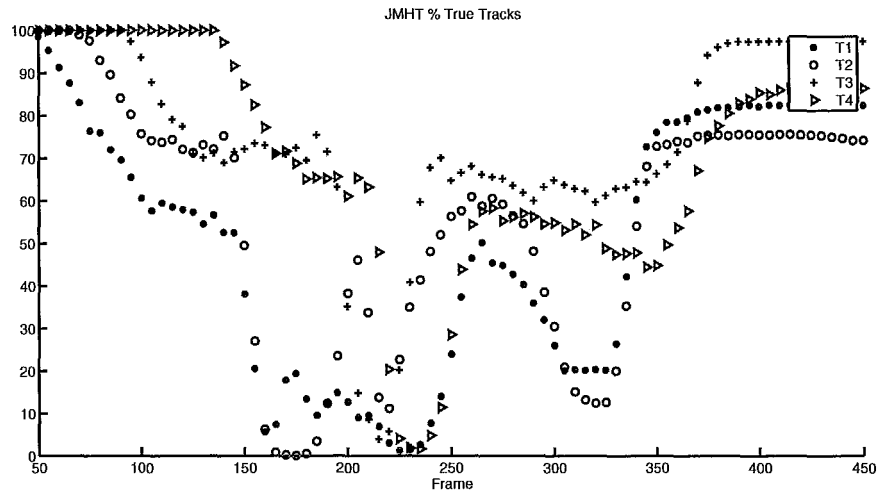


(a) True Tracks

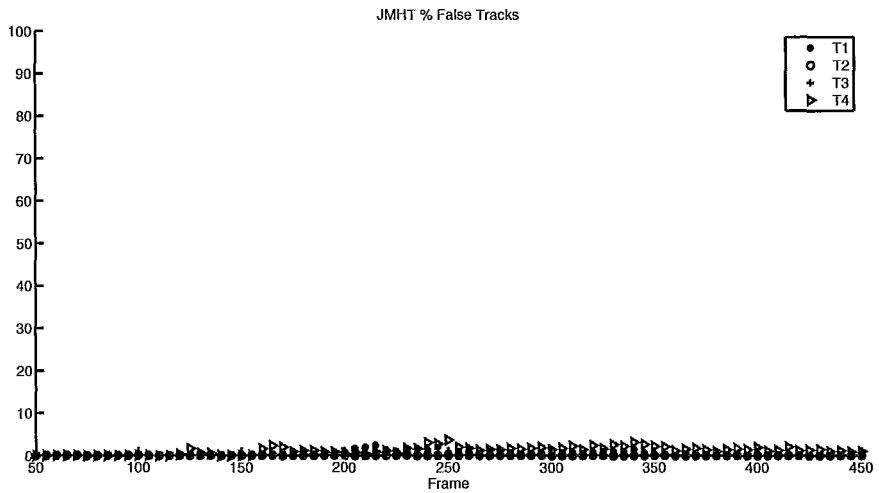


(b) False Tracks

Figure 5.6: Classical MHT augmented with feature information: By video frame, the percentage of true and false tracks over the simulations.

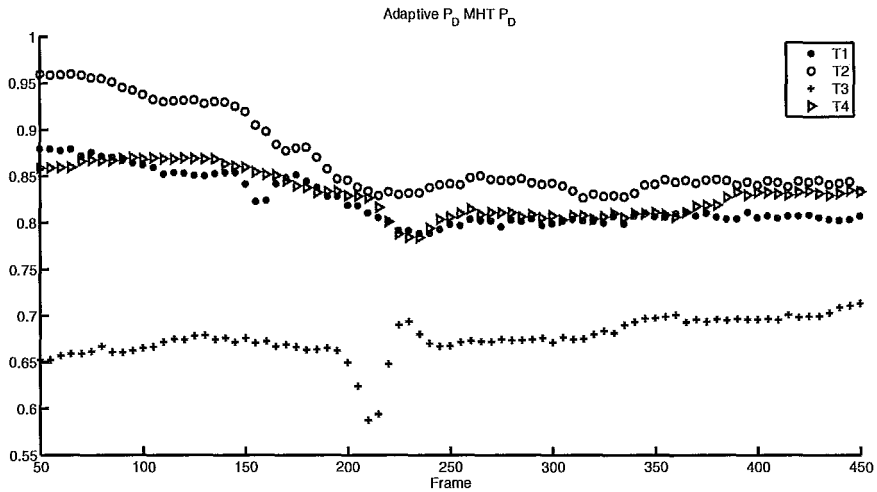


(a) True Tracks

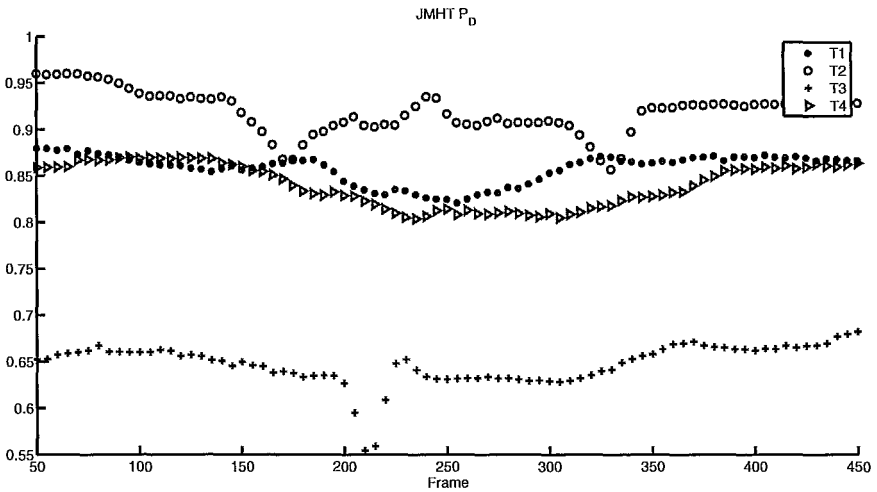


(b) False Tracks

Figure 5.7: Judicious MHT, augmented by an adaptive P_D and feature information: By video frame, the percentage of true and false tracks over the simulations.



(a) MHT_P



(b) JMHT

Figure 5.8: By video frame, the target P_D averaged over the simulations.

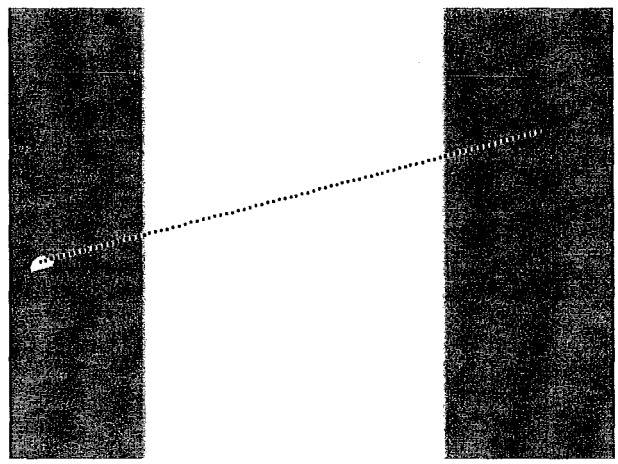


Figure 5.9: Synthetic video frame with target route. The target is referenced in the following figures as T1.

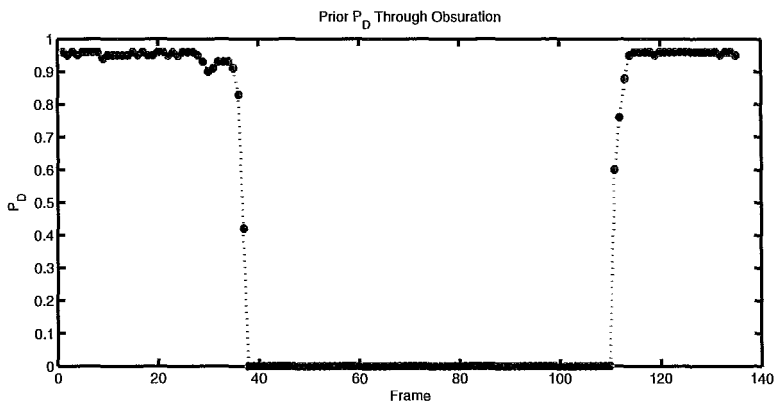


Figure 5.10: Calculated prior P_D under the judicious implementation.

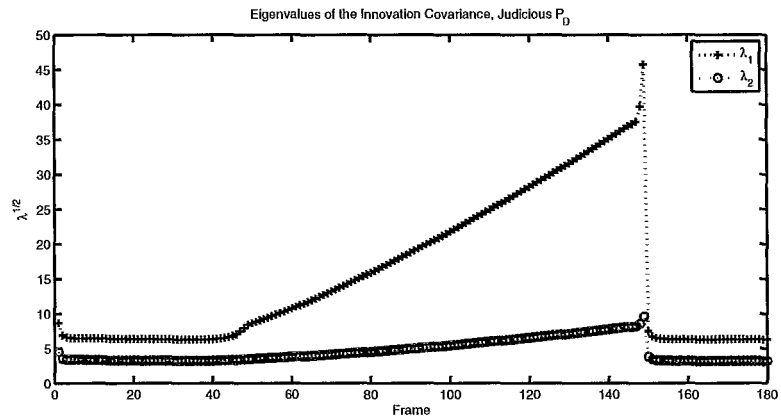


Figure 5.11: The semiaxes of the ellipsoidal gate are proportional to the eigenvalues of the innovation covariance. The slow growth is controlled by the low detection probability.

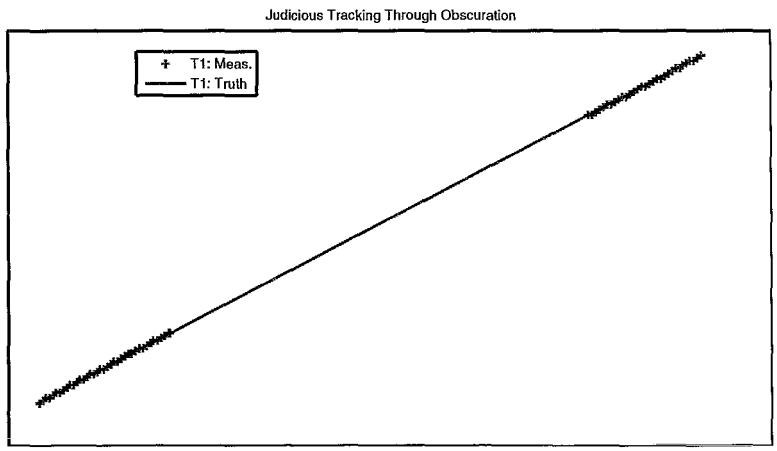


Figure 5.12: Tracking results using the full judicious implementation.

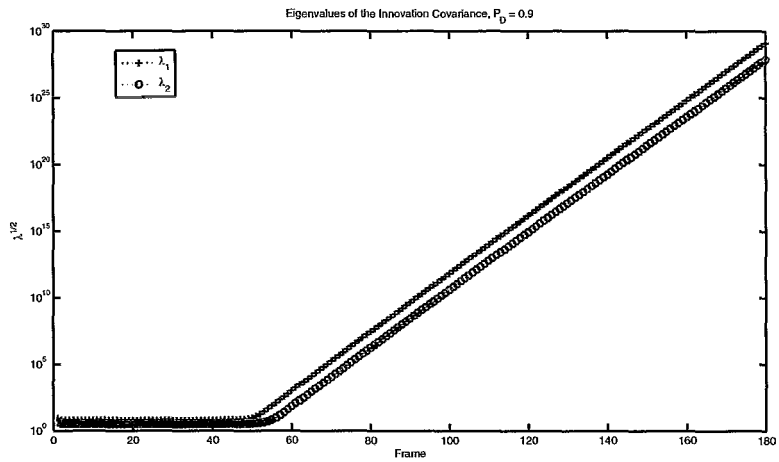


Figure 5.13: The growth in the state covariance at an artificially high P_D results in exponential growth in the validation region. Note that the y-axis is plotted on the logarithmic scale.

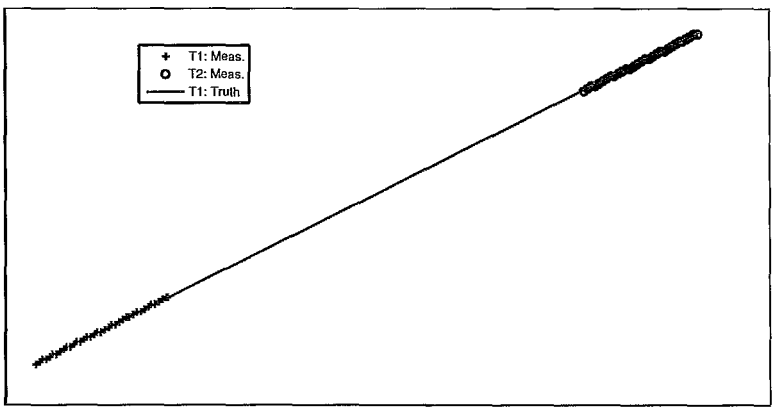


Figure 5.14: Tracking results assuming $P_D = 0.9$. The track fails due to the zero valued likelihood in an essentially infinite validation region.

Chapter 6

Conclusion

6.1 Conclusion

This thesis presented a probabilistic framework to include feature information in the MHT data association problem. In addition, it was shown that including the information available from the MHT predictions improved the feature extraction from the next round of measurements. Using this approach, the signal processor adds information to the prior probability of detection of a given track. That is, the judicious interaction between the signal processor and tracking algorithm shown in figure 1.3 ultimately increases the information available to the Bayesian prior. This prior information includes target feature information and target detection probability. Using this information, the MHT algorithm is improved as gating is achieved on both kinematic and feature aspects. Moreover, the growth of these gates and cost of a coasting assignment during missed detections is appropriately governed by a meaningful, target specific detection probability. The tracking performance through clutter and obscuration is improved as a result as is the ability of the tracker to expand the gating function for

a maneuvering target with a high probability of detection.

In the current implementation, the state covariance during coasting is increased per equation 4.17 using the prior probability of detection. For many cases this approach works well. However, in the case where the target loiters at the edge of a censored area, the prior probability of detection will be much higher than the posterior probability. That is, the predicted location will be easily segmented into foreground and background (a region that includes the normal and censored area described herein is essentially a two tone image). When the target is partially occluded by the censored area it will be segmented with this area and will not appear as its own measurement. In this case, a posterior probability of detection, the probability of detection given the assignment hypothesis of pixels within the validation gate, is more appropriate for use in growth of the covariance matrix. This is the course for our current research.

Finally, the interaction between signal processor and tracking algorithm described here is passive in the sense that the signal processor does not react to information from the tracking filter. Further effort is required to examine the prospect of an active interaction such that the signal processor changes the image acquisition to optimize the detection probability. In addition, the segmentation used herein looks for a single, global, threshold to segment foreground from background. The segmentation method should be extended to look at multiple thresholds to adaptively segment trackable features from the local background.

Bibliography

Bar-Shalom, Y. (1986). Comments, with reply, on 'Track biases and coalescence with probabilistic data association' by R.J. Fitzgerald. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-22**(5), 661 – 662.

Bar-Shalom, Y. and Li, R. (1995). *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, Storrs Ct.

Bar-Shalom, Y. and Tse, E. (1975). Tracking in a cluttered environment with probabilistic data association. *Automatica*, **11**(5), 451 – 60.

Bar-Shalom, Y., Chang, K., and Blom, H. (1989). Automatic track formation in clutter with a recursive algorithm. volume 2, pages 1402 –1408.

Bar-Shalom, Y., Li, X., and Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation*. Wiley, New York.

Bar-Shalom, Y., Blackman, S. S., and Fitzgerald, R. J. (2007). Dimensionless Score Function for Multiple Hypothesis Tracking. *Aerospace and Electronic Systems IEEE Transactions on*, **43**(1), 392–400.

Best, R. (2010). Intelligence, Surveillance, and Reconnaissance (ISR) Acquisition: Issues for Congress. http://www.fas.org/sgp/crs/intel/R41284.pdf.

Blackman, S. (2004). Multiple hypothesis tracking for multiple target tracking. *IEEE Aerosp. Electron. Syst. Mag. (USA)*, **19**, 5 – 18.

Blackman, S. and Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems.* Artech House, Norwood, MA.

Blackman, S. S. (1986). *Multiple Target Tracking with Radar Applications.* Artech House, Norwood, MA.

Blom, H. and Bloem, E. (2002). Interacting multiple model joint probabilistic data association avoiding track coalescence. volume 3, pages 3408 – 3415.

Cox, I. and Leonard, J. (1994). Unsupervised learning for mobile robot navigation using probabilistic data association. In *Proceedings of the workshop on Computational learning theory and natural learning systems (vol. 2) : intersections between theory and experiment*, pages 297–319, Cambridge, MA, USA. MIT Press.

Cox, I. J. and Hingorani, S. (1994). An efficient implementation and evaluation of Reid's multiple hypothesis tracking algorithm for visual tracking. In *International Conference on Pattern Recognition*, volume 1 of *Proc. 12th IAPR*, pages 437–442.

Cox, I. J., Miller, M. L., Danchick, R., and Newman, G. E. (1997). A comparison of two algorithms for determining ranked assignments with application to multitarget tracking and motion correspondence. *Aerospace and Electronic Systems IEEE Transactions on*, **33**(1), 295–301.

Crassidis, J. and Junkins, J. (2004). *Optimal Estimation of Dynamic Systems.* Chapman and Hall.

Drummond, O. (2000). Integration of features and attributes into target tracking. volume 4048, pages 610 – 622.

Drummond, O. (2001). Feature, attribute, and classification aided target tracking. volume 4473, pages 542 – 558.

Drummond, O. (2002). Target classification with data from multiple sensors. volume 4728, pages 377 – 393.

Drummond, O. (2004). Attributes in tracking and classification with incomplete data. volume 5428, pages 476 – 496.

Drummond, O. (2005). Tracking and classification with attribute data from multiple legacy sensors. volume 5913, pages 1 – 22.

Fitzgerald, R. (1985). Track Biases and Coalescence with Probabilistic Data Association. *Aerospace and Electronic Systems, IEEE Transactions on*, **AES-21**(6), 822 –825.

Fortmann, T., Bar-Shalom, Y., and Scheffe, M. (1980). Multi-target tracking using joint probabilistic data association. pages 807 – 12, New York, NY, USA.

Girshick, M. (1939). On the Sampling Theory of Roots of Determinantal Equations. *The Annals of Mathematical Statistics*, **10**(3), 203–224.

Gonzalez, R. and Woods, R. (2002). *Digital Image Processing.* Prentice-Hall Inc., 2 edition.

Gonzalez, R., Woods, R., and Eddins, S. (2003). *Digital Image Processing Using MATLAB.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Härdle, W. and Simar, L. (2007). *Applied Multivariate Statistical Analysis.* Springer-Verlag, 2 edition.

Kirubarajan, T., Bar-Shalom, Y., Pattipati, K., and Loew, L. (1997). Interacting segmentation and tracking of overlapping objects from an image sequence. volume vol.4, pages 3120 – 5, New York, NY, USA.

Lancaster, J. and Blackman, S. (2006). Joint IMM/MHT tracking and identification for multi-sensor ground target tracking.

Leung, H., Zhijian, H., and Blanchette, M. (1999). Evaluation of multiple radar target trackers in stressful environments. *IEEE Trans. Aerosp. Electron. Syst. (USA)*, **35**(2), 663 – 74.

Li, X. (1993). The PDF of nearest neighbor measurement and a probabilistic nearest neighbor filter for tracking in clutter. volume vol.1, pages 918 – 23.

Li, X. (1998). Tracking in clutter with strongest neighbor measurements. I. Theoretical analysis. *Automatic Control, IEEE Transactions on*, **43**(11), 1560–1578.

Li, X. and Zhi, X. (1996). PSNF: a refined strongest neighbor filter for tracking in clutter . volume 3, pages 2557 –2562.

McAnanama, J. and Kirubarajan, T. (2009). A judicious multiple hypothesis tracker with interacting feature extraction. *Proc. SPIE - Int. Soc. Opt. Eng. (USA)*, **7338**, 73380H (12 pp.).

Miller, M., Stone, H., and Cox, I. (1997). Optimizing Murty's ranked assignment method. *IEEE Transactions on Aerospace and Electronic Systems*, **33**(3), 851 – 862.

Miller, M. L., Stone, H. S., and Cox, I. J. (1995). On finding ranked assignments with application to multitarget tracking and motion correspondence. *Aerospace and Electronic Systems, IEEE Transactions on*, **31**(1), 486–489.

Morris, E. (2007). *Multivariate statistics : a vector space approach.* Institute of Mathematical Statistics, Beachwood, Ohio.

Musicki, D. and Evans, R. (2004). Joint integrated probabilistic data association: JIPDA. *Aerospace and Electronic Systems, IEEE Transactions on*, **40**(3), 1093 – 1099.

Musicki, D., Evans, R., and Stankovic, S. (1994). Integrated probabilistic data association. *Automatic Control, IEEE Transactions on*, **39**(6), 1237 –1241.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, **9**(1), 62–66.

Papanikolopoulos, N., Veeraraghavan, H., and Schrater, P. (2006). Robust target detection and tracking through integration of motion, color, and geometry. *Comput. Vis. Image Underst. (USA)*, **103**(2), 121 – 38.

Papoulis, A. (2001). *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 4th edition.

Pattipati, K. R., Popp, R. L., and Kirubarajan, T. (2000). *Multitarget-Multisensor Tracking: Applications and Advances*, chapter Survey of Assignment Techniques for Multitarget Tracking, pages 77–159. Artech House.

Reid, D. (1979). An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, **AC-24**(6), 843–854.

Sonka, M., Hlavac, V., and Boyle, R. (2007). *Image Processing, Analysis, and Machine Vision.* Thomson-Engineering.

Tissainayagam, P. and Suter, D. (2001). Visual tracking with automatic motion model switching. *Pattern Recognition*, **34**(3), 641 – 660.

Trucco, E. and Plakas, K. (2006). Video tracking: a concise survey. *IEEE J. Ocean. Eng. (USA)*, **31**(2), 520 – 9.

Wang, H., Kirubarajan, T., and Bar-Shalom, Y. (1999). Precision large scale air traffic surveillance using IMM/assignment estimators. *IEEE Transactions on Aerospace and Electronic Systems*, **35**(1), 255 – 66.

Yeddanapudi, M., Bar-Shalom, Y., and Pattipati, K. (1997). IMM estimation for multitarget-multisensor air traffic surveillance. *Proceedings of the IEEE*, **85**(1), 80 – 96.

Zetterlind, V. and Matechik, S. (2006). Hue-saturation-value feature analysis for robust ground moving target tracking in color aerial video. *Proc. SPIE - Int. Soc. Opt. Eng. (USA)*, **6235**, 62351 – 62360.