Multiple Target Tracking by Integrating Track Refinement and Data Association

Shunguang Wu, Subhodev Das, Yi Tan, Jayan Eledath and Ali Z. Chaudhry
Vision Technologies, SRI International Sarnoff
201 Washington Road, Princeton, NJ 08540, USA

Abstract— Multiple target tracking that integrates target model estimation and data association steps is described. The integration allows successive refinement of the models while reducing the uncertainty in data association. Each target is described by "weak" models of kinematics, shape and appearance. The target models are refined in a two-stage process: image-based tracklets of high purity and accuracy are generated, and geospatial tracks are extended from these tracklets. During each stage of tracking, observation data of reduced uncertainties are associated with the refined tracks in a probabilistic manner. We describe our approach in the context of a real time system that has been tested and evaluated for vehicle and human tracking in sparse, medium, and dense clutter using aerial EO/IR video.

I. INTRODUCTION

Aerial video processing (AVP) has experienced a significant market growth during the current decade and has seen successful deployment of the technology in battlefields around the globe. AVP broadly supports intelligence, surveillance and reconnaissance (ISR) type activities. Of the several facets of AVP, detection and tracking of surface (primarily ground) moving targets (i.e., certain objects of interest) have received substantial attention in the recent years, such as in DARPA's VId eo V erification of IDentity (VIVID) [1] and Flow-based Information Theoretic Tracking (FITT) programs and Australian DoD’s Analysts Detection Support System (ADSS) [2].

In order to extract actionable information, targets will have to be tracked, several of them at any given time, over considerable periods. The problem of multiple target tracking (MTT) in aerial video, as illustrated in Figure 1, is challenging due to continuous changes in target appearances, frequent occlusion/field of view (FOV) departure of targets and complex relative motion between the targets and the sensor. Assuming point representation of targets, tracking amounts to establishing correspondences between sets of points at consecutive time (of observation) instances. However, the conditional probability distribution of the predicted target state given the current target position and velocity can be highly multi-modal and can lead to incorrect correspondences or data association under high target density particularly when using deterministic methods, such as nearest neighbor filter (NNF). Statistical correspondence methods handle observation and target model uncertainties, such as measurement noise and random perturbation of maneuvering targets, during target state estimation and assign probabilistic weights to the correspondences. These methods use state space approach to model target position, velocity, and acceleration [3]. The solution of MTT requires a joint resolution of target model estimation and data association. However, most statistical methods use "strong" models of target motion, shape and appearance [4] that are difficult to satisfy given the varied nature of targets in AVP scenarios. The data association may have its own constraints that are hard to meet under a multitude of target clutter conditions. For example, the joint probabilistic data association filter (JPDAF, [5]) updates target state using all the measurements in the vicinity of the predicted location of the target that are weighted by the motion model-based assignment likelihood. The major shortcoming of JPDAF is its inability to handle new targets entering the field of view.
features are updated using an all within the JPDA framework. The shape and appearance and the results are used to update target geospatial kinematics, targets using the platform metadata. These observations are variances are computed along a tracklet or for image localized appearance features learned for the individual targets. The stopped targets are tracked using image-based correlation of representation. Motion layers containing moving targets in structural information. The tracker operates at two levels of the appearance features consist of photometric and image structural information. The tracker operates at two levels of the kinematics, shape and appearance: the kinematics include location and velocity in a reference coordinate system; the shape is described by a parametric model; while the appearance features consist of photometric and image structural information. The tracker operates at two levels of representation. Motion layers containing moving targets in stabilized video sequence are identified and tracked yielding image tracklets of high purity and accuracy. Momentarily stopped targets are tracked using image-based correlation of appearance features learned for the individual targets. The location and velocity in geospace and the corresponding covariances are computed along a tracklet or for image localized targets using the platform metadata. These observations are assigned to geospatial tracks using k-best Hungarian algorithm and the results are used to update target geospatial kinematics, all within the JPDA framework. The shape and appearance features are updated using an $\alpha - \beta$ filter. The geospatial tracks with longer track-life expectancy (TLE), high purity and accuracy, as shown in Fig. 1, are used in the prediction of image target state in the next processing cycle. The issues of track life cycle management, including track requisition, and track splitting/merging are also addressed in our framework.

The paper is organized as follows. The next sub-section reviews the background work and presents the motivation of our approach. Section 2 provides an overview of the MTT system. Section 3 presents the detection methods for both moving and momentarily stopped targets. The tracking and data association algorithms and their integration are discussed in section 4. Section 5 includes illustrative results; Section 6 has the concluding remarks.

A. Related Work

State estimation of a single target is based on the popular Kalman Filtering (point estimation) or particle filtering (density estimation). When tracking multiple targets using Kalman or particle filter, one need to deterministically associate the most likely measurement of a particular target to its state. If, however, the targets are close to each other, there is always a chance that the correspondence is incorrect. Although particle filters can support multiple hypotheses by the way of multi-modal motion and measurement models, a lack of consideration of multiple hypotheses over several time instants, say, across observation gaps, is a major drawback. Multiple-Model algorithms [7] seek to use more than one (typically 2-4) independent/interacting Kalman or particle filters to simultaneously model the motion behavior of maneuvering targets. High target density and low update rate pose serious initialization problems for these approaches. Relational Dynamic Bayesian Networks (RDBNs) employing first-order logic to model correlations between object behaviors can represent interactions among targets in high clutter environment, such as particle filter-based tracking that infersences over RDBNs [8]. More efficient solutions of inference are needed before RDBNs become common in tracking applications.

The performance of a multiple target tracker is typically measured in terms of i) TLE as a function of track purity, i.e., percentage of data in a track originating from the target that initiates the track, and ii) track accuracy, i.e., how accurate the state estimates are. The performance is characterized by a number of environmental variables, such as target density, sensor resolution, relative sensor-target geometry, update rate, etc. In a real time system, the speed at which a certain level of performance is achieved is also an important consideration. To strike a balance between performance and speed, a common strategy is to break tracks into tracklets and link these with high-level reasoning or MHT. The efficacy of this approach is dependent on the purity and accuracy of the tracklets. Additionally, tracklets may result from long periods of observation loss. [9] describes a way to maintain track identity during merges and splits just before and after long occlusions in the track linking approach. This algorithm uses an NNF for data association and demonstrates results for structured environments, such as vehicles on highways, in which tracks tend to be long and therefore predictable. Another method [10], which is the closest in implementation to our approach, derives image-based tracklets using a layer-based tracker [4] and applies a post-tracking refinement process to link the track fragments after mapping them into geospatial coordinates. Its performance is limited by the purity of the image-based tracklets and its inability to track momentarily stopped targets, which is an important consideration for video-based tracking.

B. Motivation

The types of targets available for tracking in aerial video may vary considerably in shape, size and speed and can range from large vehicles moving at considerable speed to walking human to momentarily stopped moving entities. Such variabilities forgo the use of "strong" model types for the target states. The measurements have uncertainties in their origin; these may be due to the target of interest, interfering targets, clutter, countermeasures, or false alarms in the detection process. The variations in the observed scene – day/night operations, urban/rural settings, target/clutter densities – place additional burden on a real time ISR system. The class of expectation-maximization (EM) algorithms, with its iterative expectation (E) and maximization (M) steps, is well-suited for estimating the target states. Among various algorithms developed for MTT in cluttered environment, Bayesian approaches that evaluate the posterior probabilities of data association and use them throughout the estimation process seem to be the most promising [11].

To achieve continued high performance, we propose a
hierarchical approach in which the target model estimation, using a EM algorithm, and the data association, using a Bayesian probabilistic data association (PDA) algorithm, are integrated. This integration allows incremental refinement of the target model, thereby reducing the error covariance (mean-squared error) matrix associated with the estimated state. The target model at image level uses image-based features that are continuously adapted, allowing tracking of both moving and stopped targets, and under changing environmental and viewing conditions. The corresponding target model in geospace captures the dynamics and appearance in 3D that may be quiet different from image-based observations, such as targets can be occluded in image. The improved target model, in turn, reduces the uncertainty in data association.

II. SYSTEM OVERVIEW

This paper describes a MTT approach that is part of a demonstrated prototype ISR system shown in Figure 2. The input to the ISR system is a full-motion video stream from an analog/digital EO/IR sensor and synchronized platform metadata. A preprocessing step removes the sensor noise and reduces the imaging artifacts. The bits representing each pixel datum are often corrupted by noise in the sensor hardware. The outlying values are identified in the distribution of the input data and are clipped to the significant range of the data. Since IR video is often plagued by significant inter-frame intensity variation, noise-minimized IR imagery is also subjected to local contrast enhancement. The intensity corrected frames are motion stabilized and forwarded to the MTT sub-system. Additionally, the MTT sub-system may receive filtered metadata that are optionally corrected by video geo-registration. The MTT component has the following real-time functionalities: (i) detection of moving and momentarily stopped targets, and (ii) tracking of moving and stopped targets. It has several modules accomplishing these tasks that are described below. The output of the MTT sub-system are track geolocation (latitude, longitude, height), kinematics (velocity northing and easting), shape (length and width), and appearance (spatial color clusters in RGB/HSV system) property values along with the corresponding covariances, track status (new, tentative, active, inactive, retracted, dropped) and the associated detection (image chip).

III. TARGET DETECTION

The target detection capability of our MTT sub-system is concerned with locating moving and momentarily stopped targets in stabilized video. Since this is the initial step of MTT, target detection is subjected to close scrutiny and is, therefore, not expected to report detection at every observation instance. The primary detection module is based on moving target indication (MTI) to identify targets undergoing motion independent of that of the sensor. The MTI process uses a Gaussian feature detector that can automatically adjust to the expected scale of features characterizing the likely targets – wheeled ground, tracked ground, air vehicle, and human are the target types - together with range and platform metadata.

 Significant spatiotemporal changes in these features between two observation instances are recorded. When the sensor is stationary or undergoing very small motion, such as in a geopointing gimbal, both moving and momentarily stopped targets can be detected as foreground changes against a dynamically adapting background model of the scene. The background model maintains a non-parametric intensity distribution on a per pixel basis obtained by normalizing the temporal, exponentially weighted histogram of the entire observation history. The identified change pixels from either detection method are grouped into regions or blobs using connected component labeling. The blobs satisfying temporal consistency, such as of motion, over an extended period of observation are retained as the sightings of moving targets.

IV. TARGET TRACKING

The target tracking functionality of the MTT sub-system is required to deliver accurate geospatial tracks with potentially high TLE; the tracks may be predicted in the absence of target detection. The image-based tracking receives input from the MTI module and produces image tracklets of high purity. The subsequent geospatial tracking extends the geospatially mapped image tracklets.

A. Image-based Tracking

Image-based tracking is concerned with data association in image coordinates. It addresses both moving and stopped target tracking.

1) Moving target tracking: The moving target regions in the image are further organized into motion layers each representing a tracked object characterized by weak online models of motion, shape and appearance [12]. These models improve the robustness of the tracking process under non-ideal conditions like partial or complete occlusion, which affect
the traditional layer-based tracking methods. In our approach, the layer ownership probabilities for pixels belonging to the moving objects are estimated using an EM algorithm. These probabilities are initialized using a non-parametric motion model. In the E-step, the non-parametric motion estimate derived from optical flow between the observation instances is used to warp the pixel ownership probabilities. This also involves matching of two-dimensional appearance histograms between motion layers of a previous observation and the identified MTI blobs of the current one. In the M-step, these computed ownership probabilities are used to refine the motion, shape and appearance estimates of the layers. Only layers exhibiting single ownership (i.e., unity probability) and homogeneous appearance are issued "new" track identities. Tracking involves associating the blobs in the current frame to the labeled layers by minimizing motion, shape, and appearance costs in which the association matrix is initialized using the flow-warped ownership probabilities of the layer pixels. The individual tracks are refined with Kalman filter (KF) and the layer ownerships are updated following successful data association. The tracks that fail during data association are marked as lost and are either re-acquired or dropped. The image-based tracklets are mapped into geospatial coordinates using platform metadata and are provided as input to geospatial tracking.

2) Stopped target tracking: The stopped target tracking module is triggered when an "active" (i.e., continuously tracked) target attains “active stopped” status (see Figure 4); its objective is to image locate and track the target that may have stopped. The input to this module is the predicted geo-location of the target and its MTI blobs queued from the immediately preceding observation instances. The output of the module is the geolocation of the stopped target. The key feature of this module is dynamic acquisition and adaptation of target appearance model as it undergoes kinematic state change.

In the current implementation, the target appearance model consists of intensity edge structures. The edge image is derived using Canny detector and the locations, orientations and magnitudes of all the edge points are stored in the edge list data structure: 

\[ E_l = \{e_1, \cdots, e_n\}, e_i = [x, y, \theta]^T, \]

where \( x, y, \theta \) are the location, orientation of an edge point; \( d \) is the Chamfer distance from edge point \( e_i \) to \( I_D \); \( \beta \) is the angle difference between \( e_i \) and its corresponding edge point in the search area; \( d_{max} \) and \( \beta_{max} \) are the distance and angle difference thresholds indicating a failed match, respectively. Figure 3 illustrates the intermediate steps of the matching process.

\[ \Psi(E'_l, I_D) = \frac{1}{n} \sum_{i=1}^{n} p_i^{(a)}(1 - d_i)/d_{max}, p_i^{(a)} = (1 - \beta_i)/\beta_{max}, \]

where \( d_i \) is the Chamfer distance from edge point \( e'_i \) to \( I_D \); \( \beta_i \) is the angle difference between \( e'_i \) and its corresponding edge point in the search area; \( d_{max} \) and \( \beta_{max} \) are the distance and angle difference thresholds indicating a failed match, respectively. Figure 3 illustrates the intermediate steps of the matching process.

\[ \tilde{a} = \arg \max_a \Psi(E'_l, I_D), \quad E'_l = \Gamma(E_l; \tilde{a}), \]

where \( \tilde{a} = [dx, dy, \alpha, s]^T \) is the estimated vector of translation, rotation and scale change of the model; \( \Gamma(E_l; \tilde{a}) \) is a function that transforms \( E_l \) into \( E'_l = \{e'_1, \cdots, e'_n\} \). The edge \( e'_i[x', y', \theta']^T \) is transformed by one particular hypothesized parameter \( \alpha \):

\[
\begin{align*}
x'_i &= s \cos(\alpha)x_i - s \sin(\alpha)y_i + dx, \\
y'_i &= s \sin(\alpha)x_i + s \cos(\alpha)y_i + dy, \\
\theta'_i &= \theta_i + \alpha,
\end{align*}
\]

\( \Psi(E'_l, I_D) \) is the matching score of the edge list \( E'_l \) and the distance transform image \( I_D \):

\[ \tilde{a} = \arg \max_a \Psi(E'_l, I_D), \quad E'_l = \Gamma(E_l; \tilde{a}), \]

where \( \tilde{a} = [dx, dy, \alpha, s]^T \) is the estimated vector of translation, rotation and scale change of the model; \( \Gamma(E_l; \tilde{a}) \) is a function that transforms \( E_l \) into \( E'_l = \{e'_1, \cdots, e'_n\} \). The edge \( e'_i[x', y', \theta']^T \) is transformed by one particular hypothesized parameter \( \alpha \):

\[
\begin{align*}
x'_i &= s \cos(\alpha)x_i - s \sin(\alpha)y_i + dx, \\
y'_i &= s \sin(\alpha)x_i + s \cos(\alpha)y_i + dy, \\
\theta'_i &= \theta_i + \alpha,
\end{align*}
\]

\( \Psi(E'_l, I_D) \) is the matching score of the edge list \( E'_l \) and the distance transform image \( I_D \):

\[ \Psi(E'_l, I_D) = \frac{1}{n} \sum_{i=1}^{n} p_i^{(a)}(1 - d_i)/d_{max}, p_i^{(a)} = (1 - \beta_i)/\beta_{max}, \]

where \( d_i \) is the Chamfer distance from edge point \( e'_i \) to \( I_D \); \( \beta_i \) is the angle difference between \( e'_i \) and its corresponding edge point in the search area; \( d_{max} \) and \( \beta_{max} \) are the distance and angle difference thresholds indicating a failed match, respectively. Figure 3 illustrates the intermediate steps of the matching process.

B. Geospatial Tracking

The geospatial tracking feature utilizes the target models associated with the image tracklets and refine them in its own model update step. Data association is also benefitted from the highly pure image tracklet models.

1) Target state and observation modeling: A target state is represented by its location and velocity in east-north-up (ENU) coordinate system where the ENU original point is selected as the initial location of the first target in a sequence. Additionally, the target shape and appearance are also included in the state vector:

\[ \mathbf{x}_j(k) = \{\mathbf{x}^{(s)}_j(k), \mathbf{x}^{(a)}_j(k), \mathbf{x}^{(o)}_j(k)\}, \]

where \( j = 1, \cdots, N(k) \) is the index number of \( N(k) \) targets in frame \( k \); \( \mathbf{x}^{(s)}_j(k), \mathbf{x}^{(a)}_j(k), \) and \( \mathbf{x}^{(o)}_j(k) \) are kinematic, shape, and appearance feature vectors, respectively. The kinematic vector is composed of the location and velocity of a target in the ENU coordinates, the shape is represented by an ellipse in the image space, and the six-dimensional edge measure vector is used to represent the appearance.
The observation vector for the $i$th detected target among $M(k)$ ones in frame $k$ is represented as

$$z_i(k) = \{z_i^{(m)}(k), z_i^{(s)}(k), z_i^{(a)}(k)\},$$

where $z_i^{(m)}(k)$ is the geospatial location of the object center, $z_i^{(s)}(k)$ is the shape, and $z_i^{(a)}(k)$ is the appearance observation vector, respectively. Given an image pixel and its covariance, the geolocation process can deterministically estimate its corresponding location and covariance in ENU system using sensor-platform metadata and the earth model/terrain data [13].

A constant velocity model is used to describe the propagation of the kinematic state. A KF is used to estimate the kinematic state, while an $\alpha - \beta$ filter is used to update the shape and appearance features.

We also estimate the ground heading angle of a target, $\theta(t) = \arctan \frac{a_1 + 2b_2t}{a_2 + 2b_1t}$, using a least square method to fit a second order polynomial to the most recent trajectory of the target, i.e.,

$$X(t) = a_0 + a_1t + a_2t^2, \quad Y(t) = b_0 + b_1t + b_2t^2,$$

where $(X, Y)$ is the target location in EN space, $t$ is the time, and $a_1's$ and $b_1's$ are the parameters to be estimated.

2) Data association: In the following discussion, we use $x_j(k+1|k)$ and $z_j(k+1|k)$ to represent the predicted state and predict observation of the $j$th track between frames $k$ and $k+1$. Instead of directly computing the likelihood of data association between all pairs of $x_j(k+1|k)$ and $z_j(k+1)$, the gating process screens out the potential observations of a track according to

$$\rho_i^{(l)}(z_i^{(l)}(k+1), x_j^{(l)}(k+1|k)) \leq \rho_T^{(l)},$$

where $i$ and $j$ are the indices of observation and track, respectively; $l = \kappa$, $s$, $a$; $\rho_T^{(l)}$s are their thresholds. The implementation details of these functions are presented in [13].

Given a observation inside the gate of a track, the data association cost is computed by

$$C_{ij} = \begin{cases} 
\phi_i, & \text{if } z_i \text{ to } T_j \text{ is forbidden}, \\
c_{F,A}, & \text{if } z_i \text{ is potentially a FA}, \\
c_{NT}^{(m)}, \frac{c_{ij}^{(s)}}{c_{ij}^{(a)}}, & \text{if } z_i \text{ is potentially a new track}, \\
\text{otherwise}, & 
\end{cases}$$

where $c_{ij}^{(m)}$, $c_{ij}^{(s)}$, and $c_{ij}^{(a)}$ are the probabilities of kinematic, shape and appearance matching scores, respectively. The kinematic probability is calculated from Gaussian density function, i.e.,

$$c_{ij}^{(m)} \sim \frac{1}{\sqrt{\det(S)}} \exp\left(-\frac{1}{2} d_i^2\right),$$

where $d_i$ is the normalized distance between predicted observation and actual one. The shape and appearance probabilities are derived from pairwise chip matching scores with uniform distribution.

The $k$-best Hungarian algorithm [14] is used to retain the $k$-best hypotheses track update. Unlike the whole-event JPDA algorithm [5], the $k$-best JPDA algorithm only calculates the probabilities of the $k$-best assignments/events first.

3) Track update and life cycle maintenance: Once the association cost have been derived, these are normalized to yield the association probability matrix. The $j$th track is updated according to

$$x_j(k) = \sum_{i=1}^{m} \beta_{ij} x_{ij}(k) + \beta_{0j} x_j(k|k-1),$$

$$P_j(k) = \sum_{i=1}^{m} \beta_{ij} P_{ij}(k) + \beta_{0j} P_j(k|k-1) + \sum_{i=1}^{m} (\dot{x}_{ij}\dot{x}_{ji}^T),$$

where $x_{ij} = x_{ij}(k) - x_j(k); \beta_{ij}$ is the probability of assigning observation $i$ to track $j; \beta_{0j} = 1 - \sum_{i=1}^{m} \beta_{ij}$ is the probability of false alarm (FA) for track $j$. $x_{ij}$ and $P_{ij}$ are the estimated state and its covariance due to observation $i$. Note, the above update formulas are only for the kinematic part of the target state. The appearance and shape features are updated by the best assignment.

The life cycle of a track is maintained according to eight track status states shown in Figure 4. The status of a just initialized track is “new”; if a new track is updated by an observation, its status becomes “tentative,” otherwise it becomes “retracted.” A “tentative” track can eventually become “active” if it is continued to be observed, otherwise it becomes “retracted.” If an “active” track is not updated, it may becomes “inactive occluded” or “inactive out FOV” or “active stop.” An inactive or stopped track can regain “active” status if it is updated by an observation, otherwise it is eventually “dropped” based on its duration of inactivity. Similarly, a “retracted” track can regain “tentative” status if it is updated, otherwise it too is “dropped.” Finally, a “dropped” track is deleted from the tracked target list.

![Fig. 4. Target status and their transition diagram.](image)

V. EXPERIMENT RESULTS

The MTT approach is implemented in a real time ISR system and tested on aerial platforms to track vehicle and human targets in sparse, medium, and dense clutter. Both EO and IR video streams are processed. The approach is quantitatively evaluated offline using EO/IR video clips collected.
under realistic operational scenarios. This section first presents the results of typical MTT scenarios, such as vehicle crossing/close following and move-stop-move, target occlusion, and dismount (people) tracking. Next, the TLE’s from image-based tracking and geospatial tracking are compared.

Figure 5 shows crossing/passing scenarios involving multiple vehicles. In one scenario, there are three vehicles, two of which are moving in one direction and the third one in the opposite. In the other scenario, two vehicles are traveling in the same direction. Although the MTI blobs are merged as seen in the result, the image-based tracking is able to maintain proper segmentation of the motion layers as indicated by the elliptical shape models.

Figure 6 illustrates a vehicle move-stop-move scenario in which multiple momentarily stopped vehicles are tracked along with moving vehicles.

Figure 7 describes occlusion handling, i.e., track re-acquisition, by the geospatial tracker. During occlusion, when there is no detection and the image-based tracker drops “inactive” tracklets quickly, the geospatial tracker is not updated; however, it can still predict the target location. Whenever the target is detected again and as long as it is still inside the gating of the predicted track, the detection can be potentially associated with the track based on the kinematics, shape and appearance. The current track re-acquisition strategy can handle only short-term (3-5 seconds) occlusion in medium clutter.

An example of dismount tracking is seen in Figure 8. The relative performance of image-based and geospatial tracking is illustrated in Figure 9 that shows the corresponding tracks displayed in Google Earth. Clearly, the geospatial tracking has improved TLE while maintaining purity of the contributing image-based tracklets.

Finally, to statistically evaluate the MTT, a number of tracks generated by image-based and geospatial tracking are examined and the median TLE’s are recorded. This is repeated for two data sets containing vehicles only and dismount only. The observations are summarized in Table I. The improvement in TLE from image-based to geospatial tracking is as expected. The dismount tracks have higher TLE compared to vehicle tracks because of lower clutter density in the former data set.

TABLE I

<table>
<thead>
<tr>
<th></th>
<th>Image-based Tracking</th>
<th>Geospatial Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td># of tracks</td>
<td>Median TLE (sec.)</td>
<td># of tracks</td>
</tr>
<tr>
<td>Vehicles</td>
<td>421</td>
<td>170</td>
</tr>
<tr>
<td>People</td>
<td>256</td>
<td>124</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

A MTT approach is described that integrates target model estimation and data association. In this framework, a target is described by “weak” models of kinematics, shape and appearance. These models are refined in a two-stage process: image-based tracklets of high purity and accuracy are generated, and geospatial tracks are extended from these tracklets. During each stage of tracking, observation data of reduced uncertainties are associated with the refined tracks in a probabilistic manner. The target models at image level use image-based features that are continuously adapted, allowing tracking
The stopped target tracking can be improved by including photometric (intensity/color) information in the appearance model, such as using spin image representation. The track re-acquisition step can benefit from the use of a more robust feature set and a matching strategy through better on-line, such as semi-supervised, learning. GIS information can be used to augment detection and tracking to further limit false alarms.

In order to significantly improve TLE in large semi-urban and urban environments, our approach has to be extended to use traffic flow theory with advanced modeling and sensor agnostic flow-based tracking algorithms while still operating in real-time. Our current design interfaces with a gimbal and thus our tracks will be able to steer the sensor to maximize observation when improving the TLE is the goal.

REFERENCES


