Context-Aided Tracking with an Adaptive Hyperspectral Sensor

Andrew Rice
Numerica Corporation
Dayton, OH., U.S.A.
Email: Andy.Rice@Numerica.us

Juan Vasquez
Numerica Corporation
Dayton, OH., U.S.A.
Email: Juan.Vasquez@Numerica.us

Abstract—A methodology for the context-aided-tracking of ground vehicles in remote airborne imagery is presented in which a background model is inferred from hyperspectral imagery. The functional materials in a scene are determined and become a background model. Here, a manual method of forming the model is presented, as well as a novel autonomous method which exploits the emerging class of adaptive multiple-object-spectrometer instruments. A multiple-hypothesis-tracker is introduced, which relies on background statistics to form track costs and associated track maintenance thresholds. These statistics include detectability, track density, and false measurement density. Traditionally, these statistics are uniform constants, but the advent of the background model allows for spatially-varying statistics. The context-aided-tracker, which uses these statistics, is shown to achieve an increase in performance.

Keywords: Tracking, data association, Kalman filtering, estimation, hyperspectral, context aiding, track maintenance, multiple object spectrometer, adaptive sensing.

I. INTRODUCTION

The requirement to remotely track ground vehicles within urban environments is becoming increasingly pervasive in both civilian and military contexts. These environments are typified by high vehicle density, agility, and diversity, as well as frequent occlusions. Here, two promising areas of research are combined to cope with these challenges: context-aided tracking (CAT) and adaptive hyperspectral sensing. The availability of novel adaptive hyperspectral sensors has led to new sensor resource management methods providing spatially varying background statistics of the surveillance region. These statistics are then incorporated into a multiple hypothesis tracking system to enable more robust tracking in the presence of vehicle occlusion typically encountered in urban environments.

II. BACKGROUND

A. Hyperspectral Imaging

Hyperspectral imaging (HSI) is a sensing technique which measures light in two spatial dimensions, and one spectral dimension divided into $N$ – typically more than 100 – discrete spectral bands. The resulting signal is commonly referred to as a hyperspectral cube. The spectral content of HSI data are well suited for material classification, detecting camouflaged targets, and reacquiring previously observed targets. While there are many methods to design hyperspectral imagers, dispersive spectrometers is the type most frequently employed in remote airborne sensors. These instruments take advantage of the wavelength-dependent nature of light undergoing refraction or diffraction, as with a prism or grating. In order to use a two-dimensional electro-optical focal plane array to capture this three-dimensional image, one spatial dimension is discarded by a slit-shaped aperture whose orientation is complimentary to that of the dispersion. By mechanically scanning the sensor in the cross-slit direction – as with a moving platform or a tilting mirror – the lost spatial dimension is recovered. An obvious detriment of this approach is that at any given moment, the field-of-view of the sensor is limited by the narrow slit, and the revisit rate may be quite low. Traditional dispersive HSI is ill-posed for the challenge of remotely tracking moving vehicles in urban environments due in part to this limitation, as well as the high processing burden of full HSI cubes.

If the two spatial dimensions of the resulting image are allowed to be sparse, i.e., several discrete objects of interest, then a continuous field-of-view, non-scanned instrument can be realized. For instance, in a multiple object spectrometer (MOS), the slit is replaced with a mask consisting of an open point for each object to be observed. Constraints on the distribution of points and spectral bandpass filters ensure that the spectra from multiple points do not overlap. MOS instruments have previously been applied to sensing scenarios with deterministic platform/object relationships, e.g., astronomy.

With the advent of digital micromirror device (DMD) arrays (DMA), the Rochester Institute of Technology Multi-Object Spectrometer (RITMOS) [1] instrument replaces the aperture mask with millions of small mirrors. This has the advantage of extremely fast mask reorganization, as well as repurposing the light “wasted” by the MOS to capture a panchromatic channel. This paper discusses a notional RITMOS-inspired adaptive hyperspectral sensor applied to the urban vehicle tracking problem [2].

B. Image Synthesis

This effort employs synthetic HSI data to develop and evaluate the system. The Digital Imaging and Remote Sensing Image Generation (DIRSIG) [3] model is a first-principals-of-physics based tool useful for synthesizing remote HSI data. DIRSIG accounts for object geometry, spectra, and motion; it
uses MODTRAN [4] to apply solar radiance and atmospheric transmission effects. Objects within a DIRSIG scene are “painted” according to underlying spectral-reflectance signatures. Frequently, these are the precisely-measured spectra of real-world objects such as grass, asphalt, concrete, and car paint.

C. Multiple Hypothesis Tracking

Fundamentally, tracking is the process of deriving optimal estimates of the state of sensed objects in the presence of noise. Practical tracking systems must perform several key functions: automated track initiation/deletion, multiple-track to multiple-measurement data association, and filtering. Of particular interest to the urban tracking challenge, the complexity of the data association function grows aggressively with track density. As this association is fundamental to the topic of CAT, a technique much in the spirit of [5] will be given here.1

Several key statistics must be introduced, and are classically held as constant parameters within the tracker – indeed they can be used to “tune” the system during its development. Their importance to CAT will be described in Section III-C. The probability of detecting an object (making a measurement) conditioned on its presence is $P_D$. There exists some sufficiently small spatial area $A$ which, when observed by the remote sensing system, may be measured independently from other such regions for the presence of a moving object. Then an expected density of objects per $A$ is $\beta_{NT}$ for “new track”, and $\beta_{FA}$ for false alarm.

The track cost, which is inversely proportional to track health, is defined as a recursively summed negative-log-likelihood ratio $C$. The cost is initialized for new tracks and evolves as measurement-to-track associations (M2TA) or missed detections occur:

$$C(1) = -\ln \left[ \frac{\beta_{NT}}{\beta_{FA}} \right] \quad \text{initialization} \quad (1)$$

$$C(k) = C(k-1) + \Delta C(k)$$

$$\Delta C(k) = \begin{cases} -\ln (1 - P_D) & \text{missed detection} \\ -\ln \left[ \frac{P_D p(z_k^j|x_k^i)}{\beta_{FA}} \right] & \text{M2TA} \end{cases} \quad (3)$$

The pdf $p(z_k^j|x_k^i)$ is representative of the statistical distance between the measurement $z_k^j$ and the predicted track location $x_{k}^{i(k-1)}$ of the track $T_i$ at time $k$. The track cost is a useful method of confirming and dropping tracks.

Track confirmation helps to insulate the user from false tracks. These originate from false detections and tend to be short-lived, cluttering the presented track picture. This can be particularly troublesome where false tracks occur near valid tracks; these are called redundant tracks and can convey undue ambiguity to the user. Here, a proposed track confirmation threshold $T_{conf}$ is based upon the cost of a hypothetical track which initializes and then receives $N_{conf}$ updates. Since the update cost from Eq. (3) depends upon the pdf $p(z_k^j|x_k^i)$ – which is not known for the hypothetical track in question – a related pdf $p_{conf}^{SS}(z|x)$ is formed from some steady-state benchmark instead. Hence,

$$C_{conf} = -\ln \left[ \frac{\beta_{NT}}{\beta_{FA}} \right] - N_{conf} \ln \left[ \frac{P_D p_{conf}^{SS}(z|x)}{\beta_{FA}} \right]. \quad (4)$$

All candidate (unconfirmed) tracks are compared to this threshold and are confirmed if and when

$$C \leq T_{conf}. \quad (5)$$

Of course, $C$ contains all history for that track, which is equivalent to allowing tracks unlimited time to become confirmed.

Likewise, a threshold for dropping badly behaving tracks is necessary. Much as in Eq. (4), a hypothetical track is conceived which has missed $M_{drop}$ updates out of $N_{drop}$ observations. This implies $N_{drop} - M_{drop}$ updates, which likewise necessitates a pdf $p_{drop}^{SS}(z|x)$ from a steady-state benchmark. Hence,

$$T_{drop} = -(M_{drop}) \ln \left[ 1 - P_D \right]$$

$$- (N_{drop} - M_{drop}) \ln \left[ \frac{P_D p_{drop}^{SS}(z|x)}{\beta_{FA}} \right]. \quad (6)$$

For the new tracks under test in Eq. (5), $C$ is an appropriate cost; but for mature tracks it may contain a great deal of history and can grow without bound. Since track deletion is intended to represent events which are sudden in nature – e.g., the vehicle has entered a parking garage – a form of track cost with less memory is desirable. Borrowing $N_{drop}$ from Eq. (6) to define this window of time, the test for dropped tracks becomes

$$C \geq T_{drop}, \text{ where } C = \sum_{\tau=k-N_{drop}+1}^{k} \Delta C(\tau). \quad (7)$$

While $p_{drop}^{SS}(z|x)$ is a reasonable design, each may be tuned to a desired rate of track confirmation and deletion. Other tuning parameters, $N_{conf}$, $M_{drop}$, and $N_{drop}$ are set according to tolerance for true track confirmation latency, rate of false track confirmation, prevalence of true track premature deletion, and rate of false track deletion.

Non-trivial association events – those that consist of missed measurements, false alarms, object entrances/departures, and closely spaced tracks – frequently lack an obvious M2TA solution. Instead, there is competition between tracks and measurements which must be resolved via data association. For each M2TA, there is a resulting cost which is equal to $\Delta C$. A set of these associations – a solution – assigns measurements to tracks (taking into account missed detections and new tracks), and has a composite score accounting for all tracks, normalized appropriately. Finally, many different solutions are possible, and can be ranked according to their composite scores. The global-nearest-neighbor solver simply uses the highest ranking of the solutions at each time $k$. Since the second and lower ranked solutions are discarded,

1Notably, the underlying object dynamics model is a discrete-time Markov process; the measurement model is a linear system of the object state. Both models incorporate zero-mean, white, Gaussian noise, and are given in [5].
this method has little hope of recovering from an association error.

In contrast, the multiple hypothesis tracker (MHT) preserves many of the suboptimal solutions. This results in mutually exclusive hypotheses whose tracks may disagree with respect to M2TA history. To the extent that the various hypotheses capture different solutions to any given assignment problem, that problem’s decision is “deferred,” and its solution is “soft.” A decision outcome may rise or fall in favor as its associated hypothesis incorporates new information over time. Practicality demands that the weaker hypotheses be pruned away, and that a finite length of history is maintained. Thus, the outcomes of any association problem will eventually reduce to one, which becomes “firm.”

D. Classification

Classification is the application of supervised machine-learning techniques to optimally assign identity labels to observed objects. While its applications are broad, here classification will be focused on the HSI domain. Furthermore, while the classification of HSI moving-vehicle signatures is of great utility within feature-aided-tracking (FAT) research [6], this paper will focus on background classification for CAT.

A general HSI classification architecture begins with data pre-processing and has a training stage followed by a utilization stage. Preprocessing is concerned with transforming the spectral dimension of the data from \( N \) bands (in a native radiance space), to an \( N_F \) dimensional feature space. Although not mandatory, this commonly includes radiance-to-reflectance conversion and/or dimensionality reduction, such that \( N_F \ll N \). This reduction arises from the knowledge that, for materials of interest, portions of the \( N \) bands are highly correlated. Also, some bands may have very poor signal-to-noise characteristics, e.g., water absorption in the atmosphere, and should simply be dismissed. In training, a model is derived from spectra with known or “supervised” material identity; discrete materials become “classes.” In parametric classifiers, this model assumes some underlying density function – usually Gaussian; the quality (and even attainability) of the derived parameters is subject to the population size of the training data and the goodness-of-fit to the assumed distribution. Non-parametric classifiers, such as the generalized-relevance-learning-vector-quantization-improved (GRLVQI) method [7], relax assumptions of the underlying distribution and tend to tolerate smaller training populations. The GRLVQI is a gradient-descent neural-learning method. As it trains, it uses differential-shifting to manipulate prototype vectors within the feature space. This training can become computationally expensive, although adaptations in [8] have led to significant improvements. Finally, the utilization stage of classification tests unknown signatures against the model and declares the identity of each – or defers in case of low confidence. In parametric classifiers, this test chooses the class that minimizes some statistical distance. In the case of the GRLVQI non-parametric classifier, this test chooses the class with the closest individual prototype vector.

Here, the GRLVQI classifier is applied to the urban HSI CAT challenge; the selection is due mainly to its robustness within an autonomous system.

III. CONTEXT-AIDED TRACKING

The premise of CAT is that applied knowledge of a vehicle’s environment may improve the performance of the tracker. This knowledge may be available as prepared data in a geographic information system (GIS) or inferred real-time from the surveillance data itself. The later is attractive when GIS data are outdated, denied, or difficult to co-register with the surveillance imagery.

This section will first describe an offline method of inferring context which assumes availability of full-scene HSI data and modest operator intervention. Then a novel, autonomous, online method will be introduced and applied to adaptive HSI sensing. Finally, the background statistics will be directly incorporated into the MHT to provide an innovative method for context-aiding.

A. Background Modeling

The precursor to CAT is the background model, a spatial map of the scene materials which are functionally relevant to the tracking algorithm. Here, a technique similar to [9] is used to convert HSI data into such a model; an illustration is given in Figure 1. A HSI cube\(^2\) of the scene is obtained and rectified such that the mission imagery can later be registered to the resulting model. Several bands with moderate to severe atmospheric \( \text{H}_2\text{O} \) absorption are discarded, e.g., \( 0.93 \mu \). First, the well known normalized difference vegetation index (NDVI) [10] is used to detect pixels dominated by vegetation\(^3\):

\[
\frac{L_{0.86\mu} - L_{0.66\mu}}{L_{0.86\mu} + L_{0.66\mu}} \geq 0.18 ,
\]

where the wavelengths and threshold are manually tuned for this scene, but are in the range of commonly accepted values. Next, an empirically derived tree index is used to determine which pixels among those passing the NDVI test are dominated by tree leaves:

\[
\frac{L_{0.86\mu} - L_{0.78\mu}}{L_{0.86\mu} + L_{0.78\mu}} \geq 0 ,
\]

and assuming the remaining pixels to be grass. This index is also tuned for this scene, and may not hold for scenes with different tree and grass species. With explicit dimensionality reduction (here, from \( N = 60 \) to \( N_F = 2 \)), these index methods exploit convenient material properties in an effective and computationally inexpensive way.

Subsequent spectral processing focuses on the more challenging task of classifying remaining scene elements: roads, water, and building materials. An operator manually identifies subclasses of materials and marks small training regions within

\(^2\)Radiance to reflectance conversion is not essential here, as no reference spectra will be incorporated into this method and atmospheric effects are assumed constant across the scene.

\(^3\)Using the notation \( L_\lambda \) to mean some sufficiently narrow signal band whose response is spectrally centered about light with wavelength \( \lambda \).
the cube for those materials. A functional class, e.g., road, may have many subclasses such as concrete, asphalt, and weathered asphalt; this is necessary to keep the spectral-feature-space variances low. In this experiment, there are 15 subclasses. Ten of these subclasses account for various asphalt-shingle and gravel roof treatments; four subclasses account for road surfaces. The GRLVQI classifier algorithm is trained on this population of spectral samples with corresponding subclass labels. Next, the resulting classifier model is used to assign labels to all non-vegetation pixels in the cube. The subclass labels are then discarded and replaced with their functional parent class labels. Finally, the fusion of the results from the vegetation indeces and GRLVQI classifier becomes the initial spectral background model for the scene.

Thus far, spectral domain information has been exploited for each pixel independently. This is apparent in the presence of anomalous “speckles” in the background model. These small features are frequently the result of classification errors, perhaps due to scene materials which were excluded from the training data. At best, these are of no use for the functional model. There is further salient information in the two spatial dimensions of the cube, e.g., texture and edge contrast. Recognizing this, a spatial segmentation technique is applied to each band. Similar to a raster-scan flood-fill operation, this replaces each pixel’s radiance with that of its same-band spatial-neighbor if their radiances are within some small threshold. After this replacement, many neighboring pixels will have identical values for some majority of bands in this “stack of bands” – these are called segments. Every pixel now belongs to a single segment (or is a singleton segment itself). Those segments with fewer than some number of members are discarded by assigning them to larger neighbor segments. Finally, this spatial segmentation is fused with the initial spectral background model: for each segment, all member pixels vote for a material label based on the corresponding spectral background model. This then forms the final background model.

B. Adaptive Background Modeling

The background modeling approach discussed above is effective, but comes with a high cost: the requirement for pre-tracking-time HSI acquisition, and an offline human-in-the-loop processing stage. However, the emergence of adaptive HSI sensors – such as a RITMOS-inspired DMA-based instrument – has provided a potential path for improvement. This instrument collects a full-frame panchromatic image at each step in time. A minority of pixels can be excluded from the panchromatic image on command, via flipping micro-mirrors in the DMA, and reflected into a spectrometer. One concept of employment for such a sensor is to begin the surveillance mission by investing the time to scan a dense HSI cube solely for the sake of background modeling. While this is practical and provides a just-in-time model, it is arguably not optimal: more HSI pixels are being collected than necessary. Also, in some cases a rapidly moving field-of-view is desirable or unavoidable, leaving no time for a dense cube. What is needed is an adaptive at-any-time model which initializes with as few HSI pixels as possible, yields the best model possible at any time, and converges on the model afforded by a dense acquisition. Here, a notional DMA-based MOS HSI sensor is applied to the remote ground-vehicle CAT problem and forms the basis for adaptive background modeling.

The typical background modeling process in Section III-A is modified into a form of Bayesian inference [5] as follows. Introduce an a priori library of labelled spectral signatures representing \( N_\phi \) functional background material classes \( \phi \). Allow the underlying library representation to contain subclasses as necessary, as in Section III-A. Use the library to perform an offline training process to prepare the NDVI and tree index wavelengths/thresholds, and to train the GRLVQI classifier model, such that HSI pixels can be labelled.

The online portion begins with a full-frame panchromatic radiance image, divided into regions of similar intensity via the spatial segmentation process previously described. Since this operates on a single ( PAN chromatric) band, the ability to distinguish regions with homogenous materials will be diminished versus the “stack of bands” approach. However, this segmentation provides an initial guess at the background model, with the caveat that each segment lacks a material label.

The remainder of this discussion applies to each segment in parallel\(^4\). Assume a sequence \( Z \) of labelled, single-pixel hyperspectral observations \( z \) intersecting some segment\(^5\). One, many, or none of these observations may arrive at each time \( k \). In order to strengthen a claim of independence, we require the observations to be sufficiently separated spatially or temporally within the segment. Define the pdf that the segment has the functional material label \( \phi_i \) upon incorporation of the \( 1^{st} \) through \( n^{th} \) observations as \( p(\phi_i|Z_n) \). At some loss of optimality, uniformly distributed priors are assumed:

\[
\forall i : p(\phi_i|Z_n) = \frac{1}{N_\phi} .
\] (10)

Now define the transitional pdf of receiving a specific sequence of observations \( Z_n \) conditioned upon a true class identity \( \phi_i \) as \( p(Z_n|\phi_i) \). Due to independence

\[
p(Z_n|\phi_i) = \prod_{j=1}^{n} p(z_j|\phi_i) .
\] (11)

Note that an estimate of \( p(z|\phi) \) is empirically available as a consequence of the a priori library via a confusion matrix analysis. Now process the segment’s observations in order, recursively updating \( p(\phi_i|Z_n) \) according to

\[
p(\phi_i|Z_n) = \frac{p(\phi_i|Z_{n-1}) p(\phi_i|Z_n)}{\sum_{j} p(\phi_j|Z_{n-1}) p(\phi_j|Z_n)} .
\] (12)

For the downstream CAT functionality, it is necessary to assign a single functional material label to each segment. This time-

\(^4\)For brevity, no segment index has been added to the notation.

\(^5\)Note that the library and observations are unlikely to lie in a consistent radiance space. Radiance-to-reflectance transformation is suggested but beyond the scope of this paper; see [11] for a survey of techniques.
Fused Spatial/Spectral Segmentation
GRLVQI Classifier
NDVI, Tree index
HSI cube
Spatial/spectral background model

Figure 1. An illustration of the background model generation process which takes the HSI cube as input. The NDVI index detects vegetation in general (bright green in the left image chip). The tree index refines this specifically to trees (bright green in the right image chip). The GRLVQI classifier assigns material labels to each pixel – e.g., asphalt is shown as dark blue in this chip. The spatial segmentation operates on each band individually to form regions of similar radiance. In this chip, each color represents a different region. Finally, the four intermediate products are fused into a material map. Here, it is shown with each color corresponding to a functional material type.

Varying label is simply the maximum a posteriori label
\[ \phi_{MAP}(k) = \arg \max_{i \in [1, N_{\phi}]} p(\phi_i | Z_n(k)) , \]  

and when determined for each segment, becomes the adaptive background model.

The time-varying entropy of a segment is a measure of disagreement between the segment’s observations, and is defined as
\[ h(k) = -\sum_{i=1}^{N_{\phi}} p(\phi_i | Z_n(k)) \ln p(\phi_i | Z_n(k)) . \]  

Minimizing the entropy of all segments in the model is desirable, and two methods for doing so are now defined. Since the observations \( Z_n \) for a segment are subject to measurement noise and classification error, there is hope that as \( n \) increases, \( h(k) \) will decrease. As mentioned in Section II-A, MOS spectrometer instruments restrict the quantity and placement of the HSI pixels which can be acquired at any moment. As a simple example, if two micro-mirrors from the same column and nearby rows were steered towards the spectrometer at the same time, their dispersed radiance would likely overlap, destroying both signatures. Hence, a sensor resource manager (SRM) is now conceived. Let \( u(k + 1 | k) \) be the utility of adding a new observation \( z_{n(k) + 1} \) at a future time \( k + 1 \) for some segment, and
\[ u(k + 1 | k) \overset{d}{=} \frac{h(k)}{n(k) + 1} . \]  

The SRM must then allocate spectral observations, within the constraint, in order to maximize the summed utility. Notably, this utility function can be combined with other utility functions; in a FAT system there would be a competing desire to measure the signatures of moving vehicles.

The second entropy minimization method is a segment-maintenance function. Recognizing that the original panchromatic spatial-segmentation approach has no hope of separating some materials, there will certainly be segments with heterogeneous materials – and hence, high entropy. This is mitigated by occasionally subdividing each segment having high entropy, despite a relatively large number of observations, into a set of child regions which has lower average entropy per unit area. Notably, child regions inherit observations from their parent according to their new boundaries, such that history is not lost. The subdivision process is a nontrivial optimization problem, and will not be detailed here. The time between segment-maintenance executions is intended to be roughly one order-of-magnitude greater than the time between observation frames, encouraging entropy to approach a steady state in each segment.

C. Background Statistics

Traditional uniform statistics of the background environment (\( P_D, \beta_{NT}, \) and \( \beta_{FA} \)) are so abstract as to be difficult to estimate. As such, they often degenerate into physically meaningless – albeit important – tuning variables. Part of this difficulty arises from the application of these statistics as uniform values. The CAT paradigm, however, holds the background statistics as spatially dependent. Recognizing that \( \beta_{NT}, \beta_{CAT}, \) and \( P_D^{CAT} \) are difficult to know directly, context information is used to heuristically estimate them. Here, context is primarily a question of the functional material composition of the scene – a background model – and is developed as in Section III-A or III-B. The first-order effect on \( P_D^{CAT} \) is degree of occlusion. Assuming an airborne sensor, materials which tend to obscure ground vehicles will result in lower \( P_D^{CAT} \). Tree canopies have varying density, but in urban environments dense and multi-layer canopies are rare. In oblique viewing geometries of urban scenes, vehicles may appear to be behind buildings and rooftops. Further assuming a passive imaging sensor, solar illumination also has a strong effect on detectability, which varies spatially due to shadowing. The statistic \( \beta_{CAT}^{NT} \) is treated here as a hybrid measure of detectability and hospitality (where ground vehicles can travel). While the statistic \( \beta_{FA}^{CAT} \) may indeed vary according to material, this would be due to subtleties within the motion detection algorithm. Here, \( \beta_{FA}^{CAT} \) is held constant. An empirical analysis of real remote sensing data processed with a typical motion detection algorithm has led to the values in Table I.
These statistics are formed into a spatial map, and must be drawn according to location. For $\beta_{NT}^{CAT}$ and $\beta_{FA}^{CAT}$, the location is intuitively based upon where the measurement $z_k$ falls. For $P_{D}^{CAT}$, there is some question as to whether to draw based upon the location of the measurement $z_k$, the predicted track state $x_{k|k-1}$, or the posterior track state $x_{k|k}$. The predicted track state is always available and is a good choice, whereas the measurement is meaningless in the “miss” case of Eq. (3). The posterior track state is equivalent to the predicted state in the “miss” case, but subtly different in the “update” case. The posterior represents the optimal estimate of the track at that time, and in this work serves as the reference location for $P_{D}^{CAT}$. These statistics then lead to the context-aided track cost $C^{CAT}$ and windowed cost $C^{CAT}$. The significance of the difference between $C$ and $C^{CAT}$ is readily apparent for tracks traveling through a diverse background. Illustrated in Figure 2, $C^{CAT}$ is able to provide a more accurate assessment of track health under certain circumstances.

Also of concern are the context-aided thresholds $T_{conf}^{CAT}$ and $T_{drop}^{CAT}$, which are based on hypothetical tracks. In the case of confirmation, the hypothetical $\beta_{NT}^{CAT}$ and $\beta_{FA}^{CAT}$ are located by the initializing measurement of the track under test. However, $N_{conf}$ updates are assumed to have occurred — since they are hypothetical, their positions are indeterminate, making the selection of $P_{D}^{CAT}$ questionable. Reasonable choices include $P_{D}^{CAT}$ from the location of the most recent posterior of the track under test, the average $P_{D}^{CAT}$ of that track relative to its window $N_{conf}$, or uniformly random draws of $P_{D}^{CAT}$ throughout that window. An empirical analysis suggest that another more aggressive confirmation strategy reduces the confirmation time of real tracks without a marked increase in false track confirmation: namely the minimum $P_{D}^{CAT}$ of the track under test within its window $N_{conf}$. Likewise, $T_{drop}^{CAT}$ is based on a hypothetical track with indeterminate locations. The same potential strategies exist for selection of $P_{D}^{CAT}$. Empirical analysis suggests that the minimum within the window yields a balanced but guardedly conservative track drop threshold. Two notional tracking cases have been simulated to validate these threshold choices and to further motivate the benefit of CAT versus uniform statistics.

**Case I.** A true track, which transitions from a high $P_{D}^{CAT}$ region into a low $P_{D}^{CAT}$ region and back again. This succinctly describes a primary benefit of CAT: a reluctance to drop tracks known to exist where they are less detectable. And more subtly, to smoothly transition the drop behavior as the detectability begins to improve. An example of poor transition behavior which has been resolved by the minimum-windowed method: disproportionate drop tendency given a single missed measurement as $P_{D}^{CAT}$ rises. An illustration of this case is given in Figure 3(a). The CAT method successfully maintains track throughout this scenario, while the uniform method prematurely drops the track ten frames into the outage.

**Case II.** A false track, formed entirely of false alarms occurring within a high $P_{D}^{CAT}$ region. Although erroneous, this type of event is possible under certain circumstances. Ideally, this track will drop as rapidly as possible. This represents a sort of “control experiment,” and although CAT is not equipped to hasten the drop in this case, it preferably should not prolong it. An illustration of this case is given in Figure 3(b). Both the CAT and uniform methods drop the track ten frames after the final false alarm measurement.

### IV. Experiment Design and Results

Here, an experiment has been formed around the “Megascene” scenario developed by the Rochester Institute of Technology (RIT), and distributed with DIRSIG. Suburban in nature, if offers a moderately dense road network and plentiful occlusions. A population of 37 moving vehicles has been formed, and random waypoint navigation constrained by kinematic and traffic rules has resulted in a difficult tracking challenge.

The scenario has been rendered at 10Hz temporal sampling with a notional hyperspectral instrument mounted to an airborne platform and oriented towards nadir. Platform motion has been excluded for simplicity, but is more generally resolved with registration techniques. The observation geometry and optical design yield a ground-sample-distance of 0.5m and an overall field-of-view of 0.3km². A spectral bandwidth of $0.4\mu m$ at $0.01\mu m$ resolution results in 60 bands. This is representative of a realizable silicon-based visible-light MOS instrument. Approximately 200 frames of imagery have been rendered, accounting for 20 seconds of moving-vehicle data.
The dataset has been rendered in dense-hyperspectral mode. This provides more hyperspectral pixels than would be available from any MOS instrument, and therefore is down-sampled according to the exploitation algorithm and SRM under test. Additionally, panchromatic imagery has been derived from the spectral data, and represents the video-rate imaging channel available on the MOS instrument. A traditional frame-to-frame motion detection technique has been performed on this panchromatic channel, resulting in motion detections suitable for tracking. The intentional addition of modelled noise into the data results in false-alarm motion detections. Occlusion, illumination effects, and low-contrast vehicles result in frequent missed-detections. Two tracking experiments have been performed: a uniform-statistic “control” test, and an innovative context-aided test. The vehicle population and motion are the same for both tests.

For the uniform-statistic test, values for $P_D$, $\beta_{NT}$, and $\beta_{FA}$ were set according to the final row in Table I. These values were empirically determined and are known to produce good tracking results for this combination of scenario, motion-detector, and MHT. For the context-aided test, a full hyperspectral cube was formed from the data. The manual, offline, background modeling technique described in Section III-A was applied, resulting in a functional classification of background materials in the scene. These were converted into spatially-dependent background statistics maps according to the entries in Table I.

There are many methods for assessing the performance of a ground-vehicle tracking system. First, a precursory truth-to-track association is performed. For each true vehicle $i$ at time $k$ the sufficiently close and valid tracks form the gated-set $G(i, k)$. Valid tracks are those which have, at that time, already been confirmed and have not yet been dropped. A global-nearest-neighbor assignment is performed at each $k$. This assignment is mutually exclusive (done without replacement), and forms the function $\mathcal{I}(i, k)$ which maps true object $i$ at time $k$ to a positive, natural, track identity ($\mathcal{I}(i, k) \in \mathbb{N}$) or to no track ($\mathcal{I}(i, k) = 0$). The following is a set of well-known metrics which have been identified as most likely to demonstrate the effects of context-aiding.

Track completeness is defined as

$$
\mathcal{M}_{\text{comp}}(k) = \frac{|\{i : \mathcal{I}(i, k) \neq 0\}|}{N_{\text{true}}(k)},
$$

where $|\cdot|$ is the set counting operator – here, the valid assignments, and $N_{\text{true}}(k)$ is the number of true objects at time $k$. Notably, $N_{\text{true}}(k)$ includes all tracks within the scenario area, including those which are occluded or stopped. The average completeness of all frames, $\mathcal{M}_{\text{comp}}$, is a measure of how well the tracker “covers” every true object with tracks throughout the scenario. It lies on the range $[0, 1]$, where 1 indicates ideal coverage. Notably, track identity is of no consequence to $\mathcal{M}_{\text{comp}}$. Should a track drop and immediately be replaced by a new track on the same true vehicle, i.e., an identity-swap, $\mathcal{M}_{\text{comp}}$ is not penalized.

Conversely, track purity is not concerned with coverage, but with track identity over the entire scenario $K$:

$$
\mathcal{M}_{\text{pure}}(i) = \frac{|\{k : \mathcal{I}(i, k) = \text{mode}\mathcal{I}(i, K)\}|}{|\{k : \mathcal{I}(i, k) \neq \text{mode}\mathcal{I}(i, K)\}|},
$$

which is the ratio of the frames in which a true object is assigned its most frequently occurring identity to the frames in which it is assigned any track identity. The aggregate purity is the same ratio extended to all true objects:

$$
\mathcal{M}_{\text{pure}} = \frac{\sum_{i,k} |\{k : \mathcal{I}(i, k) = \text{mode}\mathcal{I}(i, K)\}|}{\sum_{i,k} |\{k : \mathcal{I}(i, k) \neq \text{mode}\mathcal{I}(i, K)\}|}.
$$

Figure 3. An illustration of track drop-thresholds in simulated data. In (a), the track is following on object which leaves a high $P_{D}^{\text{CAT}}$ region, travels through an occcluding low $P_{D}^{\text{CAT}}$ region, and finally re-emerges into a high $P_{D}^{\text{CAT}}$ region. Notably, the uniform cost $\bar{C}_{\text{drop}}$ exceeds the uniform drop threshold $T_{\text{drop}}$ at frame 40, resulting in track loss. However, the CAT cost $\bar{C}_{\text{CAT}}$ decreases, and the CAT drop threshold $T_{\text{CAT}}^{\text{drop}}$ increases during the occlusion. This behavior makes track loss much less likely, and the track is maintained throughout the scenario. In (b), a false track has formed on several correlated false alarm measurements in a high $P_{D}^{\text{CAT}}$ region. The CAT and uniform methods have the same behavior in this scenario, dropping the track ten frames after the final false measurement. This suggests that CAT should not penalize performance in similar cases.
This innovative approach to context-aided-tracking – in the form of spatially-dependent background statistics derived from hyperspectral data – provides a demonstrable improvement in tracking performance over traditional methods. Specifically, both the completeness and purity metrics have improved by six percentage points, while spuriousness and redundancy show no appreciable difference. Perhaps the most telling statistic is the cardinality error. During this relatively short tracking experiment, the context-aided method added one extraneous track identity, while the uniform method added five. This reduction in extraneous tracks represents a significant improvement in performance. While the completeness and purity metrics indicate a modest improvement, the ability to reduce extraneous tracks in the presence of significant occlusion is one of the most difficult aspects of ground vehicle tracking in the urban setting. As these vehicles traverse many city blocks, encountering ambiguous events, it is critical to maintain both pure and complete tracks. This ability will support both real-time analysis of vehicle tracks and forensic applications that become difficult in the presence of significant false alarm rates and repeated track drops.

Table II

<table>
<thead>
<tr>
<th>Metric</th>
<th>Uniform</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{comp}$</td>
<td>0.8092</td>
<td>0.8626</td>
</tr>
<tr>
<td>$M_{pure}$</td>
<td>0.8745</td>
<td>0.9302</td>
</tr>
<tr>
<td>$M_{spur}$</td>
<td>0.0284</td>
<td>0.0481</td>
</tr>
<tr>
<td>$M_{redund}$</td>
<td>0.8905</td>
<td>0.8765</td>
</tr>
<tr>
<td>$\Delta_{card}$</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

It lies on the range $(0, 1]$, where 1 indicates that when identity assignment occurs, it remains entirely consistent. As $M_{pure} \to 0$, identity swapping occurs more frequently.

Track spuriousness is the ratio of tracks not close to any true object, divided by the number of true objects:

$$M_{spur}(k) = \frac{N_{track}(k) - \left| \bigcup_{i,j} G(i,k) \right|}{N_{true}(k)},$$

where $N_{track}(k)$ is the number of tracks at time $k$. When $M_{spur} = 0$, every track can be explained by a true object; when $M_{spur} > 0$, some tracks are false-alarm tracks. The average spuriousness across all frames is $\bar{M}_{spur}$.

Track redundancy is the ratio of tracks close to any true object, divided by the number of true objects:

$$M_{redund}(k) = \frac{\left| \bigcup_{i,j} G(i,k) \right|}{N_{true}(k)}.$$

When $M_{redund}(k) > 1$, some true objects are being overrepresented with extraneous tracks. The average redundancy across all frames is $\bar{M}_{redund}$.

The cardinality of the truth $N_{true}(K)$ is the number of unique truth objects in the entire scenario. The cardinality of the tracks $N_{track}(K)$ is the number of unique identities assigned by the tracker over the entire scenario. The difference is

$$\Delta_{card} = N_{track}(K) - N_{true}(K),$$

which is the cardinality error, and is ideally 0.

The resulting metrics for both the context-aided and uniform tracking tests are given in Table II.

V. CONCLUSIONS AND FUTURE WORK

ACKNOWLEDGEMENT

This research was sponsored by the Air Force Office of Scientific Research (AFOSR) under agreement number FA9550-08-1-0028 (AFOSR-BAA-2007-08). The authors would like to acknowledge the contributions of Dr. John Kerekes and the Digital Imaging and Remote Sensing Laboratory team at the Rochester Institute of Technology, particularly in the domains of adaptive sensing and hyperspectral exploitation.

REFERENCES


