Integrating Hyperspectral Likelihoods in a Multidimensional Assignment Algorithm for Aerial Vehicle Tracking

Burak Uzkent, Matthew J. Hoffman, and Anthony Vodacek, *Senior Member, IEEE*

***Abstract*—Tracking vehicles through dense environments is an important and challenging task that is mostly tackled using visible and near IR wavelengths. Hyperspectral imaging is known to improve the robustness of target identification, but the massive increase in data created is usually prohibitive for tracking many targets. We present a persistent real-time aerial target tracking system, taking advantage of an adaptive, multimodal sensor concept and blending the hyperspectral likelihoods with kinematic likelihoods in a multidimensional assignment framework. The adaptive sensor is capable of providing wide field of view panchromatic images as well as the spectra of small number of pixels. The proposed system does not require large amount of hyperspectral data collection as we focus on tracking fewer number of targets with higher persistency. This overcomes the data challenge of hyperspectral tracking by following dynamic data-driven application systems (DDDAS) principles to control hyperspectral data collection where most beneficial. The DDDAS framework for controlling hyperspectral data collection is developed by incorporating prior information from the filter movement predictions and information from motion detection. The proposed multidimensional hyperspectral feature-aided tracker is compared to a 2-D hyperspectral feature-aided tracker and another cascaded hyperspectral data based tracker by generating a synthetic, realistic, aerial video on a dense scene.**

***Index Terms*—Aerial tracking, adaptive sensor, adaptive prediction, hyperspectral features.**

I. Introduction

VIDEO surveillance in urban environments has been extensively studied in the literature. Traditional narrow field of view (FOV) tracking systems are based on the images provided by commercially available RGB cameras. In these cases, appearance, texture, or motion-based features can be descriptive to distinguish objects [1]–[3]. Unfortunately, such methods are not effective for wide-area surveillance systems due to poor resolution imagery. There has been a large volume of studies in wide FOV aerial tracking using state-of-the-art wide-area motion imagery sensors [4]–[7]. However, the desired persistency level has not yet been reached. Alternatively, a considerable number of other studies focus on using ground moving target indicator radars, which measure the locations of moving ground objects as well as their doppler velocities [8]–[11]. With radar, however, we are limited to only kinematic information to achieve persistent tracking. Although the sampling rate is high, kinematic data alone is more likely to fail in urban scenes with large scale occlusions and clutter. To achieve a more robust system, one can utilize more distinctive hyperspectral signatures of materials.

Hyperspectral imaging involves recording extended data in hundreds of narrow, adjacent hyperspectral bands. Extended reflectance profiles of different materials can then be further processed to identify targets. Unfortunately, it is impossible to transmit or process such high-rate volume of data in real time from traditional hyperspectral sensors. With recent advancements in sensor technology, it is becoming possible to quickly collect only small FOV hyperspectral data. As a result, such a sensor could be employed in a real-time system to augment traditional aerial tracking modalities with small amounts of hyperspectral data. An example of such a sensor, that we consider in this research, is the Rochester Institute of Technology multiobject spectrometer (RITMOS) [12]. RITMOS can be tasked to observe hyperspectrally in only targeted areas where the information is forecasted to be most beneficial by the tracker. Meanwhile, it is also capable of capturing a wide FOV panchromatic image of the scene that can be used to detect motion and register subsequent images.

Some previous work has been proposed to address the challenges of detecting, identifying, and tracking targets in dynamic scenes with the aid of an adaptive hyperspectral sensor. Wang *et al.* [13] proposes a novel sensor capable of providing an omnidirectional view and acquiring hyperspectral data efficiently. They determine moving targets by subtracting panoramic images in low resolution and then record hyperspectral data in a region of interest (ROI) determined by the tracker to identify the target. However, the proposed multimodal sensor based tracking system is not tested on a challenging scenario including intersections, severe obscurations, or parallax. Blackburn *et al.*[14] uses commercially available line scanning hyperspectral sensors and panchromatic image to achieve persistent tracking. The sensor is tasked to capture hyperspectral information of the ROI determined by the prior target density. Features are generally updated several seconds apart whereas a kinematic update is performed 10 times/s. This concept might fail in heavily cluttered scenes when only kinematic information is updated. Another study that considers a performance-driven sensor to achieve continuous tracking is [15]. They propose a novel utility

Manuscript received August 28, 2015; revised December 20, 2015 and March 26, 2016; accepted April 26, 2016. Date of publication May 17, 2016; date of current version September 30, 2016. This work was supported by the Dynamic Data Driven Applications Systems Program, Air Force Office of Scientific Research, under Grant FA9550-11-1-0348.

B. Uzkent and A. Vodacek are with the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: bxu2522@rit.edu; vodacek@cis.rit.edu).

M. J. Hoffman is with the Department of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: mjhsma@rit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2560220

function to perform spatial sampling for hyperspectral data acquisition. The utility function assigns a usefulness value to each pixel. Hyperspectral data are then recorded if the pixel is determined to be useful enough based on an empirically determined threshold. However, this study assumes perfect measurements. In cluttered scenes, false alarms or missed measurements are extremely likely to occur which might result in frequent track losses or switched track identities. In [16], the same sensor concept is utilized to fuse kinematic and hyperspectral likelihoods in a 2-D assignment algorithm. The tracker achieves persistent tracking in the case where there are not many similarly painted vehicles to the TOI. However, with the increasing number of hyperspectrally similar vehicles, target motion evolution cannot be approximated in a 2-D approach which in turn confuses the tracker. In [17], hyperspectral and kinematic outlier elimination modules are cascaded to reduce the clutter density. Next, a multidimensional assignment algorithm (MDA) is employed to better approximate target motion evolution. However, this approach requires optimum hyperspectral threshold setting to achieve persistent tracking.

This paper builds upon the works performed in [16] and[17] and proposes a novel hyperspectral and kinematic likelihoods fusion method in a multidimensional assignment framework. The assignment scores are refined by the proposed formulation of the density of hyperspectrally similar vehicles in the considered short-time window. This way, target losses and reappearances are modeled more appropriately. By considering hyperspectral likelihoods in a time window we can relax the optimum threshold setting requirement in [16] and [17]. This avoids the need to manually assign optimum thresholds for different vehicles, instead, a consistent global threshold is selected empirically. Finally, a synthetic aerial video is generated considering a similar camera mounted on top of a tall building (Fixed Platform). The digital imaging and remote sensing image generation model (DIRSIG) software is used to simulate such scenario and a large number of vehicles are tracked at separate runs to evaluate the tracker compared to [16] and[17] where only a few vehicles are demonstrated.

II. Adaptive Hyperspectral Sensor and Resource Management

A sensor capable of rich hyperspectral information can be very beneficial in consistent aerial tracking of objects. Thus, RITMOS is utilized as an adaptive performance-driven sensor to detect, identify, and track targets in highly cluttered scenes. It was originally designed in 2003 as an astronomical spectrometer and imager connected to a telescope to perform hyperspectral classification of stars [12]. RITMOS utilizes a micromirror array to reflect the light to the one of two sensors; spectrograph and panchromatic channel. Mirrors in the array can be flipped very quickly to switch the individual pixels from the panchromatic to the hyperspectral data mode. To capture the panchromatic image of a scene, an array of micromirrors reflect the light to a panchromatic imaging array. Individual micromirrors imaging the object are then tilted to reflect the light to collect the full spectra of the desired pixel. All these configurations can be performed in real time via a user controlled or automated computer interface.

The performance-driven tracking algorithm needs to be designed in a way that it matches the specifications of RITMOS. It requires about 0.1 s to obtain a panchromatic image of a scene by RITMOS. On the other hand, the full spectra of a single pixel in the visible to NIR wavelength takes 1 ms. Spatial and hyperspectral data can be collected simultaneously as long as micromirrors are tasked properly. The frame rate 1 s is used to update sufficient track estimates. Panchromatic images are used to accomplish motion detection. While the sensor collects panchromatic image, the tracking algorithm outputs the forecasting results that will be fed to the sensor resource management system for hyperspectral data acquisition. Full spectra for 100 pixels can be collected in about 0.1 s. The workflow of the tracking algorithm matching the sensor specifications is demonstrated in Fig. 1.

The total time allocated for hyperspectral data acquisition occupies a large fraction of the tracking workflow since it is still costly. Hyperspectral data is acquired in a 0.4 s time interval resulting in the acquisition of the full spectra of as many as 400 pixels. Plenty of time is also allocated for the feature matching and data association parts since hyperspectral data processing can be expensive. Within this framework, one can track multiple targets as long as they do not require an excessive number of Gaussian terms in the Gaussian sum filter (GSF). Also, one can adaptively refine Gaussian terms based on the number of tracks or uncertainty associated with each term. Terejanu *et al.*[18] explain how to adaptively refine/remove Gaussian terms at each time step.

III. Scenario Generation

The synthetic aerial video generated by the DIRSIG model is used to evaluate the proposed system. DIRSIG is designed to generate datasets in a number of modalities such as multispectral, hyperspectral, polarization, and synthetic aperture radar through the integration of the first-principles-based radiation propagation modules. However, moving vehicles cannot be simulated internally in a realistic way. Thus, the simulation of urban mobility (SUMO) traffic simulator has been integrated with DIRSIG to produce dynamic imagery for tracking scenarios [19]. Fig. 3 displays a DIRSIG generated nadir RGB image and snapshot of the traffic simulation with SUMO. Pedestrian simulation is not performed as the pedestrians are typically represented by only a few pixels in low-resolution aerial images. Generated hyperspectral data consist of 61 rectangular bands within the range of 400 nm to 1 μm with a hyperspectral resolution of 10 nm. We are interested in tracking four different vehicles at separate runs. Trajectories of these vehicles of interest can be found in [17]. The average ground sampling distance is set to be 0.7 m, yielding low resolution images. A vehicle is represented with less than 50 pixels (fig. 2) which prevents us from using appearance-based methods to detect vehicles.

Synthetic hyperspectral data are further processed to consider factors in the real-world phenomenology. These factors are filter effects, shot noise, readout noise, integration time, detector

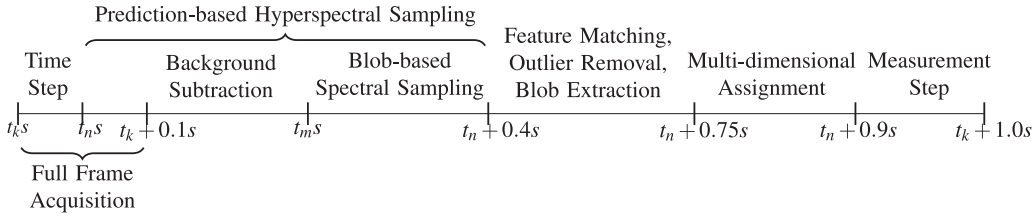

Fig. 1. Workflow of the proposed target tracking system using a performance-driven sensor.

Fig. 2. Two different vehicles, covering around 40–50 pixels in the panchromatic image, are shown traveling in the generated scene.

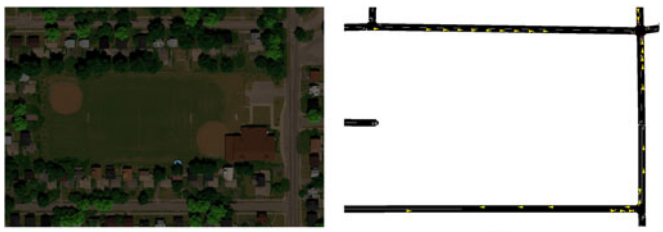

Fig. 3. (a) An RGB image (394×573 pixels) of the area of interest with only background objects. (b) Vehicular traffic simulation in the same scene with SUMO.

elements, and analog to digital converters. In the first step, sensor reaching radiance values ($\frac{W}{m^2 sr}$) are converted to irradiance values ($\frac{W}{m^2}$) on the focal plane surface by the camera equation, $E = \frac{L}{G\#}$. Since, we assume a simple lens model in this study, $G\#$ can be formulated as $G\# = \frac{4(f)^2}{\pi}$. The focal length of the lens is denoted as f and is an input parameter provided by the user in DIRSIG simulation.

The irradiance, E, at the focal plane of the lens is converted to a voltage signal, S, as

$$S = \frac{E t_{\text{int}} A_d \frac{\partial V}{\partial N} Q E \lambda^2}{hc}[V] \tag{1}$$

where t_{int} and A_d denote the integration of the sensor and the detector area, respectively. The charge to voltage ratio factor is denoted by $\frac{\partial V}{\partial N}$ whereas QE corresponds to the quantum efficiency of the detector. Finally, h is Planck's constant and c is the speed of light in a vacuum.

The voltage values are then modified to simulate two significant noise types; Readout and Shot noise. These two noise sources account for approximately 95% of the overall noise in a real-world detector element. The readout noise is a consequence of the imperfect operation of electronic devices. Conversion of the same pixel with the same induced charge does not always result in the same value from the analog to digital converter. This is one aspect of the readout noise. The second aspect is the insertion of random noise by the sensor which is added on the pixel charge. The readout noise is modeled as

$$N_{RO} = N(0, 30^2)\frac{\partial V}{\partial N}[V] \tag{2}$$

where N is a normal distribution and $30e^-$ is a common variance value that also aligns well with RITMOS.

The other major source type is shot noise, which occurs due to the random arrival nature of photons. In other words, the number of photons striking the sensor in a random time interval is likely different than the number at another, equal length, time interval. This noise factor becomes dramatic when the signal is low. The shot noise increases with increasing signal levels, however, its impact degrades since the signal level increases at a higher rate. We model shot noise using the Poisson distribution with mean (λ) equal to the variance

$$N_S = S - \frac{e^{-S}\lambda^S}{S!}[V]. \tag{3}$$

Saturation arises when a maximum signal level is reached in the readout procedure. We will refer to this noise with V_{SAT}.

Finally, the overall voltage values are converted into a quantized number S_F which accounts for the quantization and saturation noise

$$S_F = \left[2^b \frac{\max\{\min\{S + N_{RO} + N_S, V_{\text{SAT}}\}, 0\}}{V_{\text{SAT}}} + 0.5\right] 2^{-b} \tag{4}$$

where b is the number of bits used in the simulation. In this process, we do not model dark current noise since we do not consider it as a primary noise source. An extensive treatment on this radiometric sampling process is given in [20].

IV. Motion Detection

A. Median Filtering Based Background Subtraction

Background subtraction consists of subtracting the input image from the modeled background of the scene of interest. This computationaly efficient method works well in fixed platform scenarios. In nonstationary platforms, it is susceptible to parallax errors in nonplanar scenes. In this case, either additional processing is required to mitigate parallax noise [4], [21] or background subtraction needs to be avoided. As we investigate the contribution of hyperspectral likelihoods to improve identification accuracy and reduce clutter in large-scale dense scenes, we generate scenarios from a fixed platform and apply the background subtraction method.

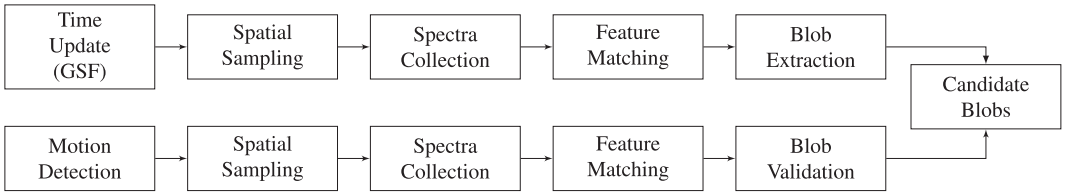

Fig. 4. Process to get candidate blobs.

The new input image is denoted as $I(x, y, k)$ whereas $I_B(x, y, k)$ represents the background frame for the given x and y coordinates of each pixel at time step k. The result, $I_D(x, y, k) = abs(I(x, y, k) - I_B(x, y, k))$, is filtered with a predetermined threshold. The major challenge lies in generating a background frame containing only background objects. Different techniques have been developed for background modeling including Adaptive or Fixed Gaussian Mixture Modeling, and Mean or Median filter [22], [23]. In Gaussian Mixture Modeling, each pixel $I_B(x, y)$ is represented by a mixture of Gaussians. The new pixel value $I(x, y, k)$ is compared with the Gaussians representing that pixel for classification. This method has proved to be robust against lighting or scene changes, however, it is not feasible for this study since it is computationally expensive. Mean or Median filters consider some number of past frames to generate a background frame by averaging out moving objects. These filters are computationally cheap which makes them viable for our tracking system. They are especially applicable for background dominated scenes. Mean or Median filters perfectly fit in our system since we only consider vehicular simulation. The background frame is then estimated as

$$I_B(x, y, k) = \text{median}\{I(x, y, k - i)\}, \quad i = 0, 1, \ldots, n - 1. \tag{5}$$

In this study, we consider nine previous scans in addition to the current scan to generate a background mask.

B. Prediction Statistics and Hyperspectral Features Based Blob Extraction

One major drawback of the median filter approach is that it is not effective in detecting slow or stopped vehicles. In order to mitigate this effect, we rely on the strength of the hyperspectral features of the target. The kernels whose hyperspectral profiles match to the TOI hyperspectral profile is stored and used to extract virtual blob or blobs. Each kernel provides the predicted coordinates of the target in addition to dimensions and velocity components. This way, each matched kernel can be used to construct a virtual blob, however, this might lead to underestimating the detected virtual blobs in data association as there might be large number of closely located blobs. This can be avoided by designing a clustering method. However, it also should be effective in separating kernels not in close proximity to each other. In this direction, we propose a clustering method based on the transition matrix types applied to each matched kernel in the GSF. These transition matrices are listed below.

1) Constant Velocity Model.
2) Nearly Constant Coordinated Turn Model (Left Turn).
3) Nearly Constant Coordinated Turn Model (Right Turn).
4) Stop Model.

More details on these transition matrices can be found in Section V. The hyperspectrally matched kernels that were applied the same category transition model is assigned to the same cluster. For instance, matched kernels with constant velocity models are clustered and a single virtual blob is extracted by averaging all the blobs in this cluster. The same process is repeated for the ones assigned Stop, Left, and Right Turn Models. This way we exploit the fact that kernels with the same motion models are more likely to be located closely compared to different motion models. By designing hyperspectral data-based virtual blob extraction, we tackle both the problem of detecting a TOI that stops for a long time and other false negatives for TOIs in the background subtraction results. Finally, the resultant foreground mask is applied the morphological closing operation to remove the noise due to illumination changes, tiny structures, and other sources and to fill in the gaps. The connected component analysis is then applied to uniquely label extracted blobs. This step is explained in more detail in [17] and the overall motion detection scheme can be visualized in Fig. 4.

V. Intersection Mask-Aided GSF

The GSF approximates a nonGaussian distribution by a finite mixture of Gaussian distributions. This way, nonlinear pdfs can be represented more accurately. The number of Gaussian kernels in the GSF is crucial in approximating complex motions. One can use a large number of kernels to ensure robust tracking, however, it results in increased computational burden. In [16], the optimum number of Gaussian kernels is reported be around 30 and we implement the same number of kernels in this study. Initially, the Gaussians are placed within three standard deviation of the center of the user selected target and each Gaussian is assigned the same weight and covariance matrix initially. The velocities are initiated as zero. This can be risky in cluttered scenes, however by eliminating vehicles through hyperspectral matching, the risk of losing the target initially is minimized since the clutter density is reduced. More details on the implementation of the GSF in a similar framework can be found in [24].

In this paper, specific motion models are adaptively removed/inserted in the GSF based on the external intersection map. Intersection extraction in panchromatic imagery is proposed in [25]. In [26], we applied an adaptive prediction model set based on the intersection type and location. However, the intersection map extraction algorithm might be susceptible to some errors due to mismatch between the OpenStreetMap data

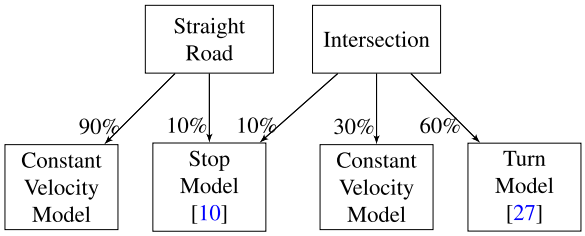

Fig. 5. Adaptive motion model set approach.

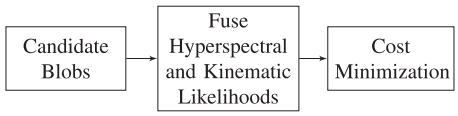

Fig. 6. Proposed assignment process.

and the surveillance scene. Thus, a model set in an intersection is implemented independent of the intersection type. The adaptive motion model set is shown in Fig. 5.

VI. Target Identification/Data Association

In this section, we propose a novel two-step data association method to robustly assign measurement to the corresponding tracks. Candidate targets are extracted after the background subtraction module. A physical gate is drawn by the mixture pdf to remove outliers. In the first step, the remaining outliers in terms of hyperspectral features are pruned by the feature matching method. Further reduction of validated measurements enables us to implement a costly MDA [28]. Without excluding outliers, the MDA algorithm would be too costly and not feasible for our tracking system. In the second step, assignments are accomplished based on the fusion of kinematic and hyperspectral likelihoods of the hyperspectrally and kinematically matched measurements. The work flow to get the candidate blobs and perform assignments are shown in Figs. 4 and 6.

In Fig. 4, the process to compute candidate blobs to be fed to the data association step is highlighted. Hyperspectral samples are acquired at detected motion areas and GSF-predicted target positions. The vector-to-vector feature matching is performed to eliminate unlikely motion blobs. Additionally, prediction led features are compared in a similar fashion to extract virtual possible target blobs as explained in the final paragraph of Section IV-A. This step is required as the median filtering-based reference model does not adequately account for stopped vehicles. In a traffic light dominated scenario, this drawback results in track fragmentation and loss.

A. Feature Matching—Hyperspectral Likelihood Estimation

In complex environments, targets are often lost when vehicles move under trees, behind buildings, or near other moving objects. In these cases, it is important to re-establish the track when a new detection occurs, as opposed to treating the redetected vehicle as a new object. A kinematic-based tracker might get easily confused due to uncorrected prior estimates and associate a series of wrong measurements. By eliminating hyperspectrally different cars, we can remove the measurements that might confuse the association algorithm. This way, we increase the likelihood of redetecting a target once it becomes visible. Without the feature matching process, we might end up with a high number of validated measurements which might lead to track termination, wrong association, and a more complex cost minimization problem in MDA. This is especially true in cluttered scenes.

The sensor is first tasked by a sensor resource management system to collect hyperspectral data in the vicinity of the center of the prior density kernels. This process is sensitive to the multimodal target motion density. If the motion is not approximated well we may end up with no target material spectra. For this reason, the sensor is additionally tasked to capture spectra around the center of the detected motions from the background subtraction module. The formulation of the proposed feature matching method is detailed in [16].

B. Multidimensional Assignment Algorithm

The MDA, first proposed by [29], is known as the real-time implementation of the multiple hyphothesis tracking filter. Originally, it was designed to handle S lists of measurements from multiple sensors. For this reason, it is also called as an S-D assignment algorithm. Later, it was formulated as a sliding window that only considers the time steps within the window to associate the measurements [30], and referred to as the multiple hypotheses tracking algorithm with a sliding window. In 2-D assignments, only the last scan, k, is considered for assignment. This methodology suffers from the lack of time depth which in turn might result in frequent jumps to wrong measurements in nonlinear motion cases.

A binary assignment function δ is defined as

$$\delta(k,\{m_s\}_{s=k-S+2}^{k},v)=\begin{cases}1 & \text{a series of measurements are}\\ & \text{assigned to } T^v(k-S+1)\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

where $m_s=0,1,\ldots,M(s)$ and $v=0,1$. M denotes the hyperspectrally validated measurement list (see Section VI-A) at scan k and $m=0$ and $v=0$ correspond to the dummy measurement[1] and nonexistant target. T represents a valid track at $k-S+1$. The multidimensional assignment algorithm is then formulated as

$$\beta(k|a)=\sum_{m_{k-S+2}=0}^{M(k-S+2)}\sum_{m_{k-S+3}=0}^{M(k-S+3)}\sum_{m_{k-S+4}=0}^{M(k-S+4)}\ldots..$$
$$\sum_{m_k=0}^{M(k)}\delta(k,\{m_s\}_{s=k-S+2}^{k},1)c(k,\{m_s\}_{s=k-S+2}^{k}) \tag{7}$$

where a contains the candidate tuples (set of measurements) and c is the cost function representing the single time-step association costs. The goal is to find the tuple minimizing the overall

[1]Dummy measurement refers to assignment of no physical target or measurement ($i=0$) to the track, T^v, whereas nondummy measurement refers to assignment of a detected target or measurement ($i>0$).

cost function. The association cost formula in our case is given by

$$c(k,\{m_s\}_{s=k-S+2}^{k}) = -ln\left(\frac{\phi(k,\{m_s\}_{s=k-S+2}^{k},1)}{\phi(k,\{m_s\}_{s=k-S+2}^{k},0)}\right) \quad (8)$$

where ϕ represents the association likelihoods. For $v=1$ and $v=0$, it is estimated as

$$\phi(k,m_s,v) = \begin{cases} \prod_s (1-P_D)^{1-u(m_s)}(P_D\,\tau(s,m_s))^{u(m_s)} & v=1 \\ \prod_s V^{-u(m_s)} & v=0 \end{cases} \quad (9)$$

where τ and V are the likelihood function and the volume of the surveillance area and P_D represents the detection probability of the target. We keep the gate (ROI) in the scan k larger than 2-D assignment case to avoid dealing with a tuple containing a measurement that was not selected for hyperspectral data acquisition. The binary function $u(m_s)=0$ in the case of a dummy measurement and $u(m_s)=1$ otherwise. We combine the kinematic and hyperspectral likelihoods as

$$\tau = \tau_k * \tau_f \quad (10)$$

where τ_k and τ_f represent the filter and hyperspectral feature based likelihoods, respectively. Estimation of τ_k is given in detail in [30]. To derive the hyperspectral likelihoods, we not only consider hyperspectral likelihoods but also the number of hyperspectrally matched objects ζ to the TOI in the gate at $k-S+2$ and $k-S+1$. Hyperspectral likelihoods for dummy measurement assignments are then formulated as

$$\tau_f(s,\{m_s\}_{s=k-S+2}^{k}) = P_S(\zeta_s,\zeta_{s-1}) \quad (11)$$

$$P_S(\zeta_s,\zeta_{s-1}) = \begin{cases} \dfrac{1}{\zeta_s+1} & \zeta_s \geq \zeta_{s-1} \\ \dfrac{\zeta_{s-1}-\zeta_s}{\zeta_{s-1}+1} & \zeta_s < \zeta_{s-1} \\ 1 & \zeta_s = 0 \end{cases} \quad (12)$$

where P_S is the probability of associating a measurement to the track T_{k-S+1}^{v}. As seen in Fig. 7, dummy measurement probability in the first case decreases with the increasing number of candidate vehicles in time step k. In the second case, dummy measurement probability increases with the increasing value of ζ_{s-1}-ζ_s. For nondummy measurement assignment, the hyperspectral likelihood is designed as

$$\tau_f(s,\{m_s\}_{s=k-S+2}^{k}) = P_S(\zeta_s,\zeta_{s-1})F_s^{m_s} \quad (13)$$

$$P_S(\zeta_s,\zeta_{s-1}) = \begin{cases} (1-\dfrac{1}{\zeta_s+1})F_s^{m_s} & \zeta_s \geq \zeta_{s-1} \\ (1-\dfrac{\zeta_{s-1}-\zeta_s}{\zeta_{s-1}+1})F_s^{m_s} & \zeta_s < \zeta_{s-1} \\ 0 & \zeta_s = 0. \end{cases} \quad (14)$$

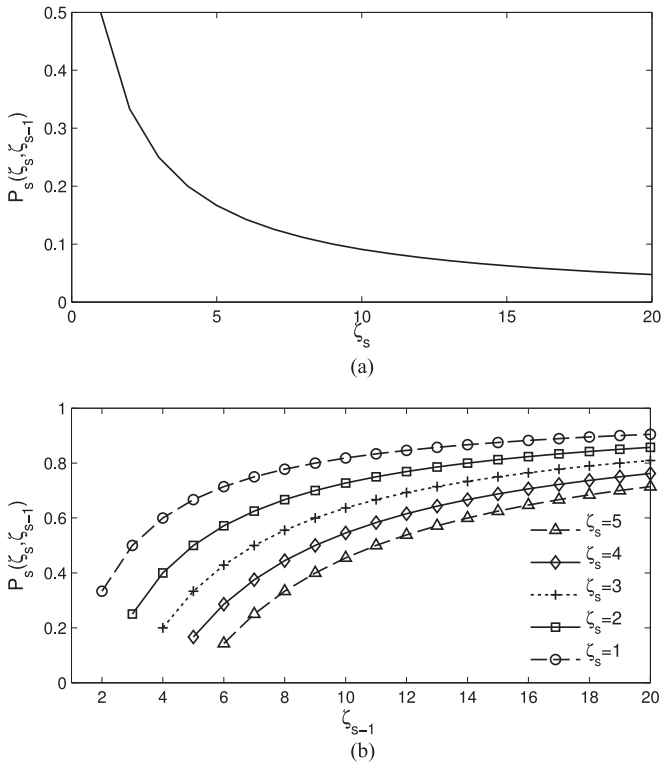

Fig. 7. First case ($\zeta_s \geq \zeta_{s-1}$) in (12) is denoted in (a) whereas (b) represents the second case ($\zeta_s < \zeta_{s-1}$).

F denotes the normalized hyperspectral likelihoods and formulated as

$$F_s^{m_s^j} = \frac{Th - f_s^{m_s^j}}{\Sigma_{i=1}^{M(s)} Th - f_s^{m_s^i}} \quad (15)$$

where f and Th denote the hyperspectral matching score and predetermined threshold used to filter hyperspectral outliers.

In this proposed framework, lost targets due to occlusions and other factors can be better handled to minimize jumps to wrong measurements. Finally, the association costs are fed to the Lagrangian relaxation algorithm in the multitarget tracking case. Since, we are interested in single target tracking, we do not deal with the one-to-one measurement-to-track constraint and use the tuple minimizing the cost function.

VII. Simulation Results

The proposed system is tested on four different vehicles (see Section III). The first track is a white painted vehicle with the length of 74 frames. It goes through a series of occlusions and stops at an intersection nearby other white vehicles. The second and third tracks belong to red and blue painted vehicles with lengths of 86 and 89 frames, respectively. They follow similar trajectories to the first track but do not travel nearby a large amount of similarly painted vehicles. Meanwhile, the fourth track with a length of 76 frames follows a different path with less severe occlusions and stops at an intersection for a long time. This scenario is challenging because of their long stop.

Fig. 8. Generated tracks to test the proposed approach. Each window has a size of 20×20 pixels. Initially, the pixels at the vicinity of the central pixels of each target are sampled hyperspectrally to build hyperspectral target features.

We report detailed results on these tracks. In addition, we test the proposed algorithm with 37 automatically generated tracks with an average length of 64.35 frames to observe how well it generalizes to different scenarios. The first frame of these tracks are displayed in Fig. 8. We applied three criterion in track generation routine as listed below.

1) The target is not fully occluded in more than two frames in the initial five frames.
2) The life of a track must be longer than 30 frames.
3) The target must be in the FOV in all frames of interest.

A high-fidelity scene is accomplished by following the radiometric sampling procedure to DIRSIG output. The panchromatic images have ≈ 29 db peak signal-to-noise ratio. Noise on hyperspectral data is measured by the SAM metric. The mean SAM value between the processed and true hyperspectral data is $\approx 2.8^{\circ}$. These values meet the RITMOS specifications.

Most of the performance evaluation metrics in tracking literature are useless in our case since different vehicles are tracked at separate runs. Thus, we focus on two metrics to measure tracking rates. The first metric is track purity (TP). It measures how many frames a tracker maintains a correct track identity within an estimated gate of the actual target position during the track life. The second one is the Target Purity (TgP) metric. It measures the ratio of the maximum number of times a ground truth is associated to a dominant track to the duration of ground truth. The TP metric, which only considers the track life, favors short tracks. On the other hand, the TgP metric considers the life of the true track. In this case, the TP score for a track has to be larger or equal to the TgP score for the same track.

The proposed hyperspectral feature-aided tracking (FAT) with multidimensional assignment algorithm is compared to the (1) FAT system with a 2-D assignment algorithm [16] and (2) another similar FAT system with multidimensional assignment algorithm [17]. Uzkent *et al.*[16] integrates the hyperspectral and kinematic features in a similar fashion in a 2-D assignment algorithm. Uzkent *et al.*[17] only utilizes hyperspectral features to eliminate outliers and employs the traditional MDA algorithm. Additionally, we compare the proposed approach to (3) kinematic only tracker (KT) and (4) hyperspectral only tracker (ST). In the KT method, τ_f in (10) is ignored whereas in ST, only SAM scores of the filter validated measurements in the last scan are considered. In the FAT and KT cases, experiments are performed with $S = 2, \ldots, 6$ (sliding window length) in the S-D assignment algorithm.

A hundred Monte Carlo runs were carried out for each TOI to minimize the randomness effect on the results. In each experiment, the target detection probability P_D is randomly drawn from the interval $0.7 \leq P_D \leq 0.9$. A track is terminated when it has not been associated with any measurement for more than 7 s. Table I, shows the TP and TgP scores for all cases. The best overall results for the FAT and KT are accomplished with the 6-D assignment.

The KT method struggles in the presence of vehicles with similar trajectories in large scale dense urban scenes. In addition, severe obscurations have a higher impact on the KT than the FAT and ST. We acknowledge that the KT could perform better in higher frame rate tracking systems where the true target probability density is more accurately approximated. On the other hand, the FAT and ST tackle these challenges by filtering outliers with similar trajectories by utilizing the hyperspectral features. This way, occlusions can be handled more robustly. The computationally efficient ST method cannot handle the scenarios including large density of similarly painted objects as it is a 2-D method and does not integrate kinematic features. This can be seen clearly in the first target case in Table I.

The proposed FAT algorithm with 6-D assignment algorithm clearly outperforms the trackers with a single modality (ST and KT) in any given case as expected. Among four different vehicles, the first target is the one with the highest challenge as it goes through a number of severe occlusions and travels near a high density of similarly painted vehicles. In this case, the proposed method outperforms the other methods reported in [17] and [16] by around 4% and 27% in TgP scores, respectively. The higher TP score for the FAT in [17] means more early track terminations compared to the same TP and TgP scores for the proposed FAT system. This is because of the high density of similarly painted vehicles nearby the first target. By fusing hyperspectral likelihoods together with kinematic likelihoods in a multidimensional assignment framework, we can differentiate the TOI with the aid of better motion evolution approximation and hyperspectral profiles. In the other cases, these three trackers perform persistent tracking. Here, it should be deeply emphasized that the FAT method in [17] requires a firm threshold setting and a different optimum threshold value for each of the four different vehicles to achieve this rates. In other words, it is more sensitive to nonoptimal threshold values as it corresponds to dealing with a greater number of objects using only kinematic likelihoods, increasing the wrong assignment probability. We can expect the proposed FAT method to outperform the other methods by a larger margin in the case of poor threshold settings.

To prove the validity of the proposed approach, we also experiment on more samples and compare our approach to other tracker types in Table II. The proposed FAT method outperforms the others by at least around 15% on average for the given 37 different tracks. As stated before, the FAT proposed in [17] requires vehicle-dependent threshold setting. It is a tedious task

TABLE I
TP AND TGP SCORES FOR TOIS WITH DIFFERENT TRACKER TYPES

	TP (%)					TgP (%)				
Tracker/ID	1st	2nd	3th	4th	Overall	1st	2nd	3th	4th	Overall
2-D (KT)	57.46	41.11	19.33	9.65	31.89	18.65	41.11	15.51	9.22	21.12
3-D (KT)	52.51	46.25	39.36	28.82	41.74	16.94	37.93	31.51	27.50	28.47
4-D (KT)	45.90	33.63	32.49	33.85	36.47	15.16	32.81	31.53	33.40	28.23
5-D (KT)	53.74	30.71	34.12	31.42	37.50	42.53	30.70	34.11	31.00	34.59
6-D (KT)	49.03	77.82	38.05	38.37	50.82	20.85	70.52	36.70	37.45	41.38
ST	20.01	87.48	65.04	50.24	55.69	4.82	65.05	87.48	43.63	50.25
2-D (FAT)	64.05	79.42	91.05	79.54	78.52	55.73	76.19	80.49	73.26	71.42
3-D (FAT)	59.31	70.40	91.96	85.50	76.79	54.41	69.67	85.11	83.97	73.29
4-D (FAT)	65.96	75.23	94.72	91.63	81.89	62.84	74.37	94.40	91.63	80.81
5-D (FAT)	59.37	81.19	97.37	90.15	82.02	54.75	80.35	97.37	89.95	80.61
6-D (FAT)	75.51	84.45	91.81	89.40	85.29	74.48	83.37	91.51	89.40	84.69
2-D (FAT) [16]	53.84	80.83	79.33	86.47	75.12	47.26	80.83	79.22	86.47	73.45
6-D (FAT) [17]	80.14	81.89	94.61	89.63	86.57	70.80	81.09	94.26	89.63	83.95

1st Track - White, 74 Frames, **2nd Track** - Red, 86 Frames, **3th Track** - Blue, 89 Frames, **4th Track** - Red, 76 Frames(Optimum threshold settings are used in 6-D FAT approach in [17].)

TABLE II
AVERAGE TP AND TGP SCORES ON 37 DIFFERENT TARGET SAMPLES

Metric/Tracker	6-D (KT)	ST	2-D FAT [16]	6-D FAT [17]	6-D FAT (Without Section IV-B)	6-D FAT
TP	24.88	39.82	39.23	44.12	32.88	57.63
TgP	24.88	35.85	39.15	43.64	29.38	57.13

(Fixed, global threshold as in the proposed approach is used in 6-D FAT [17] to get the results.)

to manually assign thresholds for 37 different tracks. Instead, we assigned a fixed global threshold value with the one used in our approach for the purpose of comparison. The main reason behind the drop in tracking rates compared to Table I is the large-scale density of slowly moving or stopped vehicles in the scene. Overall accuracy can be improved by implementing a more sophisticated detection module to reduce false negatives and false alarms. This is proved by removing the prediction-based sampling module explained in Section IV-B. In this case, tracking rates drop dramatically to around 32.88% and 29.38% as seen in Table II. Since the main contribution of this study is on smart integration of hyperspectral likelihoods in a multi-dimensional assignment algorithm, we used a computationally simple, nonparametric background subtraction method together with an efficient GSF-based hyperspectral data sampling based blob detection module.

VIII. CONCLUSION

We considered an adaptive, multimodal, performance-driven sensor capable of adaptive hyperspectral data acquisition. This allows the system to utilize hyperspectral features in real-time aerial tracking to increase the persistency of tracking in cases where a single modality struggles. Distinctive hyperspectral profiles of different materials are used to prune outliers in the gate to compensate for the hyperspectral data complexity and simplify the cost minimization problem. In addition, the hyperspectral and kinematic likelihoods of the validated measurements are fused in a Bayesian framework to better discriminate the TOI. The proposed FAT outperforms the other FAT methods utilizing 2-D and multidimensional assignment algorithm. By modeling the density of similarly painted vehicles in a time window, we can better represent dummy and nondummy measurement probabilities especially in the occlusions. In the future, we plan on generating scenarios on a nonstationary platform and remove the background subtraction module to avoid parallax error due to large scale 3-D structures in an urban scene.

REFERENCES

[1] J. Wang and Y. Yagi, "Integrating color and shape-texture features for adaptive real-time object tracking," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 235–240, Feb. 2008.
[2] J. Shin, S. Kim, S. Kang, S.-W. Lee, J. Paik, B. Abidi, and M. Abidi, "Optical flow-based real-time object tracking using non-prior training active feature model," *Real-Time Imag.*, vol. 11, no. 3, pp. 204–218, 2005.
[3] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comput. Vision Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.
[4] V. Reilly, H. Idrees, and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," in *Proc. Eur. Conf. Comput. Vis*, 2010, pp. 186–199.
[5] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 679–684.
[6] M. Keck, L. Galup, and C. Stauffer, "Real-time tracking of low-resolution vehicles for wide-area persistent surveillance," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2013, pp. 441–448.
[7] K. Palaniappan, F. Bunyak, P. Kumar, I. Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. M. Rao, and G. Seetharaman, "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in *Proc. 13th Conf. Inf. Fusion*, 2010, pp. 1–8.
[8] M. S. Arulampalam, N. Gordon, M. Orton, and B. Ristic, "A variable structure multiple model particle filter for GMTI tracking," in *Proc. 5th Int. Conf. Inf. Fusion*, 2002, vol. 2, pp. 927–934.

[9] O. Payne and A. Marrs, "An unscented particle filter for GMTI tracking," in *Proc. IEEE Aerosp. Conf.*, 2004, vol. 3.
[10] T. Kirubarajan and Y. Bar-Shalom, "Tracking evasive move-stop-move targets with a GMTI radar using a VS-IMM estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 1098–1103, Jul. 2003.
[11] A. Sinha, T. Kirubarajan, and Y. Bar-Shalom, "Autonomous ground target tracking by multiple cooperative UAVs," in *Proc. IEEE Aerosp. Conf.*, 2005, pp. 1–9.
[12] R. D. Meyer, K. J. Kearney, Z. Ninkov, C. T. Cotton, P. Hammond, and B. D. Statt, "RITMOS: A micromirror-based multi-object spectrometer," *Proc. SPIE*, vol. 5492, pp. 200–219, 2004.
[13] T. Wang, Z. Zhu, and E. Blasch, "Bio-inspired adaptive hyperspectral imaging for real-time target tracking," *IEEE Sensors J.*, vol. 10, no. 3, pp. 647–654, Mar. 2010.
[14] J. Blackburn, M. Mendenhall, A. Rice, P. Shelnutt, N. Soliman, and J. Vasquez, "Feature aided tracking with hyperspectral imagery," *Proc. SPIE*, vol. 6699, pp. 66990S–66990S, 2007.
[15] A. C. Rice, J. R. Vasquez, J. Kerekes, and M. J. Mendenhall, "Persistent hyperspectral adaptive multi-modal feature-aided tracking," *Proc. SPIE*, vol. 7340, pp. 73340M–73340M, 2009.
[16] B. Uzkent, M. J. Hoffman, and A. Vodacek, "Efficient integration of spectral features for vehicle tracking utilizing an adaptive sensor," *Proc. SPIE*, vol. 9407, pp. 940707–940707, 2015.
[17] B. Uzkent, M. J. Hoffman, and A. Vodacek, "Spectral validation of measurements in a vehicle tracking DDDAS," *Procedia Comput. Sci.*, vol. 51, pp. 2493–2502, 2015.
[18] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2151–2156, Sep. 2011.
[19] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "Sumo (simulation of urban mobility)," in *Proc. 4th Middle East Symp. Simul. Model.*, 2002, pp. 183–187.
[20] A. Rice, J. Vasquez, M. Mendenhall, and J. Kerekes, "Feature-aided tracking via synthetic hyperspectral imagery," in *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution Remote Sensing*, 2009, pp. 1–4.
[21] P. Liang *et al.*, "Vehicle detection in wide area aerial surveillance using temporal context," in *Proc. IEEE 16th Int. Conf. Inf. Fusion*, 2013, pp. 181–188.
[22] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recog.*, 2004, vol. 2, pp. 28–31.
[23] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 1999, vol. 2.
[24] B. Uzkent, M. J. Hoffman, A. Vodacek, J. P. Kerekes, and B. Chen, "Feature matching and adaptive prediction models in an object tracking DDDAS," *Procedia Comput. Sci.*, vol. 18, pp. 1939–1948, 2013.
[25] B. Chen, W. Sun, and A. Vodacek, "Improving image-based characterization of road junctions, widths, and connectivity by leveraging OpenStreetMap vector map," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, 2014, pp. 4958–4961.
[26] B. Uzkent, M. Hoffman, A. Vodacek, and B. Chen, "Feature matching with an adaptive optical sensor in a ground target tracking System," *IEEE Sensors J.*, vol. 15, no. 1, pp. 510–519, Jan. 2015.
[27] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
[28] S. Deb, M. Yeddanapudi, K. Pattipati, and Y. Bar-Shalom, "A generalized SD assignment algorithm for multisensor-multitarget state estimation," *IEEE Trans. Aerosp. Electronic Syst.*, vol. 33, no. 2, pp. 523–538, Apr. 1997.
[29] A. B. Poore, "Multidimensional assignment formulation of data association problems arising from multitarget and multisensor tracking," *Comput. Optim. Appl.*, vol. 3, no. 1, pp. 27–57, 1994.
[30] T. Kirubarajan, Y. Bar-Shalom, K. R. Pattipati, and I. Kadar, "Ground target tracking with variable structure IMM estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 26–46, Jan. 2000.

Burak Uzkent received the B.S. degree in electrical and electronics engineering from Eskisehir Osmangazi University, Eskisehir, Turkey, in 2009, and the M.S. degree in electrical engineering from the University of Bridgeport, Bridgeport, CT, USA, in 2011. He is currently working toward the Ph.D. degree at the Department of Imaging Science, Rochester Institute of Technology, Rochester, NY, USA.

Since 2012, he has been a Graduate Research Assistant in the Digital Imaging and Remote Sensing Laboratory, Imaging Science Department, Rochester Institute of Technology. His research interests include video surveillance, object recognition, detection, and medical image segmentation.

Matthew J. Hoffman received the B.A. degree in mathematics and astrophysics from Williams College, Williamstown, MA, USA, and the M.S. and Ph.D. degrees in applied mathematics and scientific computation from the University of Maryland, College Park, MD, USA.

He was a Postdoctoral Fellow in the Earth and Planetary Science Department, Johns Hopkins University, Baltimore, MD, USA, from 2009 to 2011 and is currently an Assistant Professor in the School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY, USA. His research include various applications of data assimilation, including marine forecasting, cardiac electrophysiology, and target tracking.

Anthony Vodacek (M'08–SM'13) received the B.S. degree in chemistry from the University of Wisconsin-Madison, Madison, WI, USA, in 1981, and the M.S. and Ph.D. degrees in environmental engineering from Cornell University, Ithaca, NY, USA, in 1985 and 1990, respectively.

He has held postdoctoral positions with the Joint Research Centre, Ispra, Italy, and Goddard Space Flight Center, Wallops Flight facility. From 1994 to 1998, he held research positions with the Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA. He is currently a Professor in the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA, and a Member of the Digital Imaging and Remote Sensing Laboratory. His research interests include dynamic remote sensing data driven environmental models and spectral processing methods for images of aquatic and terrestrial environments.