A JOINT APPROACH TO VECTOR ROAD MAP REGISTRATION AND VEHICLE TRACKING FOR WIDE AREA MOTION IMAGERY

Ahmed Elliethy, Gaurav Sharma

Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627
{ahmed.s.elliethy, gaurav.sharma}@rochester.edu

ABSTRACT

Modern aerial imaging platforms provide wide-area motion imagery (WAMI) at high spatial and moderate temporal resolutions making feasible a range of new applications. We consider the dual tasks of registering WAMI frames to geo-referenced vector road-maps and tracking vehicles through the progression of WAMI frames. We present a novel algorithm that performs these tasks jointly and offers improvements in both by exploiting the synergy between the tasks. Tracking for the large number of vehicles seen in urban-area WAMI is improved by auxiliary information that registration to the vector road-map provides by localizing roads within the scene. Similarly, registration of the WAMI frames to the vector map is improved by formulating the registration as a chamfer minimization between the vehicular trajectories and the road network, an approach that resolves challenges for registration posed by the fundamentally different data modalities between the aerial images and the vector road maps. Results obtained over our test datasets show the effectiveness of the proposed joint methodology. For both road network alignment and vehicle tracking, the proposed method offers a very significant improvement over available alternatives: the proposed approach yields better numerical metrics for quantification of registration accuracy and fewer false identification switches for tracked vehicles.

Index Terms— Wide area motion imagery, vehicle tracking, vector road map, geo-registration, large scale visual analytics

1. INTRODUCTION

Wide area motion imagery (WAMI) that offers a high resolution picture sequences covering a “city-scale” area within each frame at temporal rates of 1-2 frames per second has recently become available with the launch of several new aerial imaging platforms [1–3]. Prior work on WAMI has addressed individually the problems of registration of WAMI frames to a vector road-map and of vehicular tracking. Position and camera orientation estimates from the Global Positioning System (GPS) and Inertial Navigation System (INS) devices on WAMI platforms provide approximate registration of the WAMI frames to geo-referenced coordinates. In most applications, however, finer image-based registration is necessary for the road-map information to be useful. Conventional image feature matching based techniques, such as SIFT (Scale-Invariant Feature Transform) [4], and SURF (Speeded Up Robust Features) [5], cannot be directly used between the fundamentally different types of data: the WAMI data consists of image pixel values whereas the vector road map is described as lines/curves connecting a series of points. Considerable research has been done for aligning vector road maps to static aerial imagery, a process that is referred to as *conflation*. The approach in [6–8] is typical of conflation: the vector road representations are aligned with an aerial image by identifying locations of corresponding road intersection points in both representations and estimating a parametric registration transformation for the alignment. Detection of road intersections in the aerial images, which is critical for this process, is plagued by problems in natural scenes due to view point changes over a large range, shadows and occlusions caused by buildings and trees adjoining the roads and other variations in imaging conditions. To aid in the process, a recent approach [6] proposes the use of hyper-spectral aerial imagery, where spectral properties and contextual analysis can aid in detection of road intersections. A Bayes approach for road segmentation is proposed in [7] which is then followed by localized template matching to detect the road intersections. While the results from the approach are promising, it requires a large number of manually labeled training examples for each data set. In [8], corner detection is used to detect road intersections, which is often unreliable, specially in high resolution aerial images, for which simple corner detection fails because the roads are quite wide.

The utility of WAMI imagery in vehicular tracking has also been recognized and a number of efforts exist in this area. The majority of these, adopt a tracking by detection framework wherein given vehicle detections in each frame, the goal is to associate detections corresponding to the same vehicle over the entire set of aerial image frames. In [9], detections in successive frames are associated based on context similarity measure. Another context similarity approach is proposed in [10] to handle more complex cases. In [11], joint probabilistic graph matching is used to associate detections for successive frames. Due to limited vehicle appearance, these methods are prone to ID switches, as some characteristics (speed, direction, etc) for tracked vehicles are inferred from the first few frames which can result in ambiguous or incorrect associations in subsequent frames. Based on observations of these problems, other efforts approach the assignment problem globally over the entire set of video frames [12–15], and provide efficient approaches to obtain a solution in reasonable time. These approaches do not scale to WAMI because of the extremely large number of vehicles to be tracked in urban WAMI scenes. Additionally, the methods are often designed for full motion video at 30 frames per second and do not directly translate to the low temporal resolutions typical of WAMI.

Prior work has also recognized the potential benefit that can be obtained for vehicular tracking in aerial imagery upon registration to a geo-referenced road map. For example, in [11] a co-registered road network, assumed available *a priori*, is used to regularize the matching of the vehicle detections to the previous existing vehicular tracks. In [16], in a variant of the previously described approaches for registration, SIFT is used to detect correspondences between the ground features from a small footprint aerial video frame and an auxiliary geo-referenced image to geo-register the video frames. This geo-registration helps to estimate the camera pose and depth map for each frame, and this depth map is used to segment the scene into building, foliage, and roads using a multi-cue segmentation framework.

The segmentation is used to help improve tracking by reducing false vehicle detections. The approach is, however, computationally demanding.

In contrast with the prior methods that address registration and tracking for WAMI independently or individually, in this paper, we propose a joint approach for both registration and tracking. Our approach simultaneously estimates, via an alternating minimization, vehicular trajectories over a multi-frame (typically 10-15 frames) temporal window and the geometric transformation for best aligning these trajectories with the road network. The approach is developed based on a maximum *a posteriori* probability (MAP) formulation for the problem that penalizes trajectory deviations from the road network using: (1) a novel chamfer distance [17] metric for the registration accuracy that we introduced in [18] and appropriately modify for our new problem setting and (2) a successive approach to identifying and extending reliable trajectories for individual vehicles based on detections in individual frames and the alignment of the oriented trajectories with the road network and directionality. Results obtained over a test WAMI dataset highlight the effectiveness of the proposed method: compared with the alternatives we obtain better estimates of both the vehicle trajectories and the registration between the WAMI frames and the vector road map.

This paper is organized as follows. Section 2 explains our novel joint formulation for vehicle trajectory estimation and road network alignment, and the proposed solution. Results and a comparison against alternative methods are presented in Section 3. We conclude the paper in Section 4.

2. JOINT VEHICLE TRACKING AND ROAD NETWORK ALIGNMENT

2.1. Problem Formulation

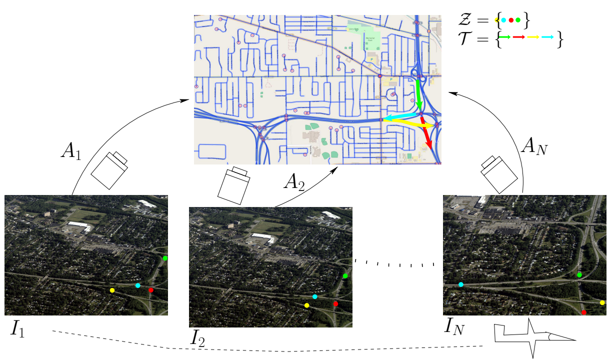

Fig. 1: Joint vehicle tracking and road network alignment. Detections and corresponding trajectories are shown in the same color on video frames and on the map. Given vehicle detections for each frame, our goal is to link these detections into trajectories and estimate the geometric transformations $A_i, i = 1, \ldots, N$

Fig. 1 illustrates our problem setting and is helpful for understanding our formal problem formulation that follows. We assume that we have available a vector map R_g defined as an orthographic projection [19] using corresponding 2D orthogonal geo-referenced coordinates (χ, ζ). The map identifies the network of roads in the geographic area, the k^{th} road r^k being represented as a sequence of spatial locations $({}^r\chi_i^k, {}^r\zeta_i^k)$ along the road and its direction represented as a sequence of orientation angles ${}^r\theta_i^k \in [-\pi, \pi]$, where $i = 1, 2, \ldots N_k^r$ for the k^{th} road. For a set of N time instants $t_1 < t_2 < \cdots < t_N$, a corresponding series of N WAMI frames $\mathcal{I} = \left(I_1(x^1, y^1), I_2(x^2, y^2), \ldots, I_N(x^N, y^N)\right)$ are captured by the moving aerial platform where (x^i, y^i) are the pixel locations along the native orthogonal coordinates for the image sensor when capturing the i^{th} image. Under a planar assumption for the imaged region, a 3×3 homography matrix $\mathbf{A}_i$ relates the image coordinates (x^i, y^i) for the i^{th} frame to the orthographic geo-referenced 2D coordinates (χ, ζ) via the homogeneous transformation relation $[\chi, \zeta, \omega]^T = \mathbf{A}_i[x, y, 1]^T$, where ω is a scaling factor [20]. For the (arbitrary number of) moving vehicles captured in the WAMI frames we define *trajectories*, where for the l^{th} vehicle, we define the trajectory as the sequence of N spatial locations $T_l = \left(({}^v\chi_1^l, {}^v\zeta_1^l), ({}^v\chi_2^l, {}^v\zeta_2^l), \ldots ({}^v\chi_N^l, {}^v\zeta_N^l)\right)$ at the time instants $t_1, t_2, \ldots, t_N$, respectively (in the geo-referenced coordinate system of R_g). We are interested in estimating the transformations $\mathcal{A} = \{\mathbf{A}_i\}_{i=1}^N$ that register the captured WAMI frames to the geo-referenced map R_g and in tracking the moving vehicles captured in the WAMI frames by estimating their trajectories $\mathcal{T} = \{T_l\}$. We consider a maximum a posteriori probability (MAP) formulation for the estimation, where the optimal estimates of the registration and the trajectories are obtained as[1]

$$\{\hat{\mathcal{T}}, \hat{\mathcal{A}}\} = \arg\max_{\mathcal{T}, \mathcal{A}} P(\mathcal{T}, \mathcal{A}|\mathcal{I}). \tag{1}$$

Our joint formulation benefits both trajectory and alignment estimation sub-problems. As we capture vehicles locations on each frame coordinate system, mapping these vehicle locations into a common reference coordinate system (R_g) and estimating trajectories in that coordinate system, allows us to leverage the rich geo-spatial information provided by R_g, which helps to estimate accurate trajectories. However, estimating an accurate registration between the WAMI frames and R_g is challenging task due to the difference between their data representations. The estimated trajectories help solve this difficult registration problem because both estimated trajectories and R_g have the same representation (binary). Thus, both trajectory, and alignment estimation sub-problems benefit from each other in a joint solution approach. Instead of solving the estimation problem in (1) directly, we split the duration of the WAMI frames into a series of temporal windows and solve the problem for each temporal window and propagate estimates between the windows. Thus we assume that the N frames previously defined represent a single temporal window used for processing. For our temporal window, the transformations $\mathcal{A} = \{\mathbf{A}_i\}_{i=1}^N$ can equivalently be represented by the transformation $\mathbf{A}_1$ and the homography matrices $\left\{\mathbf{H}_i^{i+1}\right\}_{i=1}^{N-1}$ that relate successive frames[2], where $\mathbf{H}_i^j$ transforms the image coordinates (x^j, y^j) for the j^{th} frame to the image coordinates (x^i, y^i) for the i^{th} frame. Furthermore, using the co-registered frames in the temporal window, a background model is readily obtained for the entire window (for example by using the median filter [9]) that in turn allows ready detection of the vehicle locations. Specifically, in the i^{th} WAMI frame we denote the detected vehicle locations as a sequence $\mathbf{z}_k^i = ({}^vx_k^i, {}^vy_k^i)$, $k = 1, 2, \ldots$ of points in the frame's native pixel coordinates. To proceed further, we adopt a tracking by detection framework, where tracking operates on the vehicle locations detected in each WAMI frame (using a vehicle detector) and approximates the estimation in (1) by

$$\{\hat{\mathcal{T}}, \hat{\mathbf{A}}_1\} = \arg\max_{\mathcal{T}, \mathbf{A}_1} P(\mathcal{T}, \mathbf{A}_1|\mathcal{Z}), \tag{2}$$

[1]Throughout the description, we assume that the vector map R_g is given. To simplify notation we do not indicate this conditioning.

[2]Short windows ensure that accumulated error is negligible.

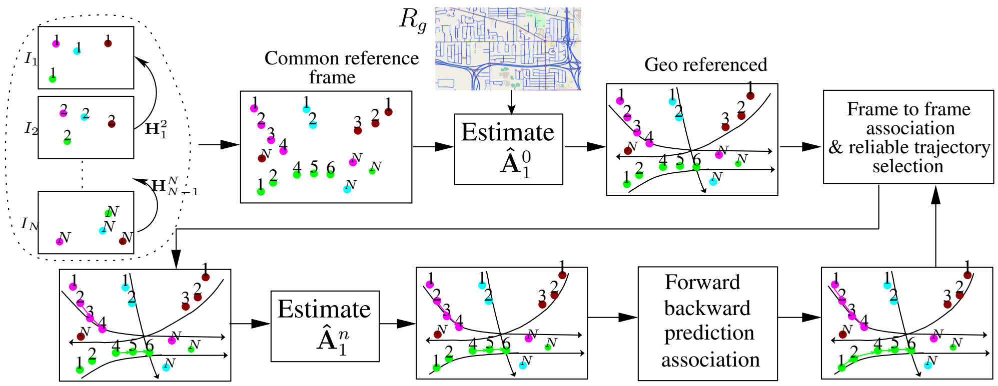

Fig. 2: Block diagram for one iteration of the proposed joint vehicle tracking and road network alignment algorithm.

where $\mathcal{Z} = \left\{\mathbf{z}_k^i\right\}_{i,k}$ is the complete set of vehicular detections. This approximation becomes exact under the assumption that the inter-frame registrations are a function of the image data and the complete set of vehicular detections $\mathcal{Z}$ constitute sufficient statistics [21]. By applying Bayes' rule, (2) becomes

$$\{\hat{\mathcal{T}}, \hat{\mathbf{A}}_1\} = \arg\max_{\mathcal{T}, \mathbf{A}_1} P(\mathcal{Z}|\mathcal{T}, \mathbf{A}_1)P(\mathcal{T}, \mathbf{A}_1), \tag{3}$$

where $P(\mathcal{T}, \mathbf{A}_1)$ is the prior joint distribution, and $P(\mathcal{Z}|\mathcal{T}, \mathbf{A}_1)$ is the likelihood distribution. Page-length restrictions limit us to conveying only the intuition of our method here, readers are referred to an upcoming journal submission [22] for details (also for publicly availability of datasets).

We assume that vehicle detections are conditionally independent given $\mathcal{T}$ and $\mathbf{A}_1$ and model the likelihood as a Bernoulli distribution as in [15]. The trajectories $\mathcal{T}$ (in the geo-referenced coordinate system of R_g) do not depend on the transformation $\mathbf{A}_1$, and therefore the prior distribution factors as $P(\mathcal{T}, \mathbf{A}_1) = P(\mathbf{A}_1)P(\mathcal{T})$. We model $P(\mathcal{T}) = P_{motion}(\mathcal{T})P_{road}(\mathcal{T})$, where $P_{motion}(\mathcal{T})$ measures the motion consistency of trajectories in $\mathcal{T}$ in terms of velocity variation, and the $P_{road}(\mathcal{T})$ measures how well trajectories are matched with roads in R_g by penalizing deviations of each trajectory from the road network in terms of both location and orientation mismatch, quantified by an oriented chamfer distance [17]. Thus, to maximize (3), we search for trajectories $\hat{\mathcal{T}}$ composed from vehicle detections that have small velocity variation over our temporal analysis window and agree with the roads in R_g in terms of location and orientation, i.e., have a low oriented chamfer distance [23] with roads in R_g.

2.2. Iterative optimization algorithm

One iteration of our proposed iterative algorithm is shown in Fig. 2. Our algorithm's goal is to estimate the transformation $\hat{\mathbf{A}}_1$ that maps vehicular detections form the WAMI frames' native coordinate system to the road network R_g coordinate system, and estimate the trajectories $\hat{\mathcal{T}}$ by linking these mapped detections together over the N WAMI frames, where $\hat{\mathbf{A}}_1$ and $\hat{\mathcal{T}}$ maximize (3). We propose an iterative solution by alternating the optimization with respect to $\hat{\mathcal{T}}$ and $\hat{\mathbf{A}}_1$, viz.,

$$\hat{\mathbf{A}}_1^{n+1} = \arg\max_{\mathbf{A}_s} P(\hat{\mathcal{T}}^n, \mathbf{A}_s|\mathcal{Z}), \tag{4}$$

$$\hat{\mathcal{T}}^{n+1} = \arg\max_{\mathcal{T}_s} P(\mathcal{T}_s, \hat{\mathbf{A}}_1^{n+1}|\mathcal{Z}), \tag{5}$$

where $\hat{\mathcal{T}}^n$, and $\hat{\mathbf{A}}_1^n$ are the estimated trajectories and registration transformation in iteration n, respectively.

As shown in Fig. 2, our algorithm uses all detections within the N frame temporal window. These detections are initially mapped to a common reference frame I_1 as described previously, then mapped to the coordinate system of R_g using the current estimate $\hat{\mathbf{A}}_1$. With the help of the road network information available in R_g, these geo-referenced mapped detections are associated on a frame-to-frame basis to estimate the initial trajectories. From these initial trajectories, we select the reliable ones (defined as having small velocity variations and good alignment with the road network) and then we progressively enlarge these reliable trajectories by iteratively linking them together and with the unassigned detections using the available road network information. With these enlarged reliable trajectories in hand, we estimate a more accurate transformation $\hat{\mathbf{A}}_1$ that best aligns these trajectories with the road network in R_g. The new estimate $\hat{\mathbf{A}}_1$ in turn helps us to recover more reliable trajectories by repeating the mentioned iterative process until no more detections are assigned to the existing trajectories.

3. EXPERIMENTAL RESULTS

(a) I_5 (b) I_{10}

Fig. 3: Road network alignment results from different methods for a sample frame. The results for our method are shown in green, for MBA in blue, SBA in purple (only in (a)), and VBA in red (only in (b)).

We evaluated our algorithm on a WAMI data set recorded using the CorvusEye 1500 Wide-Area Airborne System [24] for the Rochester, NY region. For the vector road map, we use OpenStreetMap (OSM) [25]. OSM provides each road in a road network in a vector format along with properties of each road such as its type (highway, residential, etc) and number of lanes.

To test the effectiveness of our joint methodology, we create a test sequence corresponding to our analysis temporal window of $N = 10$ frames (which maintains relevant temporal context with moderate computational burden) by cropping a region (1000×1000 pixels) that contains a forked road network with different directions and also a lot of occluders (bridges, trees, etc.), captured from oblique angle as shown in Fig. 3. We use this sequence to evaluate both the registration and tracking accuracies.

Our results are in two parts. First, we compare our method in terms of registration accuracy with three alternative methods. Next, we compare our tracking method with the method in [9]. To evaluate our registration accuracy, we compare with "Meta-data Based Alignment (MBA)" and "SIFT matching with auxiliary geo-referenced image (SBA)", and "Vehicular detections based alignment (VBA) in [18]". The MBA method obtains the alignment by using the meta-data from the NITF 2.1 format [26] files used for the WAMI frames. The SBA method tries to match SIFT features between the aerial image and an auxiliary ortho-rectified geo-referenced image taken from Google Maps. VBA registers the road network with the WAMI frame by minimizing the chamfer distance between identified vehicular detection locations and the network of road lines identified in the vector road map [18].

For a single frame, the aligned road networks obtained with proposed method and with the alternative SBA, MBA, and VBA methods are shown in Fig. 3. The proposed method offers a significant enhancement over MBA which depends only on the meta-data to get an aligned road network and over SBA which uses SIFT and auxiliary geo-referenced Google map image, and shows improvement over the VBA method. The MBA method has significant errors because of the inaccuracy of the meta-data parameters due to the limited accuracy of on-board navigation devices. The SBA method does not improve significantly because of spurious correspondences found by the SIFT matching between the aerial image and the Google map image which have significant differences due to severe view point change, different illumination, and different capturing times. Our method provides a more accurate result compared to the VBA method by incorporating both trajectory locations and directions in the road network alignment, whereas the VBA method estimates the alignment from detection locations alone [18].

To provide quantitative comparison between the different registration methods, we manually label the ground truth (GT) road network for our test sequence and calculate the chamfer distance between the ground truth road network and the road network generated from our method, the VBA, the SBA method, and the MBA method. The results summarized in Table 1 reinforce the conclusions seen visually. The proposed method has a much lower value for the chamfer distance highlighting the fact that the proposed method offers a significant improvement over both the MBA and SBA methods, and shows a moderate improvement over our recent VBA method.

	MBA	SBA	VBA	Proposed
GT chamfer distance	33.4	5.9	0.9	**0.56**
Improvement	98.3%	90.5%	37.7%	-

Table 1: Chamfer distance between the ground truth road network and the road network generated using the MBA method, the SBA method, the VBA method, and our proposed method.

To evaluate our tracking methodology, we manually label each vehicle within our test sequence and quantify tracking performance by the number of *ID-switches* for each vehicle compared to its ground truth label, and by Multiple Object Tracking Accuracy (*MOTA*) defined in [27] (with $c_s = 1$). We compare our tracking methodology with the "Frame-to Frame based Association method (F2FA)" [9] that use the Hungarian algorithm to associate vehicle detections with the estimated trajectories from frame to frame (FTF) using a cost metric that penalizes the velocity, position, and spatial context mismatch constrained by an estimated road direction. Moreover, we drop the road direction estimation step, and modify the method to exploit our accurately aligned road network resulting in a modified method "Frame-to Frame based Road constrained Association method (F2FRA)".

	F2FA method [9]	F2FRA method	Proposed method
ID-switches	3.2	2.1	**0.6**
MOTA	0.55	0.7	**0.91**

Table 2: Comparison of the proposed method with the F2FA, and F2FRA methods in terms of average number of *ID-switches* and *MOTA*. The result show that our tracking methodology provide significant enhancement for the average *ID-switches* and the *MOTA* compared to the other methods.

In Table 2, we show a comparison in terms of *ID-switches* and *MOTA* for our proposed method with the F2FA method and the F2FRA method. The average number of *ID-switches* of our proposed method is much less than the F2FA method. The F2FA method is prone to *ID-switches*, because it associates vehicle detections from FTF, a process that causes errors to propagate to following frames. Introducing our aligned road network for defining the cost function for associating detections to trajectories, reduces the *ID-switches* for the F2FA method as seen in the tabulated results for F2FRA. The proposed method improves further upon this, providing the best result among the methods evaluated. The results highlight the contribution of our method, which solves the multi-vehicle tracking problem globally over the entire temporal window. The approach employed in our method of iteratively optimizing by relying on reliable trajectories introduces a mechanism that can recover from the assignment errors resulting from incorrect FTF associations.

Our implementation is in C++ using OpenCV [28] libraries and the current unoptimized implementation requires about 0.8 seconds per frame for tracking, after FTF registration and background estimation are complete.

4. CONCLUSIONS

The joint formulation we propose for registering WAMI frames to a vector road map and for tracking vehicles within WAMI frames offers a significant improvement over prior alternative approaches that tackle these problems individually. For our ground-truth labeled WAMI dataset, compared with the approximate registration parameters obtained from WAMI metadata, our method reduced the chamfer error metric for registration inaccuracy by 98% and compared with our own recent related but non-joint approach [18] the chamfer error metric is reduced by 37%. Tracking performance is also improved by our proposed approach: tracks obtained are more consistent with the road network and with themselves, and the average number of identity switches seen are reduced by 81% compared with a recent state of the art tracking method. Similar results are obtained for a much large corpus of imagery from different sources [22]. These results demonstrate that our framework effectively utilizes the synergy between the problems of geo-registration and tracking for WAMI.

5. ACKNOWLEDGMENT

We thank Bernard Brower of Harris Corporation for making available the CorvusEye [24] WAMI datasets used in this research.

6. REFERENCES

[1] K. Palaniappan, R. M. Rao, and G. Seetharaman, "Wide-area persistent airborne video: Architecture and challenges," in *Distributed Video Sensor Networks*. Springer, 2011, pp. 349–371.

[2] E. Blasch, G. Seetharaman, S. Suddarth, K. Palaniappan, G. Chen, H. Ling, and A. Basharat, "Summary of methods in wide-area motion imagery (WAMI)," in *Proc. SPIE*, vol. 9089, 2014, pp. 90 890C–90 890C–10.

[3] R. Porter, A. Fraser, and D. Hush, "Wide-area motion imagery," *IEEE Sig. Proc. Mag.*, vol. 27, no. 5, pp. 56–65, Sept 2010.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comp. Vis. and Image Understanding.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[6] W. Song, J. Keller, T. Haithcoat, and C. Davis, "Automated geospatial conflation of vector road maps to high resolution imagery," *IEEE Trans. Image Proc.*, vol. 18, no. 2, pp. 388–400, Feb 2009.

[7] C.-C. Chen, C. A. Knoblock, and C. Shahabi, "Automatically conflating road vector data with orthoimagery," *GeoInformatica*, vol. 10, no. 4, pp. 495–530, 2006.

[8] C.-C. Chen, C. A. Knoblock, C. Shahabi, Y.-Y. Chiang, and S. Thakkar, "Automatically and accurately conflating orthoimagery and street maps," in *Proc. ACM Int. Workshop on Geographic Information Systems*. ACM, 2004, pp. 47–56.

[9] V. Reilly, H. Idrees, and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," in *Proc. European Conf. Computer Vision*, 2010, vol. 6313, pp. 186–199.

[10] X. Shi, P. Li, H. Ling, W. Hu, and E. Blasch, "Using maximum consistency context for multiple target association in wide area traffic scenes," in *IEEE Intl. Conf. Acoust., Speech, and Signal Proc.*, May 2013, pp. 2188–2192.

[11] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2010, pp. 679–684.

[12] A. Dehghan, S. M. Assari, and M. Shah, "GMMCP tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2015, pp. 4091–4099.

[13] A. R. Zamir, A. Dehghan, and M. Shah, "GMCP-tracker: Global multi-object tracking using generalized minimum clique graphs," in *Proc. European Conf. Computer Vision*. Springer, 2012, pp. 343–356.

[14] A. Andriyenko, K. Schindler, and S. Roth, "Discrete-continuous optimization for multi-target tracking," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2012, pp. 1926–1933.

[15] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2008, pp. 1–8.

[16] J. Xiao, H. Cheng, F. Han, and H. Sawhney, "Geo-spatial aerial video processing for scene understanding and object tracking," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2008, pp. 1–8.

[17] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. Int. Joint Conf. Artificial Intell.*, 1977, pp. 659–663.

[18] A. Elliethy and G. Sharma, "Vector road map registration to oblique wide area motion imagery by exploiting vehicles movements," in *IS&T Electronic Imaging: Video Surveillance and Transportation Imaging Applications*. IS&T, Springfield, VA, 2016, San Francisco, CA, 14-18 February 2016, to appear.

[19] J. P. Snyder, *Map projections–A working manual*. US Government Printing Office, 1987, vol. 1395.

[20] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: John Wiley and Sons, 2006.

[22] A. Elliethy and G. Sharma, "Joint vector road map registration and vehicle tracking for wide area motion imagery," in preparation for submission to IEEE Trans. Image Proc.

[23] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2010, pp. 1696–1703.

[24] "CorvusEye™ 1500," http://www.exelisinc.com/solutions/corvuseye1500.

[25] "OpenStreetMap," http://www.openstreetmap.org.

[26] National imagery transmission format (NITF) version 2.1. [Online]. Available: http://www.gwg.nga.mil/ntb/baseline/index.html

[27] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 31, no. 2, pp. 319–336, Feb 2009.

[28] OpenCV (Open Source Computer Vision) library. [Online]. Available: http://opencv.org/