RESEARCH ARTICLE OPEN ACCESS

Multiple Moving Object Detection from UAV Videos Using Trajectories of Matched Regional Adjacency Graphs

Dr K Sailaja MCA, M.Tech , M.Phil , Ph.D [1], P Hari Prasad [2]

[1] Professor & HOD, Department of Computer Applications
[2] Student, Department of Computer Applications
[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

Image registration has been long used as a basis for the detection of moving objects. Registration techniques attempt to discover correspondences between consecutive frame pairs based on image appearances under rigid and affine transformations. However, spatial information is often ignored, and different motions from multiple moving objects cannot be efficiently modeled. Moreover, image registration is not well suited to handle occlusion that can result in potential object misses. This paper proposes a novel approach to address these problems. First, segmented video frames from unmanned aerial vehicle captured video sequences are represented using region adjacency graphs of visual appearance and geometric properties. Correspondence matching (for visible and occluded regions) is then performed between graph sequences by using multigraph matching. After matching, region labeling is achieved by a proposed graph col- oring algorithm which assigns a background or foreground label to the respective region. The intuition of the algorithm is that background scene and foreground moving objects exhibit differ- ent motion characteristics in a sequence, and hence, their spatial distances are expected to be varying with time. Experiments conducted on several DARPA VIVID video sequences as well as self-captured videos show that the proposed method is robust to unknown transformations, with significant improvements in overall precision and recall compared to existing works.

Keywords— Object detection, Moving objects, Video sequences, Trajectories, Adjacency Graphs.

I. INTRODUCTION

Object detection has become an integral part of video surveillance systems. It serves as the fundamental enabler for important tasks such as moving object detection and tracking [1]–[4], motion segmentation [5], [6], object classification [7], event detection [8], and behavioral analy- sis [9]. In this paper, we address the problem of multiple moving object detection from video sequences captured by mounted surveillance cameras on airborne vehicles such as unmanned aerial vehicles (UAVs). In such a setting, moving object detection becomes challenging as the camera motion is independent of the moving objects' motions. Typically, UAVs fly at low altitudes, render high mobility, fast deployment, and large surveillance scope [10]. Furthermore, there is a need to cope with the undesirable yet common characteristics of UAV-captured videos such as multiple moving objects, large/small displacements of fast/slow moving objects, object occlusion (either by terrain or other objects), and objects leaving/re-entering the field of view.

Several approaches have been proposed in the past for mul- tiple objects detection from UAV videos. One popular strategy is to align each frame to its temporally adjacent frame to eliminate the effect of the camera motion. This can be achieved by using image stabilization and registration methods, where two images of the same scene taken at different times are geometrically overlaid. Image registration is the seemingly popular trend for remote sensing applications, which involves the discovery (matching) of feature correspondences between geometrically aligned image pairs [11]–[13]. In general, fea- ture detection and matching are the two fundamental steps in the majority of registration approaches where the bags- of-features representation is commonly adopted [14], [15].

However, such representations ignore spatial feature layouts and pixels value correlations due to the order-less sets of the local descriptors. This causes potential problems during the matching phase especially when one-to-one correspondences between feature points are not presented between image pairs. Other issues include illumination variations between the images as well as noise due to the poor video quality. In addition, most approaches assume that the transformations aligning the points are parametric (e.g., rigid and affine), which is not true in many real world situations, especially with a moving camera setup. Moreover, since registration techniques only process two frames at a time, it might be difficult to cater for object occlusion. This is because an object on a trajectory might suddenly disappear from the field of view. The same object, however, may re-enter the scene in a relatively distant frame. Therefore, if only two frames are considered each time, occlusion handling might not be possible.

In this paper, we propose a moving object detec- tion framework without explicitly overlaying frame pairs. Instead, correspondence matches are discovered by considering a group of frames at a time. This is followed by a labeling step that assigns regions as either belonging to the background or foreground (moving objects). Specifically, each frame is segmented into regions and subsequently repre- sented as a regional adjacency graph (RAG). Correspondence matching on a group of consecutive frames is performed by using multigraph matching where one-to-one correspondences are discovered through appearance similarity and geometrical constraints. Once correspondences are identified, a proposed graph coloring algorithm finally labels the regions as either being background or foreground objects.

II. RELATEDWORKS

In the literature, most moving object detection works use video footage from fixed cameras. This enables background stabilization and subtraction techniques to be used as the back- ground is relatively the same throughout the frame sequences. Once background pixels have been identified, they can be removed allowing foreground objects to be detected. In this section, two general background subtraction categories are discussed. They are techniques based on background modeling and those based on image registration.

Background modeling has long been applied in moving object detection where foreground objects are detected based on a reference (i.e., background) model/image. One idea is to calculate the difference between each frame sequence against the generated model where a thresholding procedure finally determines the results [16]. Temporal differencing is another alternative that takes differences between two or three suc- cessive frames to model background pixels [12]. Background- based approaches are indeed flexible and fast. Nevertheless, they only work well in a fixed camera environment where the background is expectedly constant. In a moving camera setup, however, camera motion and scene transitions exist making such background models unsuitable. Moreover, aside from the unstable backgrounds, the presence of multiple moving objects at varying speeds, slow/rapid illumination changes and/or noise from poor quality videos will also cause object detection to be problematic [11].

Image registration approaches on the other hand, dis- cover correspondences between image pairs (i.e., reference and sensed images) where a geometrical alignment is ulti- mately performed. The images involved are taken from the same scene but at different times [13]. According to Zitová and Flusser [14], the first two steps in image registration are feature detection followed by feature matching. The former involves manually or automatically detecting prominent fea- tures in both the reference and sensed images. The matching step then establishes a correspondence between features in both images using some similarity measure. Two types of image registration approaches are discussed in this section, namely, (A) area- and feature-based methods and (B) graph representation-based matching.

III. PROPOSED SYSTEM ARCHITECTURE

In this paper, we propose a novel framework for moving object detection that mainly consists of two main phases, namely: 1) correspondence matching (including occlusion han- dling) and 2) background and foreground labeling. The overall diagram of the proposed technique is presented in Fig. 1. In this paper, both appearance similarity and geometrical constraints are imposed on region-based features. If images are seen as a set of connected regions, they can hence be represented by RAGs. Representing images as graphs of regions allows the spatial relationships between pixels to also be incorporated at a higher level, making the model more robust toward local variations such as scaling, translation, rotation, illumination, and intensity changes. In addition, both unary node-to-node and pairwise edge-to-edge relationships can be integrated into the model using graph representation. Therefore, better correspondence matching can be expected. It is also worth noting that UAV-captured videos contain multiple rapidly moving objects and through time, these objects can be occluded either by terrain or other objects. An example is when an object is absent in one frame (due to being blocked by vegetation) but then re-enters the scene in a future frame. This is illustrated in Fig. 2 where an occluded vehicle in frame 680 becomes visible in frame 710. Arguably, occluded objects can only be detected by analyzing long-term trajectories. Therefore, since a sequence of frames (instead of just frame pairs) is considered at a time, the proposed framework caters to occlusion handling. Furthermore, by imposing structural and geometrical constraints on a frame sequence, which are in turn represented as a sequence of graphs, the model can be more robust toward deformations, missing or incomplete data, and outlier regions.

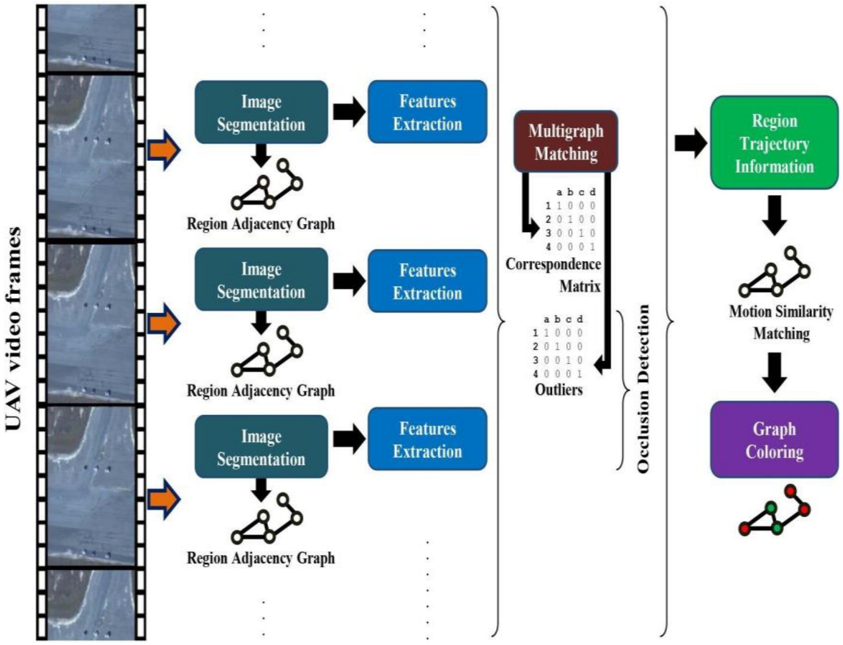

Fig.1 Proposed framework

Initially, all frames go through a segmentation process. In the desired result, a segmented region is only a part of one distinct image object. Certain segmentation approaches might yield under-segmentation, which is problematic since each object might be accidentally merged with the other objects. Resultantly, we decided to go with an over segmentation algorithm. In this paper, SLIC superpixel [34] was chosen as it is able to produce small yet uniform regions. Although large numbers of over segmented regions are generated, at least potential image objects contain many regions. But since our approach processes many frames at a time, the large number of regions can increase computational complexity, specifically during the matching phase. To solve the problem, we propose to combine homogenous regions through a merging process. Specifically, the MPEG-7 dominant color descriptor (DCD) [35] is exploited to measure the homogeneity of adjacent regions. DCD includes dominant colors information and their percentages, providing an effective, compact, and intuitive description of colors within an image region.

In the literature, establishing correspondences between two groups of points is known as point pattern matching [13]. Its objective is to remove outliers in order to estimate the transformations from inliers (inliers being points having cor- respondences in the next frame). However, this process is complicated in nonparametric and nonrigid models where images are distorted by different types of transformations [37]. In the proposed paper, we avoid estimating the transforma- tions. Instead, a set of consecutive frames is considered at a time and their graphical representation (RAG) is exploited. Specifically, correspondence discovery is treated as multigraph matching between RAGs within a set of consecutive frames. Intuitively, processing multiple frames for matching makes more sense as visually consistent regions are expectedly better discovered over a longer frame sequence (as compared to frame pairs). For graph matching to be performed, each RAG node is assigned a joint feature set consisting of the DCD, texture, and shape. DCD is the same as in the previous section. Gabor features [38], [39] are used to represent texture descriptors. For shape, the descriptor includes the first 32 (4 8) coefficients from the Elliptic Fourier Descriptor [40], seven Hu invariant moments, and three Tamura features. This combination contains detailed shape information and is also invariant toward rotation, translation, and scaling [41].

Overall, the node features effectively represent the visual attributes of the region. In the previous step, a set of region correspondences were discovered. These correspondences contain useful motion information that can be used for moving object detection. Consequently, a motion similarity graph (MSG) is constructed from the correspondences of a sequence of RAGs. Note that in each RAG, the edges connecting adjacent nodes represent weights calculated as the Euclidean distance between inter- connected nodes. If a specific edge connecting two nodes does not change more than a predefined threshold over an RAG sequence, then the resultant edge in the constructed MSG is labeled as “similar.” Otherwise, it will be labeled as “dissimilar.” This is depicted in Fig. 1 where an MSG is constructed from six consecutive graphs. A specific edge with the variable weights in the trajectory indicates that regions cor- responding to its end-nodes move independently with different motions. In other words, if the weight of edge *eij* in the first frame is different than the weights of its corresponding edges in the following frames, its connecting regions must belong to different objects due to their motion difference. Nevertheless, two objects moving next to each other share the similar motion though they are distinct objects. Obviously, this situation may not continue in the successive video frames. As soon as one of these objects moves in a different direction, it will be apparent that they are indeed different objects. Hence, the common motion is not explicitly estimated but only motion difference is observed in the trajectories. In this paper, the edge variability is investigated in a trajectory with a specific length and it is averaged over all frames in this trajectory.

Though motion similarity between image regions can be induced from the MSG, they cannot simply be labeled as either background or foreground. For instance, two neighbor- ing regions with a similar motion can be constituent parts of one moving object. Alternatively, they can also be two background regions captured by the moving camera. Such discrepancies can be treated as a graph partitioning problem where image regions corresponding to the MSG nodes are assigned to different components. Consequently, these nodes are partitioned into k +1 components, by assuming k-number of moving objects and one background region. Note, however, that the actual label (background or foreground) is not yet assigned to each node.

We propose a graph coloring algorithm to achieve the graph partitioning. The objective is to automatically assign colors to the nodes of the graph such that connected nodes belonging to different labels take differing colors by utilizing the minimum number of colors. Note that this task cannot be performed by conventional graph coloring algorithms since no adjacent node can share the same color. Although there are many ways to find a coloring solution, the number of possible colorings is unique [51]. However, the intrinsic characteristics of the problem can be analyzed to determine the final coloring scheme. Since the only relationship between nodes is motion similarity (or dissimilarity) and the size of the generated MSG is relatively small, the color range can be limited. Another useful attribute is distribution of the back- ground regions, which constrain the possible colorings for their neighbors.

The proposed graph coloring algorithm initially colors the graph with only one color (the starting color). In rare cases where there are no moving objects in the scene, the algorithm terminates with just this one color. Otherwise, the number of available colors increases until all regions are assigned their respective colors. Initially, the algorithm selects a region *A* with the most number of edges and assigns a random color to it. The next region *B* is preferably selected from the uncolored neighboring regions. If the connecting edge between *A* and *B* is labeled as "similar," *B* takes the same color as *A*. Otherwise, it will be assigned a new/different color. The algorithm will then examine all the other regions and for every region, the same color assignment rule is applied. An illustra- tion of this is given in Fig. 1 where noticeably, the minimum number of required colors will be equal to the number of objects. We can select the color that is distributed over the graph as the background color. This is because background regions are spread across the scene in UAV-captured videos. However, if background regions are separated by moving objects, the smaller background regions are considered as moving objects. To overcome this phenomenon, the nodes with the same colors are grouped together. The motion difference is then considered in these new larger regions. Notably, the motions under any transformations (rigid or nonrigid) are taken into account. In existing feature-based point matching, one main chal- lenge is object shape estimation [52]. This task is important since it defines the bounding boxes around detected moving objects. The proposed method preserves the object boundary for oversegmented region (in the earliest step). Therefore, inte-grating the connected regions with the same motions can reveal moving objects. In other words, the overall bounding box for each moving object is the combination of its constituent regions' bounding boxes.

IV. RESULTS AND DISCUSSION

The proposed moving object detection framework consists of different components. For the segmentation phase, the SLIC superpixel implementation is used. It is obtained from https://github.com/PSMM/SLIC-Superpixels. The MATLAB source code for the pairwise matching procedure is based on [46] and the code can be obtained from http://www.f- zhou.com/gm_code.html. Other components including region merging, multigraph matching and object labeling are all implemented in MATLAB on an Intel Core-i5 (3.33 GHz) system with 16-GB RAM.

We evaluate the proposed method for detecting multiple moving objects under unstable imaging conditions. Two data sets are used, namely, the standard DARPA VIVID data set [33] and two self-captured videos by a camera mounted on an UAV. DARPA VIVID is widely used as a benchmark for evaluating moving object detection algorithms. It contains video sequences with multiple small moving objects (~20 × 50 pixels), mainly vehicles of variable sizes and rotations. Most of the frame sequences have cluttered back- ground and undergo lighting changes. The vehicles move along roads or open areas and some are occluded by other vehicles or vegetation. Table I summarizes the properties of the five sequences (i.e., EgTest01–EgTest05) chosen from this data set. The two self-captured sequences are also described in Table I, namely, Seq01 and Seq02. These sequences contain aerial footage of vehicles moving around a university campus. Note that all these sequences have a frame rate of 30 fps. They are challenging videos which contain the especial difficulties in aerial imagery with known parameters such as flight altitude, focal length, and UAV speed as shown in Fig.2 and Fig.3.

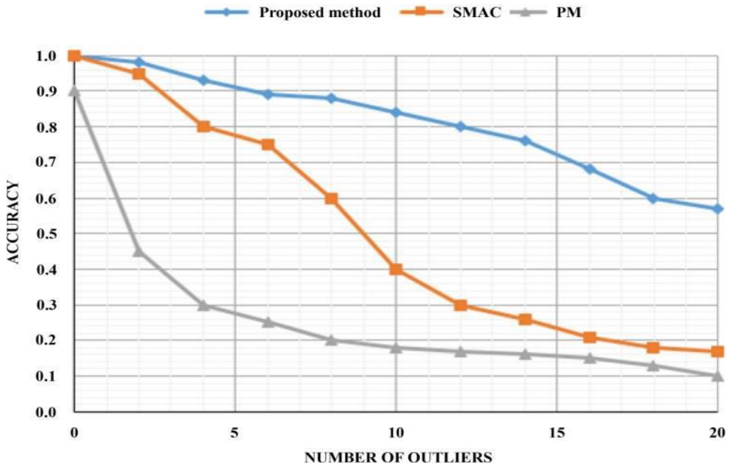

Fig.2 Comparison of the proposed method, SMAC, and PM algorithms by varying the number of outliers.

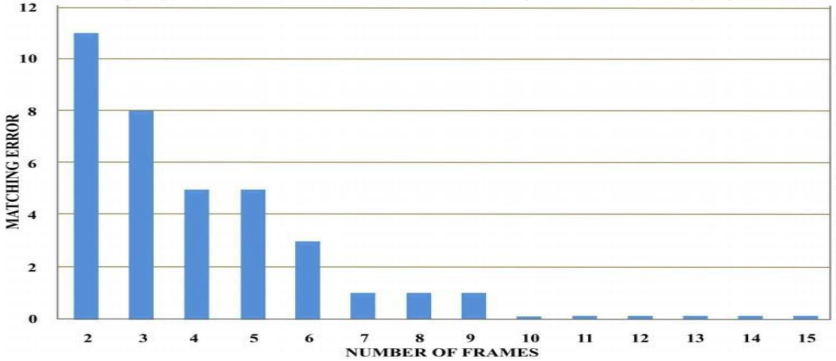

Fig.3 Matching error for different numbers of frames.

V. FUTURE SCOPE AND CONCLUSION

This paper proposes a novel approach for detecting multiple moving objects from challenging UAV-captured sequences. Video frames are first segmented into uniform regions, followed by the construction of RAGs to represent each frame. The corresponding regions are then matched between con- secutive frames by using the multigraph matching algorithm. Occluded regions are also detected in this paper. All the matched regions are then processed in groups of frames to form an MSG that keeps motion transformations of the regions in the region trajectories. Hence, multiple moving objects and background regions, which possess different motion patterns, are efficiently detected. A graph coloring algorithm finally labels objects as being background or foreground regions. The proposed method seems to benefit from the visual, spatial, and temporal features to effectively capture and represent UAV images for multiple motions estimation. Although the experiments show the proposed approach is promising for UAV environments, future works can possibly consider other domains as well.

REFERENCES

[1] Y. Tian, R. S. Feris, H. Liu, A. Hampapur, and M.-T. Sun, “Robust detection of abandoned and removed objects in complex surveillance videos,” *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 565–576, Sep. 2011.

[2] X. Zhou, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.

[3] B. Babenko, M.-H. Yang, and S. Belongie, “Robust object tracking with online multiple

instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011.

[4] W.-C. Hu, C.-H. Chen, T.-Y. Chen, D.-Y. Huang, and Z.-C. Wu, "Moving object detection and tracking from video captured by moving camera," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 164–180, Jul. 2015.

[5] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 282–295.

[6] S. R. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[7] G. Somasundaram, R. Sivalingam, V. Morellas, and N. Papanikolopoulos, "Classification and counting of composite objects in traffic scenes using global and local image analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 69–81, Mar. 2013.

[8] F. Wang, Y.-G. Jiang, and C.-W. Ngo, "Video event detection using motion relativity and visual relatedness," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 239–248.

[9] P. V. K. Borges, N. Conci, and A. Cavallaro, "Video-based human behavior understanding: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1993–2008, Nov. 2013.

[10] H. Zhou, H. Kong, L. Wei, D. Creighton, and S. Nahavandi, "Efficient road detection and tracking for unmanned aerial vehicle," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 297–309, Feb. 2015.

[11] C. Liu, P. C. Yuen, and G. Qiu, "Object motion detection using infor- mation theoretic spatio-temporal saliency," *Pattern Recognit.*, vol. 42, no. 11, pp. 2897–2906, 2009.

[12] J. Dou and J. Li, "Moving object detection based on improved VIBE and graph cut optimization," *Opt.-Int. J. Light Electron Opt.*, vol. 124, no. 23, pp. 6081–6088, 2013.

[13] B. P. Jackson and A. A. Goshtasby, "Registering aerial video images using the projective constraint," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 795–804, Mar. 2010.

[14] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.

[15] H. Yu, Y. Chang, P. Lu, Z. Xu, C. Fu, and Y. Wang, "Contour level object detection with top-down information," *Opt.-Int. J. Light Electron Opt.*, vol. 125, no. 11, pp. 2708–2712, 2014.

[16] P. Spagnolo, T. D' Orazio, M. Leo, and A. Distante, "Moving object segmentation by background subtraction and temporal analysis," *Image Vis. Comput.*, vol. 24, no. 5, pp. 411–423, 2006.

[17] G. Hong and Y. Zhang, "Combination of feature-based and area-based image registration technique for high resolution remote sensing image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 377–380.

[18] J. Le Moigne, W. J. Campbell, and R. P. Cromp, "An automated parallel image registration technique based on the correlation of wavelet features," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1849–1864, Aug. 2002.

[19] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for transla- tion, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[20] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

[21] B. Zitová, J. Flusser, and F. Sroubek, "Image registration: A survey and recent advances," in *Proc. Int. Conf. Image Process. (ICIP)*, Jan. 2005, pp. 1–52.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[23] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[24] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[25] Y. Liu, "Improving ICP with easy implementation for free-form surface matching," *Pattern Recognit.*, vol. 37, no. 2, pp. 211–226, 2004.

[26] V. Reilly, H. Idrees, M. Shah, "Detection and tracking of large number of targets in wide area surveillance," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 6313, K. Daniilidis, P. Maragos, N. Paragios, Eds. Berlin, Germany: Springer, 2010, pp. 186–199.

[27] K. Shafique and M. Shah, "A noniterative greedy algorithm for multi- frame point

correspondence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 51–65, Jan. 2005.

[28] T. Cour, P. Srinivasan, and J. Shi, “Balanced graph matching,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. 2007, pp. 313–320.

[29] 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 Recognit. (CVPR)*, Jun. 2010, pp. 679–684.

[30] R. Zass and A. Shashua, “Probabilistic graph and hypergraph matching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.