2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP) | 978-1-7281-6338-3/21/$31.00 ©2021 European Union | DOI: 10.1109/MLSP52302.2021.9596204

SMALL MOVING TARGET MOT TRACKING WITH GM-PHD FILTER AND ATTENTION-BASED CNN

Camilo Aguilar[⋆] *Mathias Ortner*[†] *Josiane Zerubia*[⋆]

[⋆] Inria, Université Côte d'Azur, Sophia Antipolis, France
[†] Airbus Defense and Space, Toulouse, France

ABSTRACT

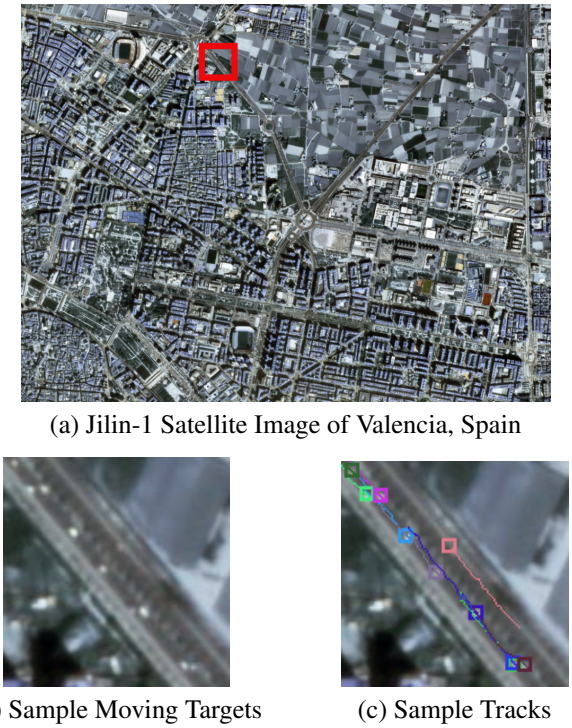
We present a multi-object tracking (MOT) approach to track small moving targets in satellite images. Our objects of interest span few pixels, do not present a defined texture, and are easily lost in cluttered environments. We propose a patch-based convolutional neural network (CNN) that focuses on specific regions to detect and discriminate nearby small objects. We use the object motion information to drive the patch selection and detect objects using a region-based CNN. In addition, we present a direct MOT data-association approach by using an improved Gaussian mixture-probability hypothesis density (GM-PHD) filter. The GM-PHD filter offers an efficient yet robust MOT formulation that takes into account clutter, misdetection, and target appearance and disappearance. We are able to detect and track blob-like moving objects and demonstrate an improvement over competing state-of-the-art tracking approaches.

***Index Terms*—** Object Detection, Satellite Object Tracking, MOT, CNN, Deep Learning, GM-PHD

1. INTRODUCTION

Multi-object tracking (MOT) constitutes a key task in numerous biological, surveillance, and remote sensing applications. Particularly, remote sensing applications have experienced an increased demand for MOT: novel Earth observation satellites are able to capture large-scale images at sub-meter resolutions that allow to detect and track small vehicles and pedestrians. For instance, the Chinese Jilin-1 satellite constellation yields ground videos at resolutions up to 0.72m, and Planet's Terra Bella Skysat-1 satellite constellation records videos up to 0.8m. Large-scale ground data will contribute to improve applications that currently rely on local ground/aerial sensors such as traffic estimation [1], traffic pattern research [2], or rare object tracking [3]. Fig. 1(a) shows an example of an image of Valencia, Spain, and Figs. 1(b)(c) show sample objects of interest and their tracks respectively.

Despite significant advancements in object tracking, small satellite target tracking remains an emerging area of research. Popular CNN approaches experience severe drawbacks when



(a) Jilin-1 Satellite Image of Valencia, Spain

(b) Sample Moving Targets (c) Sample Tracks

Fig. 1. Sample Satellite Image and Tracks.

detecting and tracking small targets in satellite images. Vehicles are often represented as blobs with an average size of 6×6 pixels, such as in Fig. 1(b). Also, texture-rich backgrounds contain blob-like objects such as air-conditioning units and electrical transformers, which generate numerous false positives for object detection. Finally, training data is sparse and often imperfect due to the difficulty of labeling small objects across numerous frames.

We approach the challenging task of small object detection by using a patch-based CNN to detect tiny moving vehicles in satellite images. We achieve this task by combining the three frame difference algorithm to drive the region selection of a CNN object detector. This approach facilitates the training of the CNN when training data is sparse because the

978-1-7281-6338-3/21/$31.00 ©2021 European Union

network learns directly from regions with objects rather than whole images and it also reduces the amount of false positives as it focuses on patches with moving targets.

In addition, we employ the Probability Hypothesis Filter (PHD) to perform object tracking. The PHD filter offers a sophisticated Bayesian data-association framework [4] that allows estimating the states of a varying number of targets given past measurements. This framework is robust to misdetections, false alarms, and appearing and disappearing objects. It also presents a significant advantage over commonly employed MOT algorithms such as probabilistic data association (PDA) [5] and multiple hypothesis tracking (MHT) [6] thanks to its relatively low computational complexity.

We propose an improved PHD filter that receives the measurements from the CNN object detector, discriminates surviving objects from appearing objects, and propagates the labels in time. The contribution of this paper is namely a track-by-detection approach by using 1) a novel method to detect small moving vehicles in satellite images and 2) the integration with an improved PHD filter to convert the measurements into tracks with a MOT statistical model.

The paper is organized as follows: Section 2 presents a literature review of state-of-the-art methods, Section 3 describes the proposed approach, first discussing the patch selection mechanism and then developing on the improvements of the PHD filter. Section 4 discusses the experiment section and Section 5 develops the conclusions and insights.

2. RELATED WORK

Common approaches for satellite object tracking involve using motion information or correlation filters. Du et al. utilized a combination of frame-difference and correlation filter to track a single object [7], Xuan et al. employed correlation filters, together with Kalman filters and linear motion to make a tracker robust to occlusions [8]. These approaches show promising results for satellite object tracking but are limited to single object tracking, require an initialization, and they generally focus on larger targets (planes/trains vs cars/pedestrians). Jiao et al. performed a survey [9] with new generation deep learning tracking approaches where numerous state-of-the-art deep learning approaches consist in learning deep features and matching them in future/past frames. However, these approaches would experience severe drawbacks when tracking small objects in satellite videos: satellite videos do not have large training datasets, targets do not exhibit distinct features, and the objects of interest are tiny relative to the satellite's field of view.

Common baselines for online multi-target trackers include Tracktor++ [10] and DeepSORT [11]. These methods obtain remarkable results in surveillance camera videos by using CNNs for detection and Kalman filters for tracking. However, their CNNs are based on Faster-RCNN and would have difficulties learning directly from the whole image.

Wei et al. presented the closest state-of-the-art approach comparable to our problem in [12]. The authors used motion information, several post-processing steps, and multiple Kalman filters to detect and track small objects in satellite videos. This approach proved promising but would underperform if cluttered objects, slow moving objects, and the data association were done using as a post processing step with the Hungarian algorithm [13]. In our approach, we incorporate deep learning for object detection, and we opt for a direct-MOT tracking with the PHD filter rather than using several individual Kalman filters.

3. PROPOSED TRACKING ALGORITHM

We use the Random Finite Set (RFS) framework to model the collection of state vectors as $\mathbf{X}_k = \{\mathbf{x}_k^1, \mathbf{x}_k^2, \dots \mathbf{x}_k^{N_k}\}$, and the frame measurements as $\mathbf{Z}_k = \{\mathbf{z}_k^1, \mathbf{z}_k^2, \dots, \mathbf{z}_k^{M_k}\}$, where N_k and M_k denote the number of objects and measurements at time k. We define a single target state vector as $\mathbf{x}_k^j = [p_k^j, v_k^j, a_k^j]^T$ where j and k are the object and time indices respectively, and $p_k^j, v_k^j, a_k^j \in \mathbb{R}^2$ denote the target's position, velocity, and acceleration vectors. Similarly, we define a measurement vector $\mathbf{z}_k^i = [\bar{p}_k^i]^T$, where $\bar{p}_k^i \in \mathbb{R}^2$ denotes the coordinates obtained by an object detector for the i^{th} measurement.

During each frame, we use the CNN-based object detector to obtain new measurements, i.e. $\mathbf{Z}_k$, and then we use the PHD filter to approximate the posterior multi-target RFS distribution: $p_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$. The summary of the proposed approach is depicted in Fig. 2.

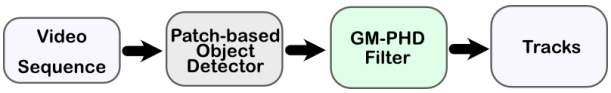


Fig. 2. Proposed Approach.

3.1. Patch-based CNN Object Detector

We propose a patch-driven approach to segment moving vehicles. This approach filters overwhelming image information and performs inference on relevant areas of interest. We first apply a motion-based detector to obtain rough object locations. In our experiments, the frame difference algorithm proved to be an effective and computationally lightweight operation and it is defined as:

$$\Delta I_k(i,j) = I_k(i,j) - I_{k-1}(i,j) \tag{1}$$

$$3FD_k(i,j) = \begin{cases} 1 & \text{if } |\Delta I_k(i,j)| + |\Delta I_{k+1}(i,j)| > T_k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Where $I_k(i,j)$ denotes the image intensity at coordinates i, j during time k and T_k is an adaptive threshold to binarize

(a) Sample Image

(b) Ground Truth

(c) 3 Frame Difference

(d) CNN Object Detector

Fig. 3. Object Detector Comparison.

the three-frame difference response. During our experiments, we obtained robust results by letting $T_k = c \cdot max(|\Delta I_k| + |\Delta I_{k+1}|)$, where $c \in (0, 1)$. Sequentially, we use binary erosion and connected components labeling to refine and transform the three-frame difference into coordinates of interest. We proceed to extract patches centered around the rough object locations to perform a refined object detection. Figs. 3 (a) (b) show a sample patch with its ground truth respectively, and Fig. 3 (c) shows the three-frame difference's output.

The three-frame difference segmentation outputs rough object locations but it does not regularize objects by area or shape and it misses slow-moving objects. For example, Fig. 3 (c) shows nearby object merging, measurement duplication for a single target, and missed objects toward the bottom of the patch. Therefore we upsample the patch and apply Faster-RCNN [14] to obtain precise object locations. Fig. 3 (d) shows that Faster-RCNN detects and discriminates nearby objects, and corrects for the missed detections. In fact, with visual inspection, the proposed object detection provides a better fit than the hand-annotated labels.

Finally, we convert the patch coordinates back to global coordinates and perform global non-maximum suppression (NMS) to discard repeated objects. This approach allows us to exploit low-availability of labeled data by focusing the network training on patches rather than whole images/videos. The summarized pipeline for this approach is depicted in Fig. 4.

3.2. Gaussian Mixture Probability Hypothesis Filter

We use the PHD filter to approximate the target's posterior RFS $p_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$ by recursively propagating its first order moment, namely its intensity function, $D_k(x)$. The GM-PHD, proposed by Vo and Ma [15] approximates a closed form solution to the PHD recursion with a weighted Gaussian mixture. It assumes that the intensity function from the previous frame has the form:

$$D_{k-1}(x) = \sum_{j=1}^{J_{k-1}} w_{k-1}^{j} \mathcal{N}(x; \mathbf{m}_{k-1}^{j}, \mathbf{P}_{k-1}^{j}) \tag{3}$$

Where ω_{k-1}^{j}, $\mathbf{m}_{k-1}^{j}$, and $\mathbf{P}_{k-1}^{j}$ are the weight, mean, and covariance for the j^{th} component and J_{k-1} is the number of components. The closed form solution for the GM-PHD prediction step is given by the equation:

$$D_{k|k-1}(x) = \lambda(x) + p_s \sum_{j=1}^{J_{k-1}} \omega_{k-1} \mathcal{N}(x; \mathbf{F}\,\mathbf{m}_{k-1}^{j}, \mathbf{Q} + \mathbf{F}\,\mathbf{P}_{k-1}^{j}\,\mathbf{F}^{T})$$

Where $\mathbf{F}$ and $\mathbf{Q}$ are the transition and motion covariance matrices in the same format as in [16], $\lambda(x)$ is the birth RFS intensity and p_s is a hyper-parameter to denotes the survival probability. Once a new set of measurements $\mathbf{Z}_k$ arrives from the object detector, we update the GM-PHD posterior following the equation:

$$D_k(x) = (1 - p_D)\, D_{k|k-1}(x) + \sum_{\mathbf{z} \in \mathbf{Z}_k} \sum_{j=1}^{J_{k|k-1}} \omega_k^{j}(\mathbf{z})\, \mathcal{N}(x;\, \mathbf{m}_{k|k}^{j}(\mathbf{z}), \mathbf{P}_{k|k}^{j})$$

Where $D_{k|k-1}(x)$ denotes the predicted GM components and p_D is a hyper-parameter to denote the detection probability. The terms $\mathbf{m}_{k|k}^{j}(\mathbf{z}), \mathbf{P}_{k|k}^{j}$ are the target-measurement association mean and covariance, which are calculated using the Kalman equations, and $\omega_k^{j}(\mathbf{z})$ denotes the updated weight for the data association and is defined as:

$$\omega_k^{j}(\mathbf{z}) = \frac{p_D\, \omega_{k|k-1}^{j} l_k^{j}(\mathbf{z})}{\kappa_k(\mathbf{z}) + p_D \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^{i} l_k^{i}(\mathbf{z})} \tag{4}$$

Where $\kappa_k(\mathbf{z})$ denotes the clutter process intensity and $l_k^{i}(\mathbf{z})$ denotes the target-measurement association likelihood.

3.3. PHD Filter Improvements

The PHD filter requires an accurate birth intensity $\lambda(x)$ modeling in order to deal with a varying number of targets. We estimate the birth intensity $\lambda(x)$ by using information from

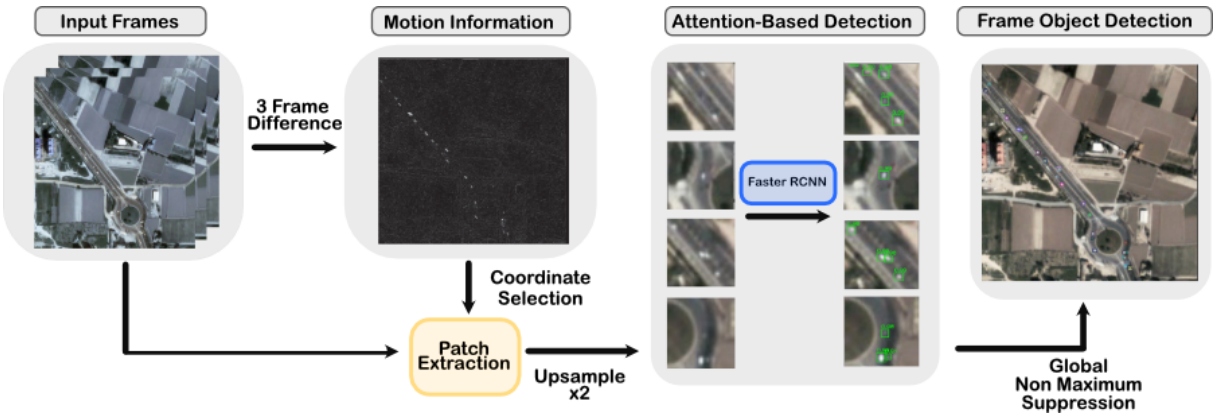


Fig. 4. Patch Based Object Detection.

the CNN object detector. We use the decomposed measurement set similar to [17], defined as $\mathbf{Z}_k = \mathbf{Z}_k^s \cup \mathbf{Z}_k^b$, where $\mathbf{Z}_k^s$ denotes the surviving targets and $\mathbf{Z}_k^b$ denotes the appearing (birth) targets. We classify a new measurement as surviving measurement (i.e. $\mathbf{z}_k^i \in \mathbf{Z}_k^s$) if it is the closest measurement to the location of a predicted GM component and if this distance is less than a threshold T_{birth}, otherwise we classify the measurement as a birth measurement (i.e. $\mathbf{z}_k^i \in \mathbf{Z}_k^b$). We approximate $\lambda(x)$ by initializing a new mixture component centered around the spatial coordinates from $\mathbf{z}_k^i \in \mathbf{Z}_k^b$ and its velocity and acceleration components set to zero.

In addition, the PHD filter does not keep tracks of labels by default, therefore we added the label tracking scheme proposed by [18]. This approach propagates labels in time in a tree-like structure without affecting the filter's performance.

4. EXPERIMENTS

4.1. Evaluation Metrics

We evaluate our approach using object detection metrics, tracking quality measures, and the ClearMOT [19] framework. We use the Hungarian algorithm [13] at each frame to match predicted hypotheses to ground truth labels and we call a hypothesis TP if it is located within a distance d_{Thrs} from the ground truth label. We choose $d_{Thrs} = 20$ pixels to account for the noise introduced by the image stabilization procedure. Similarly, we call FP any unmatched hypothesis and FN any unmatched ground truth label.

We report object detection metrics by using precision (P), recall (R), $F1$ score (F1), and the Jacard similarity index (J). These scores provide an assesment to the quality of object detection and were computed by counting TPs, FPs, and FNs over all the frames. In addition, we report tracking quality measures where we report the percentage of mostly tracked (MT) trajectories and mostly lost (ML) trajectories. We classify tracks as MT if at least 80% of the ground truth trajectory is recovered, and ML if less than 20% of the ground truth track is recovered.

Finally, we follow the ClearMOT [19] as it has become a popular and robust metric for tracking algorithms. We report the multiple object tracking accuracy (MOTA) which considers FPs, FNs, and identity switches (IDs), and we also report the multiple object tracking precision (MOTP), which considers the average distance error between the detected objects and the ground truth objects.

4.2. Satellite Dataset

The experimental satellite video can be obtained by CGSTL (available at https://mall.charmingglobe.com). This video contains 3071×4096 pixels and it represents the city of Valencia, Spain. The video contains 580 frames, lasts 29 seconds, and was imaged at 20 fps at a resolution of 1.0m by the Jilin-1 Satellite on March 7, 2017. The labels were created by the authors of [12] and they contain labels for one every 10 frames. Specifically, three regions of the image were actively labeled for evaluation purposes. These regions span 500×500 pixels and are shown in [12]. The approximate locations are: [520, 1616] for area 1, [450, 2810] for area 2, and [1074, 1895] for area 3. Due to the low availability of ground truth annotations, we tested our method on area 1 and area 2, and trained the network using 2D patches from area 3 and areas outside area 1 and area 2.

The network was trained using an NVIDIA Quadro T2000 with 4 GBs of RAM and the training time was 6 hours. In addition, we performed a pre-processing step of image stabilization by using the first frame's strong edges as reference. The stabilization step was fundamental for the three-frame difference algorithm as it contributed to significantly reduce noise introduced by the satellite motion. Experimentally, we set the three-frame difference threshold parameter $c = 0.15$, and the GM-PHD filter parameters $p_D = 0.90$ and $p_S = 0.95$.

Table 1 denotes numerical comparisons for our proposed approach vs competing methods, namely Needles in a Haystack (NIAH) [12], ViBe [20] and Gaussian Mix-

Area	**Method**	FP ↓	FN ↓	IDs↓	Precision↑	Recall↑	F1↑	J↑	MT↑	ML↓	MOTA↑	MOTP↓
1	**Proposed**	1560	**3755**	264	89.12	**77.29**	**76.01**	**70.63**	**65.31%**	10.20%	**66.3**	3.64
	NIAH [12]	935	7877	901	90.25	52.37	56.64	49.56	6.12%	6.12 %	41.30	3.66
	ViBe [20]	53993	3908	310	18.96	76.37	40.99	17.91	51.02 %	**4.08%**	-252.0	3.64
	GMM [21]	**556**	5687	**210**	**95.13**	65.61	74.32	63.48	44.90 %	18.37%	61.0	**2.50**
2	**Proposed**	**108**	**441**	**32**	87.61	63.40	**65.09**	**58.19**	**68.96%**	**17.24%**	**51.80**	3.23
	NIAH [12]	197	2953	121	**93.77**	50.08	60.55	48.47	20.69%	34.48%	44.71	2.35
	ViBe [20]	38218	2046	82	9.20	65.42	18.69	8.76	55.17%	20.69%	-582.0	3.69
	GMM [21]	1696	1915	43	70.00	**67.39**	40.07	52.29	58.62%	27.58%	37.80	**2.25**

Table 1. Tracking scores. The arrow's direction represents better scores.

(a) Raw Image (b) Ground Truth Labels (c) Sample Detection from [12] (d) Proposed Approach

Fig. 5. Area 1 Tracking Results.

ture Models (GMM) [21]. We implemented these methods following the directives in [12] and used SORT [22] as the data-association step. We stabilized the images and its respective ground truth annotations in order to reduce noise across all the methods; hence our results are slightly different than the ones reported in [12].

Our approach outperforms competing methods in comprehensive object detection and object tracking metrics. For instance, NIAH [12] and GMM [21] obtain better precision scores in Area 1, but their recall values in Area 1 are significantly lower. The addition of Faster RCNN allows our approach to discriminate nearby objects that are merged or not detected in NIAH [12] or GMM [21] such as in Fig 3. Area 2 presents numerous tiny objects such as motorcycles or small cars that are undetected by all object detectors; hence our approach obtains similar recall values. However, the combination of our object detector and GM-PHD filter allows our method to reduce FP, hence obtaining better precision scores. The trade-offs and advantages of the proposed method are denoted in Table 1 where our approach obtains the highest $F1$ scores and J scores for both areas.

The tracking performance is also outlined in the number of mostly tracked (MT) trajectories, the MOTA score, and MOTP score. Our approach obtains a slightly worse MOTP score (1 pixel difference) due to the modeling of uncertainty in the GM-PHD-filter; however, our method obtains more MT trajectories, lower identity switches (IDs), and better MOTA thanks to the employment of the GM-PHD filter. Particularly, the GM-PHD reduces clutter detections and keeps track of undetected objects for several frames.

5. CONCLUSION

In this paper, we presented a track-by-detection approach combining a patch-based CNN object detector and the PHD filter. The patch selection mechanism contributes to filter unnecessary information and to train the neural network in a dataset with sparse labels. In addition, we use the GM-PHD algorithm to track multiple targets while reducing the amount of clutter. Our approach presents an increased computational burden due to the patch selection and CNN inference, but its results outperform competing methods in both object detection and tracking scores for the challenging task of detecting small satellite objects.

6. ACKNOWLEDGMENT

The authors are grateful to the OPAL infrastructure from Université Côte d'Azur for providing resources and support. Additionally, the authors would like to thank BPI France for the financial support under the LiChiE contract.

7. REFERENCES

[1] Lynn H. Kaack, George H. Chen, and M. Granger Morgan, "Truck traffic monitoring with satellite images," in *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, Ghana, 2019.

[2] Yulu Chen, Rongjun Qin, Guixiang Zhang, and Hessah Albanwan, "Spatial temporal analysis of traffic patterns during the covid-19 epidemic by vehicle detection using Planet remote-sensing satellite images," *Remote Sensing*, vol. 13, no. 2, 2021.

[3] Michael R. Peterson, Gary B. Lamont, Frank Moore, and Patrick Marshall, "A satellite image set for the evolution of image transforms for defense applications," in *Proceedings of the 9th Annual Conference Companion on Genetic and Evolutionary Computation*, United Kingdom, 2007.

[4] R.P.S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, 2003.

[5] Yaakov. Bar-Shalom and Thomas E. Fortmann, *Tracking and data association / Yaakov Bar-Shalom, Thomas E. Fortmann*, Academic Press Boston, 1988.

[6] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[7] Bo Du, Yujia Sun, Shihan Cai, Chen Wu, and Qian Du, "Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm," *IEEE Geoscience and Remote Sensing Letters*, December 2017.

[8] Shiyu Xuan, Shengyang Li, Mingfei Han, Xue Wan, and Gui-Song Xia, "Object tracking in satellite videos by improved correlation filters with motion estimations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1074–1086, 2020.

[9] Licheng Jiao, Ruohan Zhang, Fang Liu, Shuyuan Yang, Biao Hou, Lingling Li, and Xu Tang, "New generation deep learning for video object detection: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, February 2021.

[10] Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe, "Tracking without bells and whistles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[11] Nicolai Wojke, Alex Bewley, and Dietrich Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017.

[12] Wei Ao, Yanwei Fu, Xiyue Hou, and Feng Xu, "Needles in a haystack: Tracking city-scale moving vehicles from continuously moving satellite," *IEEE Transactions on Image Processing*, vol. 29, pp. 1944–1957, 2020.

[13] H. W. Kuhn and Bryn Yaw, "The Hungarian method for the assignment problem," *Naval Res. Logist. Quart*, pp. 83–97, 1955.

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015.

[15] Ba-Ngu Vo. and Wing-Kin Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, 2006.

[16] Michele Pace, *Stochastic models and methods for multi-object tracking*, PhD thesis, Université Sciences et Technologies - Bordeaux, I, France, July 2011.

[17] Zeyu Fu, Federico Angelini, Syed Mohsen Naqvi, and Jonathon A. Chambers, "GM-PHD filter based online multiple human tracking using deep discriminative correlation matching," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[18] Kusha Panta, Daniel E. Clark, and Ba-Ngu Vo, "Data association and track management for the Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, 2009.

[19] Keni Bernardin and Rainer Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP Journal on Image and Video Processing*, 2008.

[20] Yun Yang, Deqiang Han, Jiankun Ding, and Yi Yang, "An improved ViBe for video moving object detection based on evidential reasoning," in *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2016.

[21] Adi Nurhadiyatna, Wisnu Jatmiko, Benny Hardjono, Ari Wibisono, Ibnu Sina, and Petrus Mursanto, "Background subtraction using Gaussian mixture model enhanced by hole filling algorithm (GMMHF)," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013.

[22] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016.