Anomalous Driving Detection for Traffic Surveillance Video Analysis

Hang Shi, Hadi Ghahremannezhad, and Chengjun Liu
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, 07102 USA
hs328@njit.edu, hg255@njit.edu, cliu@njit.edu

***Abstract*—Traffic safety is an important topic in the intelligent transportation system. One major factor that causes traffic accident is anomalous driving. This paper presents a novel anomalous driving detection method in videos, which can detect unsafe anomalous driving behaviors. The contributions of this paper are three-fold. First, a new multiple object tracking (MOT) method is proposed to extract the velocities and trajectories of moving foreground objects in video. The new MOT method is a motion-based tracking method, which integrates the temporal and spatial features. Second, a novel Gaussian local velocity (GLV) modeling method is presented to model the normal moving behavior in traffic videos. The GLV model is built for every location in the video frame, and updated online. Third, a discrimination function is proposed to detect anomalous driving behaviors. Experimental results using the real traffic data from the New Jersey Department of Transportation (NJDOT) show that our proposed method can perform anomalous driving detection fast and accurately.**

I. Introduction

Traffic video analysis is a compelling topic in computer vision. How to use AI technologies to improve traffic safety is always a major concern of society. Anomalous driving behaviors are one of the most dangerous behaviors in traffic. Every year, thousands of lives are lost in traffic due to anomalous driving. In this paper, we propose an anomalous driving detection method, which can detect anomalous driving behaviors in traffic surveillance videos fast and accurately.

Many methods are proposed to detect the anomalies in videos bases on spacial feature representations [1], [2], [3]. This kind of approach is generally applied to detect all kinds of anomalies without specification. People want to utilize anomaly detection methods in real-world applications. Therefore, more methods are proposed with a specific concentration. To concentrate the attention on anomalous driving behaviors, we proposed our anomalous driving detection method, which integrates the MOT method, the GLV model, and the discrimination function.

We first propose a new multiple object tracking (MOT) method to track the motions of the vehicles in traffic. Many MOT methods are proposed based on a track-by-detection procedure, which needs high computational power. Our proposed MOT method is based on spatial and temporal feature representations, which can process the video frames fast and accurately. By considering the spatial and temporal distance between the objects in two adjacent frames, our MOT method can quickly match the same object in two adjacent frames.

Second, we use a novel Gaussian local velocity (GLV) modeling method to model the normal driving behaviors in traffic. In order to detect the anomalies in traffic, we consider modeling the normal behaviors. Everything that cannot be classified in the normal class is an anomaly. As we know, the vehicles in traffic should be driving in a queue. Every vehicle that passes the same location should follow similar speeds and directions, otherwise, traffic accidents may happen. Therefore, we can use a Gaussian distribution to model the normal driving speed and direction at a specific time and location. The Gaussian distribution can be updated online while the speed may change over time.

Third, we use a discrimination function to distinguish the anomalous driving behaviors in videos. With the GLV model, we can describe the normal driving behaviors in traffic. Then we can use a discrimination function to classify each moving vehicle in traffic into a normal class or an anomalous class. Figure. 1 shows the workflow of our proposed anomalous driving detection method.

There are three major contributions of our proposed method. First, the method does not need any training process. The foreground segmentation, MOT process, and GLV modeling method are all unsupervised methods. The models are built when the video is processing. The whole process doe not require any manual label work which increases the general ability of our method. Second, the computational complexity of our proposed method is low and the processing speed is fast. Recently, many anomaly detection methods are proposed based on deep neural networks. However, the implementation of this kind of methods requires high-performance GPU, and some of them still cannot process the video in real-time. Unlike the deep neural networks, our statistical modeling method has low computational complexity. Even on a regular desktop PC, our method can reach a processing speed of 60 frames per second or faster. Third, our method is robust and generalized. We tested our method on dozens of real traffic videos with different illumination conditions, weather, and resolutions. Our method is able to detect anomalous driving behaviors accurately and robustly. The accuracy is an important criterion to evaluate an anomalous driving detection

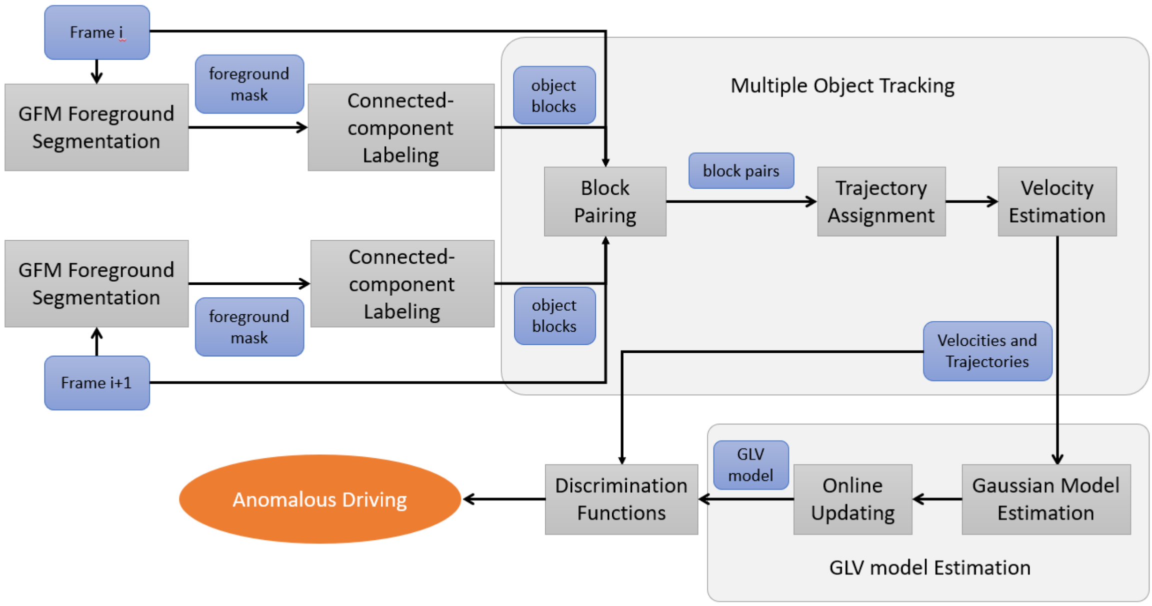

Fig. 1. The workflow of our proposed anomalous driving detection method using the new MOT method and the novel GLV model.

method. Our method reaches a 100% detection rate without a false alert during testing.

II. Related Work

A. Foreground Segmentation

Most video analytic tasks in computer vision applications involve segmenting each of the video frames into moving foreground and stationary background as an initial step. In recent years, foreground segmentation has been one of the most studied fields in computer vision with applications in target tracking, behavior understanding, multimedia, and intelligent video surveillance. The accuracy of foreground segmentation as an initial preprocessing step has a major influence on the overall performance of the application in its following steps. However, this accuracy can be challenged by changes in illumination conditions, noises, and clutter in the background, occlusion, flickering camera, etc. A significant amount of work has been carried out on improving the reliability and robustness of foreground segmentation methods.

Many Unsupervised methods extract the moving foreground by applying spatial and temporal features to model the background and subtract it from the foreground [4]–[7]. In [8], an adaptive Gaussian mixture model (AGMM) is applied to extract foreground objects and is followed by a BP neural network as a post-process step to deal with challenging scenes with illumination changes or moving shadows. Wang et al. [9] apply the scene division method for foreground segmentation based on a spatio-temporal stochastic update strategy in GMM. [**?**] proposes a global foreground modeling (GFM) method, which considers the global information of the foreground. In recent years, many supervised foreground segmentation methods have been proposed based on deep learning algorithms [10], [11]. However, the supervised methods depend on models that are trained on a number of visually similar videos and perform purely when tested on unseen videos with different visual aspects. As foreground segmentation is the very first step of our proposed system, we choose to use fast and efficient unsupervised foreground detection methods to segment the foreground information from frames.

B. Multiple Object Tracking

Multiple Object Tracking (MOT) is one of the essential steps in most video analytic tasks. Tracking multiple objects in videos at the same time involves the detection of objects at each video frame and the association of the detected objects across multiple consecutive frames. There has been a significant amount of research conducted on Multiple Object Tracking (MOT) in recent years. Tang et al. [12] deal with the MOT problem as a Minimum Cost Subgraph Multicut Problem and apply the Kernighan-Lin algorithm to solve it. [13] utilizes a fully differentiable graph-based framework and a Message Passing Network (MPN) to propagate the node features throughout the graph. Maksai et al. [14] use behavioral patterns to propose a non-Markovian approach in order to impose global consistency and further improve upon state-of-the-art tracking algorithms. In [15], a deep prediction-decision network is developed in a collaborative deep reinforcement learning (C-DRL) method, which simultaneously detects and predicts objects under a unified network.

With the improvements in object detection methods in recent years, tracking by detection has been the most studied approach in multi-target tracking. In [16], the multi-object state is modeled as a labeled random finite set and using the Bayes recursion to reduce false negatives and false positives and propagate the multi-object filtering density forward in time. A CNN-based framework is proposed by Chu et al. [17], which uses single object tracking (SOT) to enrich detections in MOT. Jorquera et al. [18] use the Probability Density Hypothesis (PHD) filter and Determinantal Point Processes (DPP) to deal with data association uncertainty, noise, and false alarms and improve the detection accuracy.

C. Anomaly Detection in Video

In most video surveillance systems, anomaly detection is an important component. Recently, anomaly detection in videos has been an active research area with applications in intelligent video surveillance and secure related video analytic tasks. Since the occurrence of abnormal actions in real-world video analysis applications is infrequent, detecting anomaly in videos automatically reduces a significant amount of manual work.

Some of the recent research efforts have tried to detect the anomalies in videos. A deep CNN is proposed by Nguyen et al. [19], which is a combination of a reconstruction network that determines the main structures that appear in video frames and an image translation model which associates motion templates to such structures. In [20], anomaly detection in videos is dealt with within a video prediction framework in which a future frame is predicted based on spatial and temporal constraints and then compared to its ground truth to detect abnormal incidents. An end-to-end network is proposed by Tang et al. [21] that conducts future frame prediction which enlargers the reconstruction errors to help with the identification of abnormal events followed by reconstruction which helps enhance the predicted future frames from normal events.

Besides the methods that are aiming at detecting anomalies in general videos, many methods are proposed to detect anomalies in traffic videos. Doshi and Yasin proposed an unsupervised method to detect anomalies in traffic videos [22]. [23] presents a 3-stage pipeline for anomaly detection in traffic videos. This kind of approach considers the specific scenarios in traffic, and are aiming at landing in real-world applications. In this paper, we propose an anomalous driving detection method, which uses a Gaussian model to model the normal driving behaviors, and detects the anomaly by a discrimination function in traffic videos.

III. The New Multiple Object Tracking Method Based on Temporal and Spacial Features

Object tracking is an important topic in video processing. Many methods are proposed to track the moving objects in videos [24], [25], [26]. Multiple object tracking (MOT), which is an extension of single object tracking, plays a more important role in AI applications. People want to catch the trajectories of each individual moving items with MOT, and further tackle some high-level tasks with the tracking results, such as action recognition [27], anomaly detection [28], and object counting [29]. In this section, we propose a new MOT method using the spatial and temporal information in videos, which can tracking multiple objects fast and accurately.

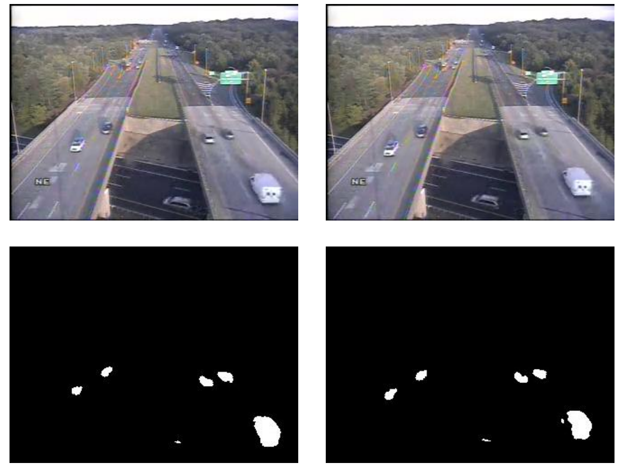

Fig. 2. Two adjacent video frames from the NJDOT video and their corresponding foreground masks.

To utilize the temporal information in videos, we apply the GFM foreground segmentation method to extract the foreground mask [4], [5]. The GFM method is an unsupervised foreground segmentation method, which can detect the foreground pixels fast and accurately. In the GFM method, foreground pixels are detected using the Bayes decision rule for minimum error, which is described as [30]:

$$C\left(x_{i,j}\right) = \begin{cases} 1, & p(\mathbf{x_{i,j}}|\omega_f)P(\omega_f) \geqslant p(\mathbf{x_{i,j}}|\omega_b)P(\omega_b) \\ 0, & Otherwise \end{cases} \tag{1}$$

where $x_{i,j}$ represents the pixel at location (i,j), $p(\mathbf{x}|\omega_f)$ and $p(\mathbf{x}|\omega_f)$ means the conditional probability density functions (CPDF) for the foreground and background, respectively, $P(\omega_f)$ and $P(\omega_b)$ denote the prior probability of the foreground and background, respectively, and C is a binary mask where $C\left(x_{i,j}\right) = 1$ means the pixel at location (i,j) is classified as foreground. The foreground mask represents the moving components in a video. For two adjacent frames $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$, we can get the foreground masks using the GFM method. Figure. 2 shows two adjacent video frames from an NJDOT traffic video and the corresponding foreground masks detected by the GFM method.

In each foreground mask, there are some foreground pixels, which can indicate the moving objects. A connected-component labeling method [31] is applied to the foreground mask to label every connected region in the foreground mask with a block id. Then we can get two sets of block $S_i = \{B_{i,1}, B_{i,2}, \dots, B_{i,m}\}$, and $S_{i+1} = \{B_{i+1,1}, B_{i+1,2}, \dots, B_{i+1,n}\}$ from the two adjacent frames. For every B in set S_i, we pair it with one block B' in set S_{i+1}, which can minimize the distance function $Dist(B, B')$.

$$Dist(B, B') = \frac{Euc\_Dist(B, B')}{Cos\_Dist\left(B, B'\right)} \tag{2}$$

$$Euc\_Dist(B, B') = e^{|Cent_B - Cent_{B'}|} \tag{3}$$

$$Cos\_Dist(B, B') = \frac{f_B}{||f_B||} \cdot \frac{f_{B'}}{||f_{B'}||} \tag{4}$$

where $Cent_B$ and $Cent_{B'}$ means the centroids of block B and B', f_B and $f_{B'}$ means the spacial feature vector extracted from the video frame at the corresponding location of block B and B'. In this paper, we select the mean and variance of the block area as the spatial features. Note that both the mean and variance are non-negative numbers, the feature vectors are all fall in the same quadrant. Therefore the Cosine similarity is always positive. The novel distance function $Dist(B, B')$, on one hand, considers the temporal information, which is the Euclidean distance between the blocks, on the other hand, considers the spatial information, which is the Cosine similarity between the blocks.

After all the blocks in set S_i have paired with a block in set S_{i+1}, we apply the Algorithm. 1 to convert the block pairs to trajectories.

Algorithm 1 pseudocode of MOT method

if i is the first frame **then**
  Create a new trajectory for each block in S_i.
end if
for all paired blocks $\{B, B'\}$ **do**
  if The value of the $Dist(B, B')$ is over the threshold **then**
    Discard the pair and mark the trajectory associated with B as END.
  else
    Assign block B' to the trajectory of B, and add the connection between B and B' into the trajectory.
  end if
end for
for all blocks in S_{i+1} that do not have paired blocks in S_i **do**
  Crease a new trajectory.
end for

Different from the detection based tracking methods, our proposed MOT method does not need to detect each individual item, which can save computational power. Our proposed method is based on a motion-tracking procedure and further involves the spatial features to define an innovative distance function. The combination of the temporal and spatial features can improve the tracking performance with a limited computational requirement.

IV. The Anomalous Driving Detection using a Novel Gaussian Local Velocity Model

Anomalous driving is one of the most dangerous behaviors in traffic. Many traffic accidents are caused by anomalous driving behaviors, such as wrong-way driving, sudden lane merge, and stopped in traffic. To detect the anomalous driving behaviors, we propose the novel Gaussian local velocity (GLV) model to model the normal driving behaviors and use a set of discrimination functions to identify the anomalies in traffic surveillance videos.

Fig. 3. The moving trajectories detected by our new MOT method. The red lines indicate the moving trajectories of the vehicles in two seconds.

Our proposed GLV model uses an unsupervised online updating strategy to establish the model. First, we use the velocities achieved from the MOT method in each frame as input feature vectors to estimate an initial GLV model. The model is updated every frame to satisfy the speed change of traffic. Finally, we apply the discrimination functions to the trajectories of each individual foreground object achieved from the MOT method to determine if the driving behavior of that object is anomalous.

Our proposed method is mainly focusing on traffic surveillance video analysis. As we know, vehicles in traffic are moving along the traffic lanes, every vehicle normally moves along a similar direction at the same location. Therefore, we can build a Gaussian distribution for every location in a frame to model the normal velocity. For each location i, j in a frame, the feature vector $\mathbf{x}_{i,j}$ is composed of the magnitude $s_{i,j}$ and the angle $\theta_{i,j}$. The GLV model at location i, j can be described as:

$$V_{i,j} = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}_{i,j} - \mathbf{M}_{i,j})^t \Sigma_{i,j}^{-1} (\mathbf{x}_{i,j} - \mathbf{M}_{i,j})\right\}}{(2\pi)^{d/2} \mid \Sigma_{i,j} \mid^{1/2}} \tag{5}$$

where d is the dimensionality of the feature vector $\mathbf{x}_{i,j}$, $\mathbf{M}_{i,j}$ is the mean vector, and $\Sigma_{i,j}$ is the covariance matrix. The GLV model is updated with every moving object passed by the location.

The GLV model describes the normal moving behaviors in traffic. If the motion feature of an object is far from the mean vector of the GLV model, it can be considered as an anomalous motion. We further propose an anomalous driving detection method based on the GLV model.

As we know, the normal driving vehicles should follow a similar direction to that of the traffic flow. If a vehicle drove in the wrong direction that is away from the traffic flow, it may cause a traffic accident. Our anomalous driving detection is

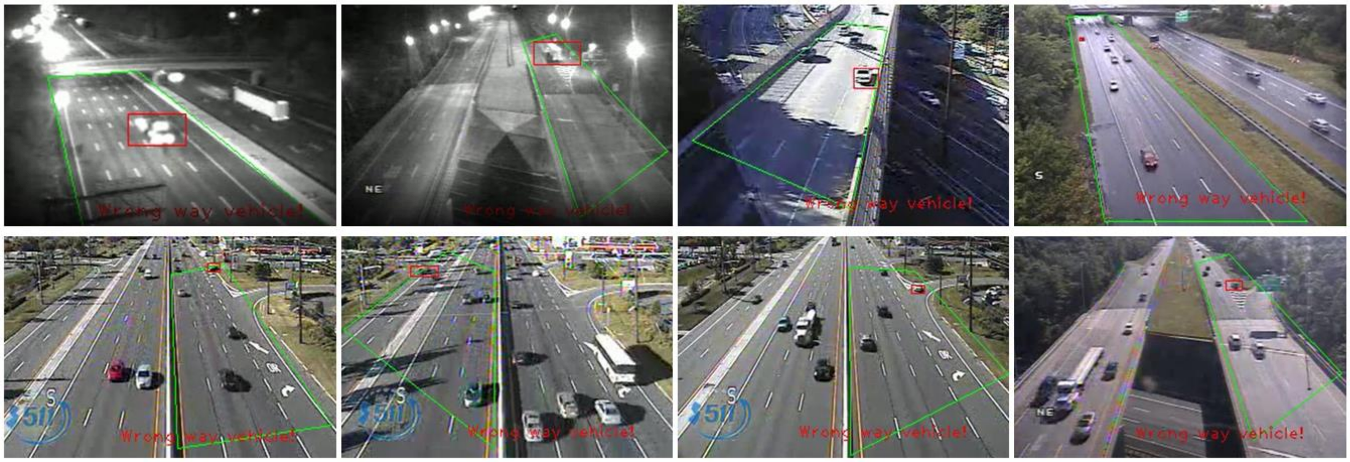

Fig. 4. The wrong-way driving vehicles detected by our proposed method. The red rectangles are the wrong-way driving vehicles detected by our proposed method, the areas in the green lines are the region of interest (ROI).

aiming at detecting and alerting this kind of wrong-way driving behavior. We propose a discrimination function to identify the anomalous driving based on the GLV model:

$$D(\mathbf{x}_{i,j}) = \frac{\mathbf{x}_{i,j}}{||\mathbf{x}_{i,j}||} \cdot \frac{\mathbf{M}_{i,j}}{||\mathbf{M}_{i,j}||} \tag{6}$$

where $\mathbf{x}_{i,j}$ is the motion feature vector of a vehicle, and $\mathbf{M}_{i,j}$ is the mean vector of the GLV model at location i, j. If the discrimination function $D(\mathbf{x}_{i,j}) < 0$, we classify that vehicle as an anomalous driving. If the duration of the anomalous driving behavior of a specific vehicle is longer than the threshold, we identify it as dangerous and send an alert.

V. Experiments

To evaluate our proposed anomalous driving detection method, we run experiments on the real traffic video sequences from the New Jersey Department of Transportation (NJDOT). We used a desktop with an Intel Core i7-8700 Processor to implement our proposed method. Our method can process 65 frames per second (fps) for the videos with the spatial resolution of 352×240, which is a general resolution of the traffic surveillance videos.

We first present the results achieved by our proposed MOT method. The processing speed is a very important factor of an MOT method. For a video frame with a resolution of 352×240, the average processing time of our proposed MOT method is 13 ms, which is 77 frames per second or fps. Figure. 3 shows the vehicle moving trajectories detected by our proposed MOT method. The red lines show 2 seconds of the vehicle moving trajectories. We can see the normal moving vehicles have relatively straight moving trajectories along the road.

The GLV model is built after the moving trajectories are extracted. By utilizing the GLV model, we detect anomalous driving behaviors in traffic videos. The most commonly seen anomalous driving in traffic is wrong-way driving. Drivers may back up the vehicles in traffic due to missing the existing ramp or entering a wrong ramp. Our proposed method is able to detect these dangerous behaviors in order to minimize the traffic accident happening. Figure. 4 shows some wrong-way driving vehicles detected by our anomalous driving detection method. We tested our proposed methods on dozens of different real traffic scenarios, our method can reach a 100% detection rate without a false alert. We can see our method is able to deal with both the night videos and daytime videos.

VI. Conclusion

We proposed a novel anomalous driving behavior detection method for traffic surveillance video analysis in this paper. The method integrates a new multiple object tracking (MOT) method, a novel GLV model, and an anomalous driving discrimination function to detect and alert the wrong-way driving behaviors. The MOT method utilizes the spatial and temporal information in the video and is able to process the video fast and accurately. The GLV model is built locally using the Gaussian distributions and is updated online. The discrimination function can classify each moving vehicle into a normal driving class or an anomalous driving class. There are three advantages of our proposed method. First, the method proposed is based on statistical modeling. The estimation of the statistical models does not require any labeled data. This can reduce manual labeling work and increase the generalization ability of our method. Second, the computational complexity of our proposed method is low. The anomalous driving detection can process over 60 frames per second on a normal PC. Third, the anomalous driving detection method is accurate and robust to the real-world situation. We tested on dozens of different traffic video scenarios, all the anomalous drivings can be detected without false alerts. The experimental results using the New Jersey Department of Transportation (NJDOT) real traffic videos show the feasibility of our proposed method.

VII. Acknowledgments

This paper is partially supported by the NSF grant 1647170.

REFERENCES

[1] N. Patil and P. K. Biswas, "A survey of video datasets for anomaly detection in automated surveillance," in *Sixth International Symposium on Embedded Computing and System Design, ISED 2016, Patna, India, December 15-17, 2016*. IEEE, 2016, pp. 43–48.

[2] K. Cheng, Y. Chen, and W. Fang, "Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5288–5301, 2015.

[3] H. Vu, T. D. Nguyen, T. Le, W. Luo, and D. Q. Phung, "Robust anomaly detection in videos using multilevel representations," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 5216–5223.

[4] H. Shi and C. Liu, "A new foreground segmentation method for video analysis in different color spaces," in *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20-24, 2018*. IEEE Computer Society, 2018, pp. 2899–2904.

[5] ——, "A new global foreground modeling and local background modeling method for video analysis," in *Machine Learning and Data Mining in Pattern Recognition - 14th International Conference, MLDM 2018, New York, NY, USA, July 15-19, 2018, Proceedings, Part I*, ser. Lecture Notes in Computer Science, P. Perner, Ed., vol. 10934. Springer, 2018, pp. 49–63.

[6] S. Isik, K. Özkan, S. Günal, and Ö. N. Gerek, "Swcd: a sliding window and self-regulated learning-based background updating method for change detection in videos," *Journal of Electronic Imaging*, vol. 27, no. 2, p. 023002, 2018.

[7] S.-h. Lee, G.-c. Lee, J. Yoo, and S. Kwon, "Wisenetmd: Motion detection using dynamic background region analysis," *Symmetry*, vol. 11, no. 5, p. 621, 2019.

[8] X. Ou, P. Yan, W. He, Y. K. Kim, G. Zhang, X. Peng, W. Hu, J. Wu, and L. Guo, "Adaptive gmm and bp neural network hybrid method for moving objects detection in complex scenes," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 02, p. 1950004, 2019.

[9] Z. Wang, C. Cheng, and J. Yang, "Scene division-based spatio-temporal updating mixture gaussian model for moving target detection," in *Proceedings of the 2018 7th International Conference on Software and Computer Applications*, 2018, pp. 169–172.

[10] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognition*, vol. 76, pp. 635–649, 2018.

[11] L. A. Lim and H. Y. Keles, "Learning multi-scale features for foreground segmentation," *Pattern Analysis and Applications*, pp. 1–12, 2019.

[12] S. Tang, B. Andres, M. Andriluka, and B. Schiele, "Subgraph decomposition for multi-target tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5033–5041.

[13] G. Braso and L. Leal-Taixe, "Learning a neural solver for multiple object tracking," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[14] A. Maksai, X. Wang, F. Fleuret, and P. Fua, "Non-markovian globally consistent multi-object tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2544–2554.

[15] L. Ren, J. Lu, Z. Wang, Q. Tian, and J. Zhou, "Collaborative deep reinforcement learning for multi-object tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 586–602.

[16] D. Y. Kim, B.-N. Vo, B.-T. Vo, and M. Jeon, "A labeled random finite set online multi-object tracker for video data," *Pattern Recognition*, vol. 90, pp. 377–389, 2019.

[17] Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu, "Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4836–4845.

[18] F. Jorquera, S. Hernández, and D. Vergara, "Probability hypothesis density filter using determinantal point processes for multi object tracking," *Computer Vision and Image Understanding*, vol. 183, pp. 33–41, 2019.

[19] T.-N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[20] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection – a new baseline," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[21] Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, and J. Yang, "Integrating prediction and reconstruction for anomaly detection," *Pattern Recognition Letters*, vol. 129, pp. 123–130, 2020.

[22] K. Doshi and Y. Yilmaz, "Fast unsupervised anomaly detection in traffic videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 624–625.

[23] K. M. Biradar, A. Gupta, M. Mandal, and S. K. Vipparthi, "Challenges in time-stamp aware anomaly detection in traffic videos," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 2019, pp. 13–20.

[24] L. Fan, Z. Wang, B. Cai, C. Tao, Z. Zhang, Y. Wang, S. Li, F. Huang, S. Fu, and F. Zhang, "A survey on multiple object tracking algorithm," in *IEEE International Conference on Information and Automation, ICIA 2016, Ningbo, China, August 1-3, 2016*. IEEE, 2016, pp. 1855–1862.

[25] Y. Xu, X. Zhou, S. Chen, and F. Li, "Deep learning for multiple object tracking: a survey," *IET Comput. Vis.*, vol. 13, no. 4, pp. 355–368, 2019.

[26] H. Ooi, G. Bilodeau, and N. Saunier, "Supervised and unsupervised detections for multiple object tracking in traffic scenes: A comparative study," in *Image Analysis and Recognition - 17th International Conference, ICIAR 2020, Póvoa de Varzim, Portugal, June 24-26, 2020, Proceedings, Part I*, ser. Lecture Notes in Computer Science, A. Campilho, F. Karray, and Z. Wang, Eds., vol. 12131. Springer, 2020, pp. 42–55.

[27] J. Choi, Y. Cho, K. Cho, S. Bae, and H. S. Yang, "A view-based multiple objects tracking and human action recognition for interactive virtual environments," *IJVR*, vol. 7, no. 3, pp. 71–76, 2008.

[28] M. Han and Y. Gong, "Real-time multiple-object tracking and anomaly detection," in *Storage and Retrieval Methods and Applications for Multimedia 2005, San Jose, CA, USA, January 18, 2005*, ser. SPIE Proceedings, R. Lienhart, N. Babaguchi, and E. Y. Chang, Eds., vol. 5682. SPIE, 2005, pp. 173–182.

[29] C. R. del-Blanco, F. Jaureguizar, and N. García, "An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications," *IEEE Trans. Consumer Electronics*, vol. 58, no. 3, pp. 857–862, 2012.

[30] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.

[31] L. He, Y. Chao, K. Suzuki, and K. Wu, "Fast connected-component labeling," *Pattern recognition*, vol. 42, no. 9, pp. 1977–1987, 2009.