A Multi Sensor Real-time Tracking with LiDAR and Camera

Surya Kollazhi Manghat
Department of Electrical and Computer Engineering
Purdue School of Engineering and Technology Indianapolis
sukoll@iu.edu

Mohamed El-Sharkawy
Department of Electrical and Computer Engineering
Purdue School of Engineering and Technology Indianapolis
melshark@iupui.edu

***Abstract*—Self driving cars are equipped with various driver-assistive technologies (ADAS) like Forward Collision Warning system (FCW), Adaptive Cruise Control and Collision Mitigation by Breaking (CMbB) to ensure safety. Tracking plays an important role in ADAS systems for understanding dynamic environment. This paper proposes 3D multi-target tracking method by following a lean way of implementation using object detection with aim of real time. Object Tracking is an integral part of environment sensing, which enables the vehicle to estimate the surrounding object's trajectories to accomplish motion planning. The advancement in the object detection methodologies benefits greatly when following the tracking by detection approach. The proposed method implemented 2D tracking on camera data and 3D tracking on LiDAR point cloud data. The estimated state from each sensors are fused together to come with a more optimal state of objects present in the surrounding. The multi object tracking performance has evaluated on publicly available KITTI dataset.**

Index Terms*— *Autonomous Vehicles, Camera, LiDAR, Tracking, KITTI, Object Detection, FCW, ADAS, State Estimation, Multi-sensor Fusion.

I. Introduction

The Autonomous cars combine a variety of sensors like LiDAR, camera, radar, sonar, GPS, odometry and inertial measurement units to perceive their surroundings with enhanced data quality. Interpreting the sensory information aids in perception of the environment, which can be used for Navigation and Control of a vehicle. Object Detection and Tracking are the main methods to perceive information from sensors. The Object Detection gives details of the presence of objects in a frame, at the same time Object Tracking goes beyond Object Detection and it monitor objects using a dynamic model allotted to it. An autonomous vehicle must be aware of the position and dynamic information of all moving objects encountered in the environment to come up with a path planning algorithm or any ADAS algorithms. By using a series of measurements in each video frame made over time, motion tracking can estimate, predict present and future locations.

The detection and tracking of objects of surrounding assist the intelligent autonomous vehicle for forward collision warning, path panning, Adaptive Cruise Control etc. In forward collision avoidance technique, camera based object tracking helps to identify the potential collision threats by giving their relevance in the planned path of the car. Knowing the presence

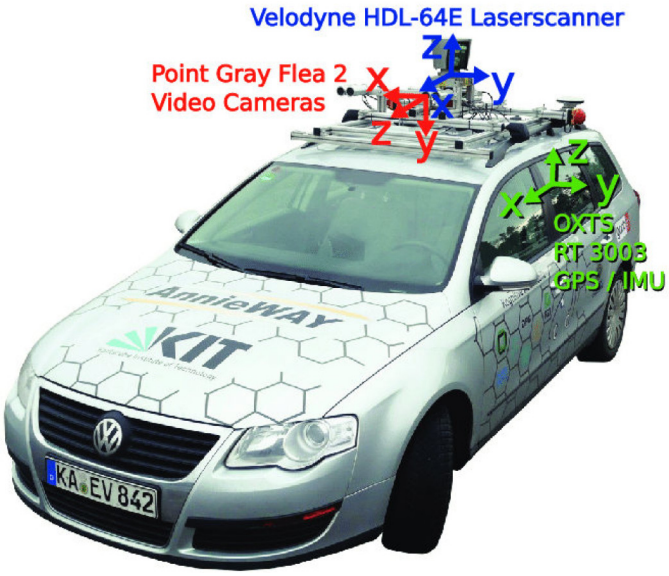

Fig. 1. KITTI Data Recording Platform Setup [7]

of moving objects around the vehicle helps a driver assistant system to alert the driver of collisions chances and hazard [1]. Other application in the autonomous transportation systems is the design of an optimal trajectory for one vehicle under normal conditions and in unexpected conditions such as when overtaking a single moving vehicle on a predetermined road. Path planning for autonomous vehicles is to plan in real time a collision-free path in the presence of dynamically moving objects and with a limited sensing range. This way camera based target tracking can assist the autonomous vehicle to drive through all area in the camera range and avoid obstacles.

Visual Multi Object Tracking estimate the object trajectories according to image sequence identities. Among numerous types of proposed MOT methods, Tracking by detection is the popular one for online tracking as it is less complex. Tracking by detection methods can be broadly divided in to online and offline tracking (batch or semi batch) methods. The offline/batch methods use current and future frame for the detection results. Tracklets are generated by linking the detections from the frames and associating iteratively to construct the trajectory of objects in the entire sequence. But Online

This is the author's manuscript of the article published in final edited form as:

Manghat, S. K., & El-Sharkawy, M. (2020). A Multi Sensor Real-time Tracking with LiDAR and Camera. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 0668–0672. https://doi.org/10.1109/CCWC47524.2020.9031247

MOT algorithms estimate the trajectory using the detections from past and current frames, are more applicable to real time applications such as ADAS, FCW and Navigation.

Target tracking has been a research area for decades with numerous applications [21, 17]. Various methodologies have been introduced to solve the efficiency in real time [1, 2] and the occlusion problem. There are offline tracking algorithms which evaluate on past and future frames to generate efficient tracklets. But when comes to real time applications tracking should be online tracking methods [1, 2, 9]. With recent advancement in the detection domains including CNN based [17] and traditional feature vectors approaches, the missed detections can be decreased, the precise bounding box can be located. The advancement in the detection and higher frame rates simplifies the tracking method. The simple IOU tracker [9] introduced a method of data association with out using visual information, thus reported a decrease in the computational complexity and processing time. The method proposed by Bochinski, Erik, et al.[31] integrated visual information to handle longer occlusion with the increase of complexity in computation. Numerous MOT methods directly utilize the first- order or the second-order independent motion models to locate objects (Bae and Yoon 2014) and associate accurately. Here we present a 3D Multi Object Tracking system which uses the detected object and in case of occlusion it utilizes the predicted track of objects. This real-time application uses online MOT with real-time as the motivation. The complete system has implemented with combining 2D tracklets from 2D Multi Object Tracking system and 3D tracklets from 3D Multi Object Tracking system. The dataset used is publicly available KITTI, datasets are captured by driving around the mid-size city of Karlsruhe, Germany (Figure 1 shows the setup of KITTI platform).

II. Methodology

Multi Object Tracking can be viewed as a combination of Object Detection, Propagating the detection using Motion Model, Data Association and Managing the Tracklets. The implemented MOT method has explained through these steps. Figure 3 gives the overall flow of the methodology, which uses image and point cloud detections as input. Tracker has shown using each step involved in its process.

Object Detection: The first and foremost step of Tracking by Detection methods are detecting the objects in each frame. The proposed 2D object detection uses visual based detection algorithm. For the better accuracy and fastness of the MOT system, we have chosen state-of-the-art detectors from official KITTI. The 2D detection uses Faster-RCNN results and 3D detection on LiDAR point cloud uses Point RCNN results.

The 2D Object Detector- Faster-RCNN in the MOT, uses images as input and output the best fit bounding box of the detected objects in each frame. The appearance features are ignored other than the detected bounding box for tracking to decrease the overall complexity. The position [x, y, w, h] and 2D bounding box's size [w, h] are used for processing. This paper exploit recent advances in visual object tracking to solve the problems of online tracking by detection method, rather than aiming to be robust to detection errors. Object detection algorithms have been improved remarkably in recent years. This results in improved object detectors, which helps to improve accuracy and complexity of tracking algorithm. The 3D detector-Point RCNN, runs detecting algorithm on LiDAR point cloud and gives the object location in 6 coordinates which are [x, y, z , w, h, l] where x, y , z are coordinates of center of 3D box. The position [x1, y1, x2, y2] and size of the 2D bounding box from 2D detector are used to calculate the center pixel of the object location. x1, y1, x2, y2 are extreme left, right, bottom and top coordinates. The tracking is done using center coordinates, width and height.

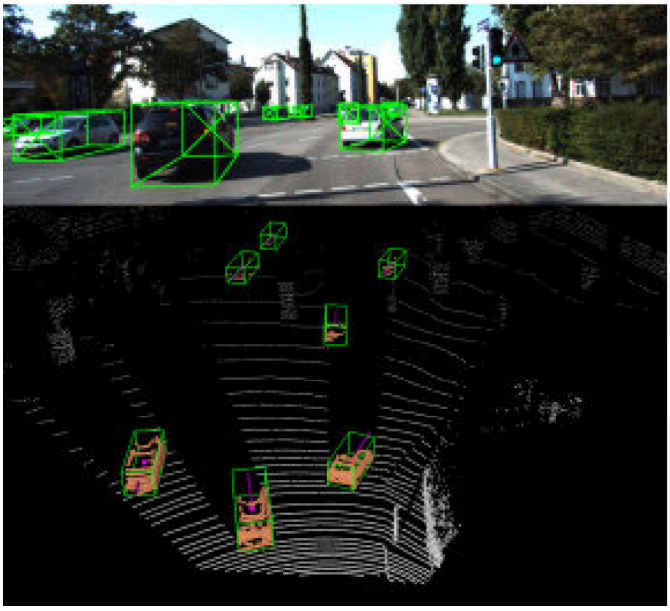

Fig. 2. Object Detection [19]

The detections are available for 3 categories of KITTI dataset - car, pedestrian and cyclist. We aim to concentrate on tracking scenario of car category.

Estimation Model: The estimation model is representation of object motion from frame to frame.The state estimation helps to come up with a likely position of object in the future frame, which reduces the search area, hence increases the accuracy of association. The popular motion models are categorized in to linear and non-linear motion models. The linear motion model follows a linear movement with constant velocity or constant turn rate. The non-linear model can represent a non linear model accurately than linear one.This works under the assumptions that the noise is Gaussian. The method has 2 state estimation model. One for the 2D detection in the image plane and the second is 3D detection in the point cloud.

The proposed 2D MOT system uses Particle Filter, which is a sampling based recursive Bayesian algorithm on image detections. Each of the particle is selected to represent a possible state of object. The filter assumes a uniform distribution of the particles at the start. This model represents

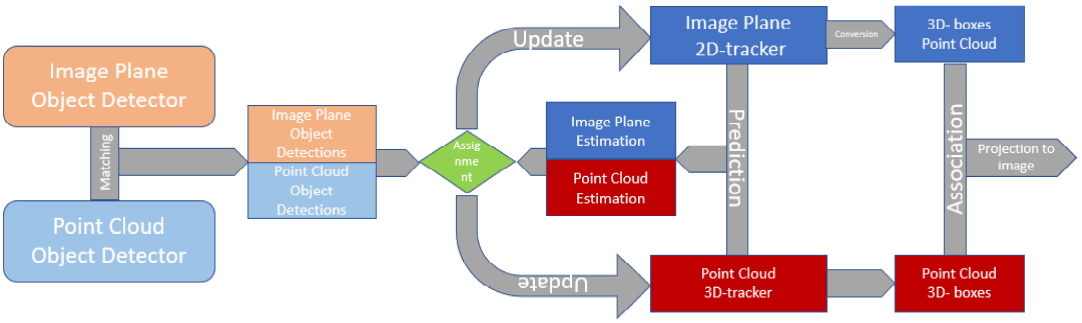

Fig. 3. Block Diagram of our Multi Object Tracking implementation.

the posterior density function using a set of random samples with associated weights. This method compute the estimate with these samples and the weights. The 3D locations have Kalman filter as state estimation model. The models predict and correct the future states and optimal possible state is found with smallest possible variance error. The state of each particle is modelled as [x, y, w, h, x_{vel}, y_{vel}, w_{vel}, h_{vel}], where x and y represent the horizontal and vertical pixel location of the centre of the target, while the scale w and h represent the width and height of the target's bounding box respectively. The 3D MOT tracking is done in complete 3D space. The state of the model is represented using [x, y, z, l, w, h, x_{vel}, y_{vel}, z_{vel}]. Prediction of tracker's filter gives a reduced region which helps decreasing the missed rate and reduced uncertainty of measured noise.

Data Association: The most crucial step of Multi-Target tracking is data association. This method is used to identify the resemblance between sensor measurements/detections and the pre-existing tracks. Incorrect assignment can decrease in accuracy remarkably. Generally multi-scan methods are recommended in situations where there are a lot of false alarms and missed detection. But delaying the association to include future information will negatively affect the use in real-time applications. So for associating tracklets Online trackers uses past and current frames and done in faster way possible.

If the sensor has acceptably high frame rates, detections of an object in successive frames possess great overlap IOU (intersection-over-union) [1].

$$IOU(a,b) = \frac{Area(a) \cap Area(b)}{Area(a) \cup Area(b)}$$

If the above assumptions are satisfied, tracking will be simpler and can be implemented even without using feature information. We use a simple IOU tracker which essentially continues a track by associating the detection with the highest IOU to the tracklets predicted from the previous frame.The 3D multi object tracking and 2D multi object tracking uses same theory for data association. The condition for measurement to be considered for association is σ_{IOU}. if the given threshold σ_{IOU} is met the measurement can be considered for association.The affinity matrix from IOU is solved using bipartite graph algorithm. The total computational complexity of the algorithm is made low compared to other trackers. As no visual information used in 2D data association, it will result in fast filtering procedure. Figure 4 shows how the data association allows to track object with same identity through out it's life in the video sequence.

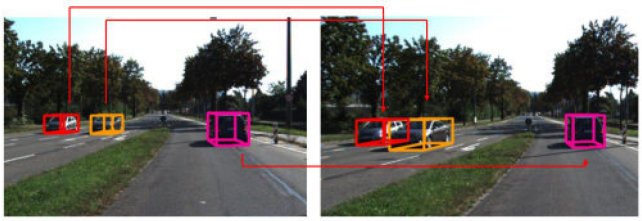

Fig. 4. IOU Data Association

Managing Track Identities: This stage is important in multi object tracking to make the overall system less complex. The tracklet management algorithm is inspired from SORT algorithm [9]. The implemented method has modifications applied for a lean implementation of Multi Object Tracking. The tracker has different modes and actions according to the result of it's data association. The modes of tracker are active, tracked state, inactive and lost. Each of these state has two actions corresponding to assignment decision.

When new objects enter the frame and leave, unique identities need to be created and destroyed accordingly. Any detection with overlap less than σ_{IOU} -(The minimum criteria for the association) is considered as an un-tracked object and created a new identity for it. The tracker is initialized with the bounding box parameters [x, y, w, h] and velocity set to zero. Tracks which are having missing measurements for N_{lost} frames are stopped. The same algorithm applies to both 3D tracklets and

2D tracklets. This will prevent the growth in number of tracks and accumulating error from the prediction with out having any detection to correct. The T_{lost} is set to 3 frames for object re-identification if available and did not make it a high value as this will increase the total tracks, thus computation of the tracks which might have left the frame already.

III. 2D to 3D Conversion

The proposed Multi object Tracking uses 3D boxes to associate between sensor datas. In this case the 2D tracks from camera, should be converted to 3D. The input of the model is mono-camera image and a 2D bounding box containing the target objects, the method outputs the 3D bounding box estimated for the targets. The architecture is inspired from the work of Mousavian et al. in [10] which utilizes a deep network and geometry for the calculation of 3D bounding box from one image. The input image is cropped to 2D bounding box, re-sized and fed to the ResNet-34 CNN that extract feature vector. This is then input to the fully connected network of 3 layers to output (2*8 + 3+ 1) values. Algorithm is conceptually simple, but this method outperforms complex and computationally expensive algorithms. The first 16 values in the output are the regressed residual for the pixel values of the projected eight corners of 3D bounding box on image plane. The next three values are regressed residuals for the size dimension of the 3D box. The last value is the regressed residual for the the distance of 3D box center from the camera center. The regressed residuals are used to find the 3D box parameters. This is estimated using geometry, by minimizing the difference between the regressed pixel coordinates and the ones obtained by projecting the estimated 3D box onto the image plane. The architecture is implemented in python using PyTorch. The proposed method used the ResNet-34 model as in the extended FrustumPointNet [5] architecture and the network training is done using the Adam optimizer. Later the fusion of estimated states from camera and LiDAR is done by probabilistic weighing method.

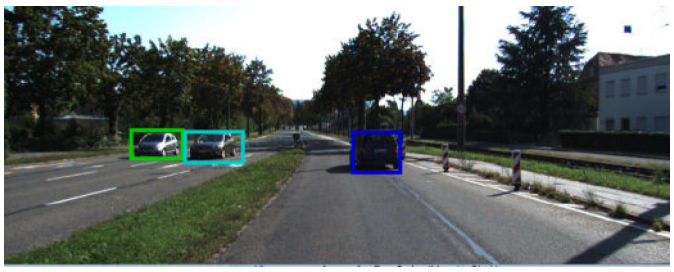

Fig. 5. Tracking in 2D. The tracklets with unique identity has given unique color of bounding box.

IV. Results

The Multi Object tracking methodology implemented should be evaluated on the following criteria to talk about its efficiency. It should detect all the objects and estimate the location of the objects in the all the frames as precisely as possible. It should also keep track of these objects over time. Each detected object is given unique ID which stays constant throughout the sequence in the tracking. The proposed method is tested on CLERAMOT [30] metrics. The proposed tracker method can run at 180 Hz (frames per second) on Intel i7 2.5GHz machine which is faster than many existing methods like Complexer-YOLO [11], BeyondPixels [6] on KITTI. Figure 5, 6, 7, 8 shows the visual results of tracking. Table 1 shows the quantitative results.

The accuracy of the tracker is discussed on:

MOTA(↑): Multi-object tracking accuracy

MOTP(↑): Multi-object tracking precision

MT(↑): number of mostly tracked trajectories. i.e. target has

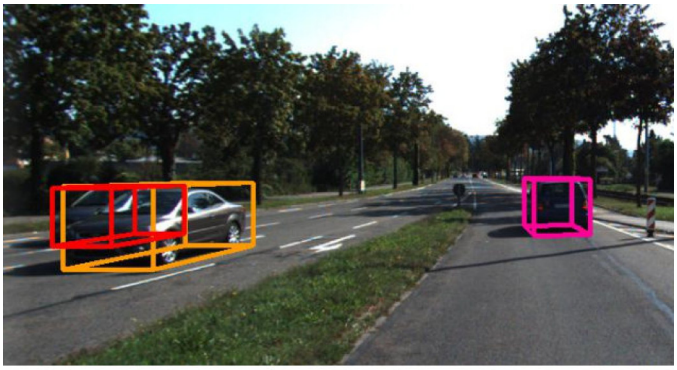

Fig. 6. The tracklet with occlusion due its orientation against ego vehicle has been tracked

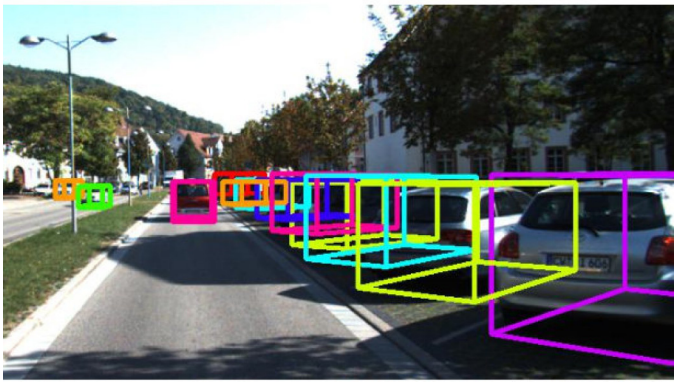

Fig. 7. Tracking done in cluttered scenario

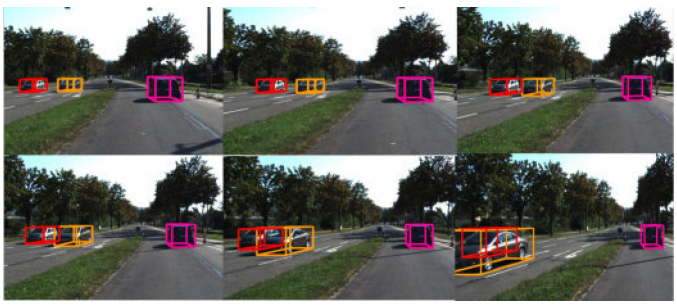

Fig. 8. The tracklets with unique identity has given unique color of bounding box

TABLE I
OUTPUT OF KITTI DATASET ON MOT TRACKER ARCHITECTURE

Method	MOTA	MOTP	MT	ML
Proposed Tracker- KITTI	84.62%	85.81%	74.1%	2.48%

the same label for at least 80% of its life span.
ML(↓): number of mostly lost trajectories. i.e. target is tracked for at most 20% of its life span.
Fusion of LiDAR data increased the efficiency of MOT compared to the MOT implemented [1] using mono-camera.

V. CONCLUSION

This paper proposes an approach to implement Multi-Target Tracking method by following efficient method for real time implementation. Most of the complex Multi Object Tracking methods achieve high efficiency at the cost of run time performance. But for an autonomous vehicle the real time processing is most critical. The proposed method considered this in every stage of its implementation and reduced the processing complexity of tracker algorithm. The proposed model used one camera and LiDAR for tracking.

REFERENCES

[1] Surya Kollazhi Manghat, Mohamed El-Sharkawy. "Forward Collision Prediction with Online Visual Tracking." 2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES). IEEE, 2019.

[2] Erik Bochinski, Volker Eiselein, Thomas Sikora. "High-Speed tracking-by-detection without using image information", 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2017

[3] Katare, Dewant, and Mohamed El-Sharkawy. "Autonomous Embedded System Enabled 3-D Object Detector:(with Point Cloud and Camera)." 2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES). IEEE, 2019.

[4] Catlin, Donald E. Estimation, control, and the discrete Kalman filter. Vol.71. Springer Science and Business Media, 2012.

[5] Qi, Charles R., et al. "Frustum pointnets for 3d object detection from rgb-d data." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[6] Sharma, Sarthak, et al. "Beyond pixels: Leveraging geometry and shape cues for online multi-object tracking." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.

[7] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.

[8] Zhang, Xinyu, et al. "Real-time vehicle detection and tracking using improved histogram of gradient features and Kalman filters." Proceedings of International Journal of Advanced Robotic Systems, 2018.

[9] Bewly, Alex, et al. "Simple Online and Realtime Tarcking." Proceedings of the IEEE International Conference on Image Processing (ICIP), 2017.

[10] Arsalan Mousavian, Dragomir Anguelov, John Flynn, Jana Kosecka. "3D Bounding Box Estimation Using Deep Learning and Geometry." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[11] M. Simon, K. Amende, A. Kraus, J. Honer, T. Samann, H. Kaulbersch, S. Milz, and H. M. ¨Gross. Complexer-YOLO: Real-Time 3D Object Detection and Tracking on Semantic Point Clouds. CVPRW, 2019.

[12] Yoon, Ju Hong, et al. "Structural Constraint Data Association for Online Multi-object Tracking." Proceedings of International Journal of Computer Vision. 2018.

[13] Faragher, Ramsey. "Understanding the basis of the Kalman filter via a simple and intuitive derivation." IEEE Signal processing magazine 29.5 (2012): 128-132.

[14] Wang , Li et al. "Evolving Boxes for Fast Vehicle Detection" IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 1135-1140.

[15] Girshik, Ross "Fast R-CNN"2015 IEEE International Conference on Computer Vision (ICCV) , 2015.

[16] Dongliang, Zheng,"Planning and Tracking in Image Space for Image-Based Visual Servoing of a Quadrotor" IEEE Transactions on Industrial Electronics, 2018.

[17] Boksuk, Shin. "Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities" Published in Journal Machine Vision and Applications archive Volume 29 Issue 1, January 2018 Pages 95-112

[18] Zhang. "Global Data Association for Multi-Object Tracking Using Network Flows" Published in IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[19] Shi, S., Wang, X., and Li, H., 2018. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. In Computer Vision and Pattern Recognition. arXiv preprint arXiv:1812.04244 [online] [https://arxiv.org/pdf/1812.04244.pdf] [Accessed on 10th November 2019].

[20] Voigtlaender, P., Krause, M., Osep, A., 2019. MOTS: Multi-Object Tracking and Segmentation. arXiv preprint arXiv:1902.03604v2. [online] [https://arxiv.org/pdf/1409.1556.pdf] [Accessed on 10th November 2019].

[21] Girdhar, Rohit. "Detect-and-track: Efficient pose estimation in videos." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[22] Wang, Qiang. "Fast online object tracking and segmentation: A unifying approach." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[23] Xu, Danfei, Dragomir Anguelov, and Ashesh Jain. "Pointfusion: Deep sensor fusion for 3d bounding box estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[24] Garcia, Fernando. "Sensor fusion methodology for vehicle detection." IEEE Intelligent Transportation Systems Magazine 9.1 (2017): 123-133.

[25] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[26] Venkitachalam, Sreeram. "Realtime Applications with RTMaps and Bluebox 2.0." Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2018.

[27] Qi, Charles R. "Frustum pointnets for 3d object detection from rgb-d data." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[28] Wen, Longyin. "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking." arXiv preprint arXiv:1511.04136 (2015).

[29] Chen, S. Y. "Kalman filter for robot vision: a survey." IEEE Transactions on Industrial Electronics 59.11 (2011): 4409-4420.

[30] Bernardin, Keni and Stiefelhagen, Rainer. "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics." EURASIP Journal on Image and Video Processing, Volume 2008, Article ID 246309, 10 pages.

[31] Bochinski, Erik, Tobias Senst, and Thomas Sikora. "Extending IOU based multi-object tracking by visual information." 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2018.