# Multi-object Tracking for Urban and Multilane Traffic: Building Blocks for Real-World Application

Nikolajs Bumanis[1][a], Gatis Vitols[2][b], Irina Arhipova[2][c] and Egons Solmanis[2]
[1]*Faculty of Information Technologies, Latvia University of Life Sciences and Technologies, 2 Liela str., Jelgava, Latvia*
[2]*WeAreDots Ltd., Elizabetes str. 75, Riga, Latvia*

Keywords: Multi-object Tracking, MOT Algorithms, Traffic Detection.

Abstract: Visual object detection and tracking is a fundamental research topic in computer vision with multiple applications ranging from object classification to multi-object tracking in heavy urban traffic scenarios. While object detection and tracking tasks, especially multi-object tracking, have multiple solutions, it is still unclear how to build the real-world applications using different building blocks like algorithms, filters, base neuron networks, etc. The issue becomes more sophisticated as most of the recently proposed solutions are based on existing methodologies, frameworks and applicable technologies; however, some are showing promising results using contradictory realization. This paper addresses issues and research trends of multi-object tracking, while depicting its building blocks and currently best solutions. In result, a potential building blocks for real-world application in the framework of Jelgava city in Latvia is presented.

## 1 INTRODUCTION

Visual object detection and tracking is becoming a common analysis tool for applications in the field of urban or mixed urban traffic (Jodoin et al., 2014; Ooi et al., 2018; Feng et al., 2019), pedestrian tracking (W. Luo et al., 2014; Barthélemy et al., 2019), video surveillance (Olszewska, 2016; Tsakanikas & Dagiuklas, 2018), illegal parking detection (Lee et al., 2009; Tang et al., 2020), illegal trespassing (G. Li et al., 2019) and other traffic anomalies (Ibadov et al., 2019).

While object detection aims to localize and classify object in a particular video frame, object tracking algorithms predict and associate observations across multiple frames, thus providing data for traffic analysis tasks. There are various surveys (Yilmaz et al., 2006; Ciaparrone et al., 2020) and literature reviews (W. Luo et al., 2017; Verma, 2017) addressing object detection and object tracking.

According to Ciaparronea et al. (Ciaparrone et al., 2020) multi-object tracking (MOT) is task of video analysis with the aim to track objects that belong to particular category with the assumption that there is no prior information about these objects, while Luo et al. (W. Luo et al., 2017) states that MOT is an estimation problem. Li et al. (X. Li et al., 2010) adds that compared to single object tracking (SOT), MOT focuses on the determination of the individual trajectories of multiple objects and that is a complex issue due to interactions among multiple objects.

The MOT algorithm is build using computer vision (CV) blocks that perform object detection, object classification and object tracking. While there are go-to principles of choosing these blocks, novel and contradictory ideas are still emerging. These ideas often lead to new MOT algorithms outperforming existing solutions (Mykheievskyi et al., 2020).

The aim of this paper is to identify current trends of MOT and determine appropriate CV building blocks for application in road traffic scenarios, including surveillance and planning.

This paper is structured as follows: section 2 covers multi-object tracking process; section 3 includes video pre-processing, object detection, feature extraction and object tracking, while simultaneously focusing on challenges associated

[a] https://orcid.org/0000-0002-1884-7731
[b] https://orcid.org/0000-0002-4131-8635
[c] https://orcid.org/0000-0003-1036-2024

Bumanis, N., Vitols, G., Arhipova, I. and Solmanis, E.
Multi-object Tracking for Urban and Multilane Traffic: Building Blocks for Real-World Application.
DOI: 10.5220/0010467807290736
In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 1*, pages 729-736
ISBN: 978-989-758-509-8

with these steps; section 4 briefly introduces MOT benchmarks, MOT solutions developed and proposed in the framework of last four (2017-2020) years addressing urban and/or multi-lane traffic and addresses MOT solution building block for Jelgava city urban traffic tracking.

## 2 MULTI-OBJECT TRACKING

MOT algorithms can be classified (W. Luo et al., 2017; T. Yang et al., 2017) from the perspective of initialization method, processing mode and type of output.

There are currently two strategies to approach object initialization - tracking based on detection, commonly referred to as “tracking-by-detection” or “association based tracking” and Detection free tracking, alternatively referred to as “category free tracking” (Fragkiadaki & Shi, 2011; W. Luo et al., 2017; T. Yang et al., 2019; B. Yang et al., 2020).

Tracking-by-detection approach firstly uses an object detector to obtain the classification hypothesis of each object per frame and then connects successful hypotheses into trajectories. This object detector must be pre-trained, thus limiting potential classes the algorithm can detect and track (Ciaparrone et al., 2020). However, this process is typically performed automatically. Conversely, Detection-free tracking requires manual initialization of all objects in the first frames (Fragkiadaki & Shi, 2011), thus limiting the amount of objects being tracked. Detection-free tracking has found its niche in sports related tracking like basketball (Fragkiadaki & Shi, 2011; Lin et al., 2015), while the Tracking-by-detection seems to be a go-to (Milan et al., 2013; Ciaparrone et al., 2020; L. Liu et al., 2020; B. Yang et al., 2020) approach for real-time object tracking.

The processing mode refers to object detection, feature extraction and object tracking, respectively. There are two processing modes defined - online tracking and offline tracking. According to Luo et al. (W. Luo et al., 2017), these methods differ in the aspect of processing the observation from future frames. Online method processes video frame by frame, using past and up-to-date frames to build trajectories on the fly, and offline method uses batch approach to process past, up-to-date and future frames in order to find optimal trajectory.

To relate this to intelligent transportation applications, both of the methods can be used - online tracking to monitor the current movement of the cars and pedestrians, and offline to optimize traffic in the framework of a city or its part. Realizing both of these methods come to resolving ambiguities in associating object detections (Xiang et al., 2015). In case of real time application of multi-object tracking in Jelgava, it is now clear that Online Tracking-by-detection approach must be selected.

From the output perspective, MOT algorithms are classified into deterministic and probabilistic (W. Luo et al., 2017).

## 3 MULTI-OBJECTS TRACKING IN ACTION

The general steps of multi-object tracking are the detection of all objects in the frame and creating associations between detected objects across multiple joined frames. In addition, it is often required to prepare raw video for accurate processing. This includes video pre-processing, object segmentation and object occlusion handling (Joy & Vijaya Kumar, 2018).

### 3.1 Video Pre-processing

Video pre-processing is performed due to issues as object occlusion, illumination, complexity of background and image quality parameters like resolution, noise, blur, etc. Pre-processing includes de-noising, stabilization and enhancement for image quality and illumination issue reduction (Balasubramanian et al., 2014).

One of the approaches to solving these issues is key frame extraction (Zheng et al., 2015; Joy & Vijaya Kumar, 2018) that aims to extract only those frames that differ in terms of features provided. For a smart city traffic surveillance problem this can be done using Frame-Per-Warp method (Zheng et al., 2015). However, key frame extraction relates more to SOT, rather than MOT. Depending on video resolution the video frames may contain a lot of imperfections in the form of noise that refers to the random dot pattern that is superimposed on the picture. The source of that noise is fog, rains, dust, unrealistic edges, corners, invisible lines and blurred objects. The noise leads to reduction of image quality; therefore, reducing the likelihood of a correct object detection (Halkarnikar et al., 2010). There are various noise reduction and removal algorithms such as VBM3D algorithm (Balasubramanian et al., 2014; Ehret & Arias, 2020) or Adaptive Spatial-Temporal Filtering proposed by Yahya et al. (Yahya et al., 2015). Video can also be enhanced by mathematical manipulation affecting each separate pixel’s colour, hue, saturation, etc., by using Particle Swarm Optimization (Seixas Gomes de Almeida & Coppo Leite, 2019).

It can be concluded that while the issues are still common, various solutions are already ready to be implemented as separate CV blocks.

### 3.2 Multiple Object Detection

Object Detector in the framework of object tracking is the same detector used to classify objects in typical classification problems. Both SOT and MOT have multiple challenges to address: scale changes, occlusions, object similarities, initialization and termination of trajectories.

There is a requirement for a high real-time performance and accuracy for traffic object detection. Traditional CV approaches are slow, and cannot compare to modern deep learning methods. There are single and two stage deep learning methods. Two stage methods such as R-CNN (Girshick, 2015; S. Ren et al., 2017) family methods, firstly extract multiple different size feature regions with proposals that are then classified by CNN using various sequences depending on the particular method's architecture. For optimization purposes bounding box regression models (Y. He et al., 2019) and SoftMax estimation are typically used.

It is common knowledge that single stage methods like YOLO (Redmon et al., 2016) and SSD (W. Liu et al., 2016) are faster, but lack accuracy. This is also proven in our previous research addressing person re-identification (Arhipova et al., 2020; Bumanis et al., 2020). Researchers also propose (Soviany & Ionescu, 2018; L. Liu et al., 2020) the use of both single and two-stage methods by employing image difficulty estimation prior to object detection (Ionescu et al., 2017), resulting in the choice of either fast single-stage method for simple images or slower two-stage method for complex images.

One of the issues of CNN-based algorithms is the scaling problem. Traffic related objects such as buses, cars and bicycles have different scales thus introducing challenges to neural network training. One of the solutions is increasing neural network depth and training neural network using different resolution images (Hu et al., 2019); however it may lead to higher training errors (Simonyan & Zisserman, 2014) due to accuracy saturation.

To address the scaling problem Hu et al. (Hu et al., 2019) proposed an advanced Region of Interest (RoI) pooling method, based on multi-branch decision network, that can be used with existing object detection architectures and aims to feature map vehicles with small scales. Another approach, proposed by Feng et al.  (Feng et al., 2019) is a 32-layer multibranch network that divides the typical object detection sequence into three different branches.

It can be concluded that modern (Feng et al., 2019; Hu et al., 2019; K. He et al., 2020) object detection algorithms provide sufficient performance when it comes to object and image scale issues.

### 3.3 Segmentation and Object Occlusion Challenge

Object segmentation may be considered as extension of object detection. One of the challenges segmentations comes across is object occlusion.

Occlusion is typically mentioned in combination with bounding boxes, when a bigger bounding box overlaps with a bounding box of a smaller object thus creating a so-called blob, the same goes for object shapes in case of object segmentation. There are three occlusion types – self occlusion, scene occlusion and object-over-object occlusion. The first one is more applicable to person related detection and tracking problems, whereas the other two can be applied to urban and dense traffic.

There are various ways to address this issue, for example Benamara et al. (Benamara et al., 2016) followed the merge-split approach that freezes objects as soon as occlusion is detected, creates an object group consisting of frozen objects and determines the step - either to split or merge objects.

Alternatively, Yang et al. (T. Yang et al., 2005) divides detected objects into three states - before, during and after occlusion. In order to estimate a background model Yang et al. applied a two-level pixel motion analysing algorithm and a pairwise mechanism to merge or split objects depending on object detection and segmentation results. The splitting was done by employing corresponding feature modules to assign labels to split objects based on Kullback-Leibler distance.

To reduce the occlusion (Yu et al., 2016) used the Reliability of Tracklets for Particle Based multi-object tracking framework.

### 3.4 Feature Extraction

To track a particular object, this object must be identified across multiple frames correctly. Firstly, the similarity between objects in one frame must be measured and secondly the identity information based on measured similarity must be recovered. According to Luo et al. (W. Luo et al., 2017) classification, the former refers to the modelling visual features, while the latter addresses the interpretation. The main goal is to be able to separate closely positioned small objects (Beaupré et al., 2018).

Modelling of appearance typically refers to visual characteristics and statistical measurements. Visual characteristics can be depicted using features, such as colour, shape, edge, position colour histogram, histogram oriented gradients that are based on single or multiple cues (Badal et al., 2018; W. Li et al., 2019); all in order to help distinguish between objects effectively when they are similar (W. Li et al., 2019).

It is also possible to fuse information from multiple different video sources that provide, for example, typical RGB images (C. Y. Ren et al., 2017), infra-red images (Zhang et al., 2020), or LiDAR point clouds (Muro et al., 2019; Rangesh & Trivedi, 2019). When it comes to visual appearance the methods focus on either local and regional features or depth estimation (Mauri et al., 2020). One of the examples of the former is utilization of optical flow that performs linking detection responses into short tracklets before data association (T. Yang et al., 2019), like it was done in (Beaupré et al., 2018) for urban traffic. The latter, regional features, are common for object detection algorithms and are based on bounding boxes (Y. He et al., 2019; L. Liu et al., 2020). Luo et al. (W. Luo et al., 2017) defines three levels of regional features - zero order, i.e. colour histograms, first order, i.e. gradient-based histogram representations, and second order, i.e. region covariance matrix.

The choice of appearance modelling comes to the needs of the final tracking system - if the efficiency and low computation cost is priority then local feature-based methods are better, while for occlusion and illumination issue reduction region covariance matrix should be used.

### 3.5 Tracking

There are various algorithms that can be used to track objects; some of them (X. Li et al., 2010; Kulkarni & Rani, 2018) use original Kalman filter, but relatively newer ones (Y. M. Song et al., 2019; Y. Song & Jeon, 2020) utilize Gaussian Mixture Probability Hypothesis Density (GMPHD) filter. The Kalman filter in combination with Iterative-Hungarian Algorithm was utilized in work by Bima and Adiprawita (Sahbani & Adiprawita, 2017).

The lifetime of a tracked object can be modelled according to the Markov Decision Process (Xiang et al., 2015; T. Yang et al., 2017, 2019).

When an object is detected it enters the initial stage of an “Active” state. The object can then transition from this state to either “Tracked” or “Inactive” depending on the object detector result. According to definition by Xiang et al., it's ideally when a true positive from an object detector transitions the object into “Tracked” state, while at the same time false alarm should be a reason to transition into “Inactive” state.

While in “Tracked” state the same object is being continuously tracked until it transitions to state “Lost”. If an object is lost it can either stay in this state forever, move to the state “Inactive” with particular time passing, or be re-detected and become “Tracked” again. The objects cannot be “re-detected” from “Inactive” state, instead they are considered to be new objects.

## 4 MULTI-OBJECT TRACKING COMPONENTS FOR REAL-WORLD APPLICATION IN PROCESS OF URBAN TRAFFIC MONITORING

When it comes to comparing various MOT trackers, MOT specific metrics must be used. These metrics are applied to publicly available data sets with fixed resolution, number of frames, number of objects (i.e. sparse or dense traffic).

Most popular MOT trackers benchmarks are KITTI (*The KITTI Vision Benchmark Suite*) and MOTChallenge (Leal-Taixé et al., 2015; *MOT Challenge*). There is also a community project by computer vision developers called Papers with Code (*The Latest in Machine Learning | Papers With Code*, 2021). Some specific cases like tracking from drones may benefit from benchmarks like large scale variance highway dataset LSVH (Krajewski et al., 2018). These benchmarks give access to grouped data sets to test problem specific MOT tracker solutions, and while MOTChallenge focuses more on pedestrian tracking, KITTI provides both pedestrian and vehicle data sets. For example, KITTI provides separate data sets for car tracking and road lane detection and tracking; what’s especially alluring is possibilities to choose different road types, i.e., urban marked road, urban unmarked road, urban multiple marked lanes or even combination of the three, thus allowing to test developed tracking solutions to specific regional scenario (small city, large city, highways, etc). These benchmarks are based on CLEAR MOT metrics (Bernardin & Stiefelhagen, 2008), whereas KITTI benchmark also applies classical pixel-based metrics to evaluate methods (for example, road detection) that output confidence maps.

## 4.1 Benchmark Leaders

The following section describes the current (for 9th of December, 2020) KITTI and/or MOTChallenge benchmark leaders. Based on the KITTI metrics for urban road detection, currently the best performing algorithm for single-lane urban roads and second best for multilane urban roads is PLARD by Chen et al. (Chen et al., 2019). The PLARD method has the MaxF score of 97.05% for single-lane and 97.77% for multi-lane urban roads accordingly.

As for car tracking, based on results depicted in KITTI benchmark's leader boards, the best method is SRK_ODESA proposed by Mykheievskyi et al. (Mykheievskyi et al., 2020). SRK_ODESA has the MOTA score of 90.03%. Instead of a classical DBT embedding function authors used Learned Local Features Descriptors (LLFD). According to Mykheievskyi et al. (Mykheievskyi et al., 2020) the difference is that the objective of typical DBT embedding function is to produce compact and separable object manifolds, when the loss formulation is adopted from metric learning. The same loss combined with a different sampling approach in the case of LLFD is expected to result in extended and non-separable manifolds.

The second-best car tracking tracker in KITTI benchmark's leader boards is the online type tracker called CenterTrack proposed by Zhou et al. (Zhou et al., 2020) that, similarly to SRK_ODESA, CenterTrack does not follow the typical pipeline of tracking-by-detecting, and instead focuses utilizes object detection algorithm on an actual image and previously acquired detection as a pair. CenterTrack has the MOTA score of 67.3% using MOT17 data set and the MOTA score of 89.4% using KITTI tracking benchmark.

## 4.2 Urban Traffic Solutions

There are various methods and algorithms addressing urban traffic scenarios. Some of them are:

- Mixed urban traffic multi-object tracking method proposed by Yang et al. (Y. Yang & Bilodeau, 2016) that is based on background separation using Kernelized Correlation Filter (KCF) tracker. The method achieved the MOTA score of 85.5% for car tracking using Rene-Levesque data set (1280x720px, 1000 frames) and 82.5% for mixed traffic tracking using St-Marc data set (1280x720px, 1000 frames).
- Multi-object tracking algorithm for urban traffic scenes proposed by Ooi et al. (Ooi et al., 2018) utilizes Region-based Fully Convolutional Network (RFCN) for object detection that was trained and refined using the MIO-TCD data set (Z. Luo et al., 2018). Object association was realized using the Hungarian algorithm. Algorithm achieved the best MOTP score of 74.88% using the labelled Sherbrooke data set, and the MOTP score of 68.93% using the unlabelled Rouen data set.
- Urban traffic monitoring was proposed by Arinaldi et al. (Arinaldi et al., 2018). The method is based on approach of using region based convolutional neural networks, Faster RCNN respectively, while the speed of cars is calculated by performing short term tracking using the KCF based tracker.
- Multi-object tracking using point clouds provided by LiDAR technology was proposed by Sualeh et al. (Sualeh & Kim, 2019). The method is based on object identification performed by SSD algorithm together with DBSCAN algorithm for silhouette analysis based on LiDAR point cloud. The method was evaluated using KITTI's LiDAR data set, and achieved a result of the MOTA score around 90% for different scenarios.

## 4.3 Urban Traffic Solution for Jelgava

Jelgava is a small urban city with around 60k population with urban roads up to 6 lanes (3 per direction). The traffic in Jelgava can be both - sparse (during the day and evening) and dense (rush hour, commute hours). The MOT solution for Jelgava city must be built in accordance with following parameters: relatively low number of cars (30 per frame), slow moving speeds (0-60 km/h).

Based on the analysis of MOT process and existing solutions, the following MOT solution building blocks can be identified: a) MOT framework – Tracking-by-detection or Detection-Free tracking, b) tracking processing mode - Online or Batch (offline), c) Object detector (may include data association and feature extraction), e) object tracker. Based on the analysis the following building blocks are considered to be appropriate for Jelgava city urban traffic tracking:

a) Tracking-by-detection approach;
b) Online processing mode for real-time data analysis and Offline processing mode for traffic flow analysis;
c) One-step deep learning detector for Online processing mode (for example, YOLO) and

Two-step deep learning detector for Offline processing mode (for example, Faster R-CNN).

d) Object tracker based on Kalman filter and Hungarian algorithm.

# 5 CONCLUSIONS

Novel multi-object tracking algorithms rarely focus on improving object detection step, feature association and occlusion processing altogether; instead, they propose a novel niche for one step while simultaneously using already known and proven blocks for other steps. Based on analysis the best object detectors in respect to precision are two-step deep learning methods, while one-step deep learning methods excel in processing speed. It was also concluded that novel multi-object tracking solutions tend to improve already existing solutions by mixing different previously alternative approaches, i.e., creating so called hybrid approaches. The solution for Jelgava city urban traffic scenario can be build using existing multi-object tracking solutions; however, any sophisticated scenario will require in-depth analysis of existing and development of a new multi-object tracking solutions. Future work includes developing task specific multi-object tracking solution for Jelgava city.

# ACKNOWLEDGEMENTS

The research leading to these results has received funding from the project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/18/A/003 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.12 "Multi-object detection and tracking in vehicle traffic surveillance: fusing 3D-LiDAR and camera data".

# REFERENCES

Arhipova, I., Vitols, G., & Meirane, I., 2020. Long Period Re-identification Approach to Improving the Quality of Education: A Preliminary Study. *Advances in Intelligent Systems and Computing*, *1130 AISC*, 157–168.

Arinaldi, A., Pradana, J. A., & Gurusinga, A. A., 2018. Detection and classification of vehicles for traffic video analytics. *Procedia Computer Science*, *144*, 259–268.

Badal, T., Nain, N., & Ahmed, M., 2018. Online multi-object tracking: multiple instance based target appearance model. *Multimedia Tools and Applications*, *77*(19), 25199–25221.

Balasubramanian, A., Kamate, S., & Yilmazer, N., 2014. Utilization of robust video processing techniques to aid efficient object detection and tracking. *Procedia Computer Science*, *36*(C), 579–586.

Barthélemy, J., Verstaevel, N., Forehead, H., & Perez, P., 2019. Edge-Computing Video Analytics for Real-Time Traffic Monitoring in a Smart City. *Sensors*, *19*(9), 2048.

Beaupré, D.-A., Bilodeau, G.-A., & Saunier, N., 2018. Improving Multiple Object Tracking with Optical Flow and Edge Preprocessing. *ArXiv*, http://arxiv.org/abs/1801.09646.

Benamara, A., Miguet, S., & Scuturici, M., 2016. Real-time multi-object tracking with occlusion and stationary objects handling for conveying systems. *Lecture Notes in Computer Science*, *10072 LNCS*, 136–145.

Bernardin, K., & Stiefelhagen, R., 2008. Evaluating multiple object tracking performance: The CLEAR MOT metrics. *Eurasip Journal on Image and Video Processing*, *2008*.

Bilinski, P., Bremond, F., & Kaaniche, M. B., 2009. Multiple object tracking with occlusions using HOG descriptors and multi resolution images. *IET Seminar Digest*, *2009*(2).

Bumanis, N., Vitols, G., Arhipova, I., & Meirane, I., 2020. Deep learning solution for children long-term identification. *Research for Rural Development 2020: Annual 26th International Scientific Conference Proceedings*, *35*, 268–273.

Chen, Z., Zhang, J., & Tao, D., 2019. Progressive LiDAR adaptation for road detection. *IEEE/CAA Journal of Automatica Sinica*, *6*(3), 693–702.

Ciaparrone, G., Luque Sánchez, F., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F., 2020. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, *381*, 61–88.

Ehret, T., & Arias, P., 2020. Implementation of the VBM3D Video Denoising Method and Some Variants. *ArXiv*, *http://arxiv.org/abs/2001.01802.*

Feng, J., Wang, F., Feng, S., & Peng, Y., 2019. A Multibranch Object Detection Method for Traffic Scenes. *Computational Intelligence and Neuroscience*, *2019*.

Fragkiadaki, K., & Shi, J., 2011. Detection free tracking: Exploiting motion and topology for segmenting and tracking under entanglement. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2073–2080.

Girshick, R., 2015. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1440–1448.

Halkarnikar, P. P., Khandagle, H. P., Talbar, S. N., & Vasambekar, P. N., 2010. Object detection under noisy condition. *AIP Conference Proceedings*, *1324*(1), 288–290.

He, K., Gkioxari, G., Dollár, P., & Girshick, R., 2020. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397.

He, Y., Zhu, C., Wang, J., Savvides, M., & Zhang, X., 2019. Bounding box regression with uncertainty for accurate object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 2883–2892.

Hu, X., Xu, X., Xiao, Y., Chen, H., He, S., Qin, J., & Heng, P. A., 2019. SINet: A Scale-Insensitive Convolutional Neural Network for Fast Vehicle Detection. *IEEE Transactions on Intelligent Transportation Systems*, *20*(3), 1010–1019.

Ibadov, S. R., Kalmykov, B. Y., Ibadov, R. R., & Sizyakin, R. A., 2019. Method of Automated Detection of Traffic Violation with a Convolutional Neural Network. *EPJ Web of Conferences*, *224*, 04004.

Ionescu, R. T., Alexe, B., Leordeanu, M., Popescu, M., Papadopoulos, D. P., & Ferrari, V., 2017. How hard can it be? Estimating the difficulty of visual search in an image. *CoRR*, *arXiv,* http://arxiv.org/abs/1705.08280.

Jodoin, J. P., Bilodeau, G. A., & Saunier, N., 2014. Urban Tracker: Multiple object tracking in urban mixed traffic. *2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014*, 885–892.

Joy, F., & Vijaya Kumar, V., 2018. A Review on Multiple Object Detection and Tracking in Smart City Video Analytics. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(2S2), 2278–3075.

Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L., 2018. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-November*, 2118–2125.

Kulkarni A., Rani E., 2018. KALMAN Filter Based Multiple Object Tracking System. *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development*, *8*(2), 1–6.

Leal-Taixé, L., Milan, A., Reid, I., Roth, S., & Schindler, K., 2015. *MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking*. arXiv, http://arxiv.org/abs/1504.01942.

Lee, J. T., Ryoo, M. S., Riley, M., & Aggarwal, J. K., 2009. Real-time illegal parking detection in outdoor environments using 1-D transformation. *IEEE Transactions on Circuits and Systems for Video Technology*, *19*(7), 1014–1024.

Li, G., Song, H., Liao, Z., & Deng, K., 2019. An Effective Algorithm for Video-Based Parking and Drop Event Detection. *Complexity*, *2019*.

Li, W., Mu, J., & Liu, G., 2019. Multiple Object Tracking with Motion and Appearance Cues. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 161–169.

Li, X., Wang, K., Wang, W., & Li, Y., 2010. A multiple object tracking method using Kalman filter. *2010 IEEE International Conference on Information and Automation, ICIA 2010*, 1862–1866.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M., 2020. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, *128*(2), 261–318.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. Y., & Berg, A. C., 2016. SSD: Single shot multibox detector. *Lecture Notes in Computer Science*, *9905 LNCS*, 21–37.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., & Kim, T.-K., 2014. *Multiple Object Tracking: A Literature Review*. *V*(212), 1–18. arXiv, http://arxiv.org/abs/1409.7618.

Luo, Z., Branchaud-Charron, F., Lemaire, C., Konrad, J., Li, S., Mishra, A., Achkar, A., Eichel, J., & Jodoin, P. M., 2018. MIO-TCD: A New Benchmark Dataset for Vehicle Classification and Localization. *IEEE Transactions on Image Processing*, *27*(10), 5129–5141.

Mauri, A., Khemmar, R., Decoux, B., Ragot, N., Rossi, R., Trabelsi, R., Boutteau, R., Ertaud, J. Y., & Savatier, X., 2020. Deep learning for real-time 3D multi-object detection, localisation, and tracking: Application to smart mobility. *Sensors*, *20*(2).

Milan, A., Schindler, K., & Roth, S., 2013. Challenges of ground truth evaluation of multi-target tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 735–742.

*MOT Challenge*. Retrieved December 7, 2020, from https://motchallenge.net/

Muro, S., Matsui, Y., Hashimoto, M., & Takahashi, K., 2019. Moving-object tracking with lidar mounted on two-wheeled vehicle. *In proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO), 2,* 453–459.

Mykheievskyi, D., Borysenko, D., & Porokhonskyy, V., 2020. Learning Local Feature Descriptors for Multiple Object Tracking. *Proceedings of the Asian Conference on Computer Vision (ACCV)*.

Olszewska, J. I., 2016. Tracking the invisible man: Hidden-object detection for complex visual scene understanding. *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART), 2,* 223–229.

Ooi, H.-L., Bilodeau, G.-A., Saunier, N., & Beaupré, D.-A., 2018. Multiple Object Tracking in Urban Traffic Scenes with a Multiclass Object Detector. *Lecture Notes in Computer Science*, *11241 LNCS*, 727–736.

Rangesh, A., & Trivedi, M. M., 2019. No Blind Spots: Full-Surround Multi-Object Tracking for Autonomous Vehicles Using Cameras and LiDARs. *IEEE Transactions on Intelligent Vehicles*, *4*(4), 588–599.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 779–788.

Ren, C. Y., Prisacariu, V. A., Kähler, O., Reid, I. D., & Murray, D. W., 2017. Real-Time Tracking of Single and Multiple Objects from Depth-Colour Imagery Using 3D Signed Distance Functions. *International Journal of Computer Vision*, *124*(1), 80–95.

Ren, S., He, K., Girshick, R., & Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 658–666.

Sahbani, B., & Adiprawita, W., 2017. *Kalman filter and Iterative-Hungarian Algorithm implementation for low complexity point tracking as part of fast multiple object tracking system*. 109–115.

Seixas Gomes de Almeida, B., & Coppo Leite, V., 2019. Particle Swarm Optimization: A Powerful Technique for Solving Engineering Problems. In *Swarm Intelligence - Recent Advances, New Perspectives and Applications*. IntechOpen.

Simonyan, K., & Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

Song, Y. M., Yoon, K., Yoon, Y. C., Yow, K. C., & Jeon, M., 2019. Online Multi-Object Tracking with GMPHD Filter and Occlusion Group Management. *IEEE Access*, *7*, 165103–165121.

Song, Y., & Jeon, M., 2020. *Online Multi-Object Tracking and Segmentation with GMPHD Filter and Simple Affinity Fusion*. arXiv, http://arxiv.org/abs/2009.00100

Soviany, P., & Ionescu, R. T., 2018. Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction. *Proceedings - 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2018*, 209–214.

Sualeh, M., & Kim, G.-W., 2019. Dynamic Multi-LiDAR Based Multiple Object Detection and Tracking. *Sensors*, *19*(6), 1474.

Tang, H., Peng, A., Zhang, D., Liu, T., & Ouyang, J., 2020. SSD real-time illegal parking detection based on contextual information transmission. *Computers, Materials and Continua*, *62*(1), 293–307.

*The KITTI Vision Benchmark Suite*. Retrieved December 7, 2020, from http://www.cvlibs.net/datasets/kitti/.

*The latest in Machine Learning | Papers With Code*, Retrieved March 3, 2021 from https://paperswithcode.com/

Tsakanikas, V., & Dagiuklas, T., 2018. Video surveillance systems-current status and future trends. *Computers and Electrical Engineering*, *70*, 736–753.

Unaldi, N., Arigela, S., Asari, V. K., & Rahman, Z., 2008. Nonlinear technique for the enhancement of extremely high contrast images. *Visual Information Processing XVII*, *6978*, 697803.

Verma, R., 2017. A Review of Object Detection and Tracking Methods. *International Journal of Advance Engineering and Research Development*, *4*(10).

Xiang, Y., Alahi, A., & Savarese, S., 2015. Learning to Track: Online Multi-object Tracking by Decision Making. *2015 IEEE International Conference on Computer Vision (ICCV)*, 4705–4713.

Xu, Q., Chavez, A. G., Bulow, H., Birk, A., & Schwertfeger, S., 2019. Improved Fourier Mellin Invariant for Robust Rotation Estimation with Omni-Cameras. *Proceedings - International Conference on Image Processing, ICIP*, *2019-September*, 320–324.

Yahya, A. A., Tan, J., & Li, L., 2015. Video Noise Reduction Method Using Adaptive Spatial-Temporal Filtering. *Discrete Dynamics in Nature and Society*, *2015*.

Yang, B., Tang, M., Chen, S., Wang, G., Tan, Y., & Li, B., 2020. A vehicle tracking algorithm combining detector and tracker. *Eurasip Journal on Image and Video Processing*, *2020*(1), 1–20.

Yang, T., Cappelle, C., Ruichek, Y., & El Bagdouri, M., 2017. Multi-object tracking using compressive sensing features in markov decision process. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10617 LNCS*, 505–517.

Yang, T., Cappelle, C., Ruichek, Y., & El Bagdouri, M., 2019. Online multi-object tracking combining optical flow and compressive tracking in Markov decision process. *Journal of Visual Communication and Image Representation*, *58*, 178–186.

Yang, T., Li, S. Z., Pan, Q., & Li, J., 2005. Real-time multiple objects tracking with occlusion handling in dynamic scenes. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, *I*, 970–975.

Yang, Y., & Bilodeau, G.-A., 2016. Multiple Object Tracking with Kernelized Correlation Filters in Urban Mixed Traffic. *Proceedings - 2017 14th Conference on Computer and Robot Vision, CRV 2017*, *2018-January*, 209–216.

Yilmaz, A., Javed, O., & Shah, M., 2006. Object tracking: A survey. In *ACM Computing Surveys* (Vol. 38, Issue 4).

Yu, R., Zhu, B., Li, W., & Kong, X., 2016. A particle filter based multi-person tracking with occlusion handling. *ICINCO 2016 - Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics, 2(127)*, 201–207.

Yu, F., Li, W., Li, Q., Liu, Y., Shi, X., & Yan, J., 2016. POI: Multiple object tracking with high performance detection and appearance feature. *Lecture Notes in Computer Science*, *9914 LNCS*, 36–42.

Zhang, X., Ye, P., Leung, H., Gong, K., & Xiao, G., 2020. Object fusion tracking based on visible and infrared images: A comprehensive review. *Information Fusion*, *63*, 166–187.

Zheng, R., Yao, C., Jin, H., Zhu, L., Zhang, Q., & Deng, W., 2015. Parallel Key Frame Extraction for Surveillance Video Service in a Smart City. *PLOS ONE*, *10*(8), e0135694.

Zhou, X., Koltun, V., Krähenbühl, P., Austin, U. T., & Labs, I., 2020. Tracking Objects as Points. *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12349*, 474-490.