**INTRODUCTION**

Object detection technology has been successfully applied to find the size and position of target objects appearing on images or videos. Several applications have appeared mainly in self-driving of vehicles, CCTV monitoring and security system, cancer detection, etc. Object tracking is another area in image processing to be achieved by unique identification and tracking the positions of identified objects over time. However, to track objects, it is necessary to define object class and position first in a firstly given static image by object detection. Therefore, it can be said that the results of object tracking should be deeply dependent on the performance of the object detection involved. This object tracking technology has been successfully utilized for tracing of targeted pedestrian and the moving vehicle, accident monitoring in traffic camera, criminal and security monitoring in the certain local area of concern, etc. In the traffic control field, a case study on analysis and control of traffic conditions by automatic object detection has carried out in this paper. The summaries are given as follows. According to, an on-road vehicle detection system for the self-driving car was developed. This system detects vehicle object and classifies the type of vehicle by Convolutional Neural Network (CNN). The vehicle object tracking algorithm tracks the vehicle object by changing the tracking center point according to the position of the recognized vehicle object on the image. Then, the monitor shows a localized image like a bird’s viewpoint with the visualized vehicle objects, and the system calculates the distance between the driving car and the visualized vehicle objects. This process of the system enables to objectively view the position of the vehicle object so that it can help assistance of the self-driving system. As a result, it can localize the vehicle object in vertical 1.5m, horizontal 0.4m tolerance at the camera. In, another deep learning-based detection system in combination with CNN and Support Vector Machine (SVM) was developed to monitor moving vehicles on urban roads or highways by satellite. This system extracts the feature from the satellite image through CNN using the satellite image as an input value and performs the binary classification with SVM to detect the vehicle BBox. Besides, Arinaldi, Pradana, and Gurusinga developed a system to estimate the speed of the vehicle, classify vehicle type, and analyze traffic volume. This system utilizes BBox obtained by object detection based on videos or images. The algorithm applied to the system was compared with the Gaussian Mixture Model SVM and faster RCNN. Then it appears that faster R-CNN was able to detect the position and type of vehicle more accurately. In other words, it could be said that the deep learning-based object detection approach is superior to the algorithm based object detection system. As a conclusion, all of the development cases in this paper deal with object information, showing outstanding performance with deep learning. However, they all were hard to assign unique IDs to the detected objects and track them by keeping the same ID over time.

Therefore, in this paper, an attempt is made for generate an object detection & tracking system (ODTS), that can obtain moving information of target objects by combining object tracking algorithm with the deep learning-based object detection process. The full ODTS procedures will be described in details in the following section. Also, the tunnel accident detection system in the framework of ODTS will be taken into consideration. This system is used for detecting accident or unexpected events taking place on moving object and target local region on CCTV.

a lean implementation of a tracking-bydetection framework for the problem of multiple object tracking (MOT) where objects are detected each frame and represented as bounding boxes. In contrast to many batch based tracking approaches , this work is primarily targeted towards online tracking where only detections from the previous and the current frame are presented to the tracker. Additionally, a strong emphasis is placed on efficiency for facilitating realtime tracking and to promote greater uptake in applications such as pedestrian tracking for autonomous vehicles. The MOT problem can be viewed as a data association problem where the aim is to associate detections across frames in a video sequence. To aid the data association process, trackers use various methods for modelling the motion and appearance of objects in the scene.

The methods employed by this paper were motivated through observations made on a recently established visual MOT benchmark. Firstly, there is a resurgence of mature data association techniques including Multiple Hypothesis Tracking (MHT) and Joint Probabilistic Data Association (JPDA) which occupy many of the top positions of the MOT benchmark. Secondly, the only tracker that does not use the Aggregate Channel Filter (ACF) detector is also the top ranked tracker, suggesting that detection quality could be holding back the other trackers. Furthermore, the trade-off between accuracy and speed appears quite pronounced, since the speed of most accurate trackers is considered too slow for realtime applications. With the prominence of traditional data association techniques among the top online and batch trackers along with the use of different detections used by the top tracker, this work explores how simple MOT can be and how well it can perform. Keeping in line with Occam’s Razor, appearance features beyond the detection component are ignored in tracking and only the bounding box position and size are used for both motion estimation and data association. Furthermore, issues regarding short-term and long-term occlusion are also ignored, as they occur very rarely and their explicit treatment intro duces undesirable complexity into the tracking framework.

We argue that incorporating complexity in the form of object re-identification adds significant overhead into the tracking framework – potentially limiting its use in realtime applications. This design philosophy is in contrast to many proposed visual trackers that incorporate a myriad of components to handle various edge cases and detection errors. This work instead focuses on efficient and reliable handling of the common frame-to-frame associations. Rather than aiming to be robust to detection errors, we instead exploit recent advances in visual object detection to solve the detection problem directly. This is demonstrated by comparing the common ACF pedestrian detector with a recent convolutional neural network (CNN) based detector . Additionally, two classical yet extremely efficient methods, Kalman filter and Hungarian method, are employed to handle the motion prediction and data association components of the tracking problem respectively. This minimalistic formulation of tracking facilitates both efficiency and reliability for online tracking, In this paper, this approach is only applied to tracking pedestrians in various environments, however due to the flexibility of CNN based detectors , it naturally can be generalized to other objects classes. The main contributions of this paper are:

• We leverage the power of CNN based detection in the context of MOT.

• A pragmatic tracking approach based on the Kalman filter and the Hungarian algorithm is presented and evaluated on a recent MOT benchmark.

• Code will be open sourced to help establish a baseline method for research experimentation and uptake in collision avoidance applications.

Traditionally MOT has been solved using Multiple Hypothesis Tracking (MHT) or the Joint Probabilistic Data Association (JPDA) filters, which delay making difficult decisions while there is high uncertainty over the object assignments. The combinatorial complexity of these approaches is exponential in the number of tracked objects making them impractical for realtime applications in highly dynamic environments. Recently, Rezatofighi et al., revisited the JPDA formulation in visual MOT with the goal to address the combinatorial complexity issue with an efficient approximation of the JPDA by exploiting recent developments in solving integer programs. Similarly, Kim et al. used an appearance model for each target to prune the MHT graph to achieve state-of-the-art performance. However, these methods still delay the decision making which makes them unsuitable for online tracking. Many online tracking methods aim to build appearance models of either the individual objects themselves or a global model through online learning. In addition to appearance models, motion is often incorporated to assist associating detections to tracklets . When considering only one-to-one correspondences modelled as bipartite graph matching, globally optimal solutions such as the Hungarian algorithm can be used . The method by Geiger et al. uses the Hungarian algorithm in a two stage process. First, tracklets are formed by associating detections across adjacent frames where both geometry and appearance cues are combined to form the affinity matrix. Then, the tracklets are associated to each other to bridge broken trajectories caused by occlusion, again using both geometry and appearance cues. This two step association method restricts this approach to batch computation. Our approach is inspired by the tracking component of, however we simplify the association to a single stage with basic cues as described.