Development and Key Technologies of the Cognition system of Autonomous Driving

Yuyang Zhang[a], Junliang Feng[b], Mingni Huang[c], Qinghao Liu[d], Wenxuan Fu[e]

Chang'an Dublin International College of Transportation, Chang'an University, Xi'an, China

[a]China15655230306@163.com, [b]junliang.feng@outlook.com, [c]mingni.huang@ucdconnect.ie, [d]qh_liu18@163.com, [e]1481343302@qq.com

Abstract. With the rapid development of automotive electronics technology and artificial intelligence, autonomous driving has become an important solution for future transportation challenges. As an integral component of autonomous driving systems, the intelligent cognitive system has emerged as a prominent area of technological research. This paper focuses on the development and key technologies of autonomous driving cognitive systems, based on published literature. The architecture of cognitive systems for autonomous driving is typically divided into two subsystems: the environmental perception system and the assessment and prediction system. Our study provides a detailed description of the methods used in both subsystems. We then conduct a comparative analysis and discussion to evaluate their strengths and limitations. Overall, we hope that this study will provide timely and valuable information to a rapidly evolving field concerning cognitive technology for autonomous vehicles.

Keywords: Autonomous vehicles, cognition, perception, assessment and prediction.

1. Introduction

Automatic driving system (ADS), also known as autonomous vehicle technology, has made a breakthrough with the development of artificial intelligence technology. Since the mid-1980s, many universities, research centers, automobile companies, and companies in other industries worldwide have been researching and developing autonomous vehicles. The famous projects are Eureka Project ETHEUS [1], one of the earliest studies on autonomous driving technology carried out in Europe from 1987 to 1995, and Navlab mobile robot [2], etc.

The remarkable progress of ADS essentially depends on the advancement and support of various disciplines, including the Internet of Things and embedded systems, sensors and data acquisition, wireless communication, artificial intelligence, and machine learning. Most ADS divide the complex task of autonomous driving into subcategories and utilize a range of sensors and algorithms across different modules. The typical architecture generally divides the cognition system into two modules, as shown in Figure 1: perception subsystem, assessment and prediction subsystem.

In this paper, we present the development of the cognition system of the autonomous driving. Our aim is to provide a structured and comprehensive overview of the current hardware and software practices associated with autonomous driving. The remainder of this paper is structured as follows. In Chapter 2, we summarize the research on related technologies of environment-aware localization systems, including sensor hardware systems, localization systems for automatic driving vehicles, detection and recognition systems, and recognition and tracking systems. In Chapter 3, we introduce the related technologies of assessment and prediction system, which include risk assessment, surrounding driving behavior prediction, and driving style recognition. Finally, the conclusion is drawn in the last section.

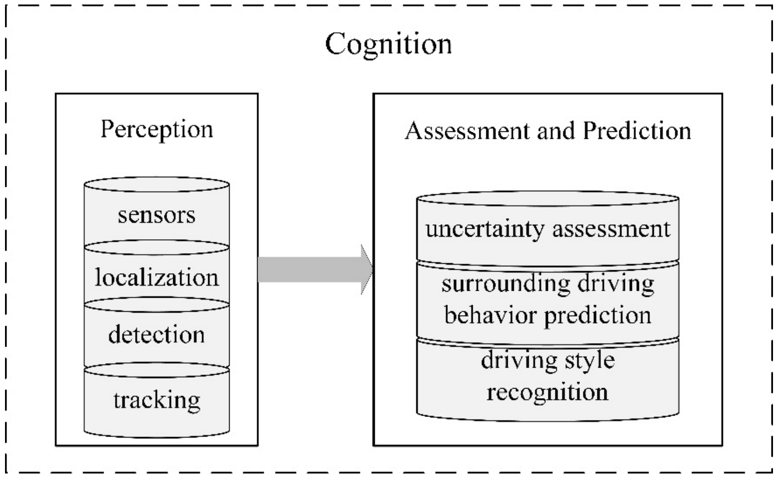

Figure 1. Overview of the cognition system of autonomous driving

2. Perception

Perception systems for self-driving cars generally refer to the integration of conventional positioning technologies (e.g., GPS or odometers) with advanced sensing technologies (e.g., LIDAR, radar, or cameras) to accurately determine the vehicle's position, both in terms of its coordinates and its surroundings. The goal of the sensing system is to extract more detailed and accurate guidance information about the position and direction of the vehicle, considering the specific features and conditions of the driving environment. This includes both static features, such as roads, buildings, and traffic signs, as well as dynamic features, such as other vehicles, pedestrians, and changes in traffic conditions. The system typically combines multiple positioning techniques to enhance accuracy and reliability.

2.1 Sensor System

SExteroceptive sensors, proprioceptive sensors, and system communication sensors are the main types of sensors used in autonomous driving systems. Among these, exteroceptive sensors play a crucial role in sensing the surrounding environment and gathering dynamic real-time information. [3] The exteroceptive sensors mainly include different kinds of cameras and radars, which usually provide 2D or 3D image information, and perform processing such as classification and complementary estimation on moving and stationary input data. With the development of [4] neural networks (e.g., convolutional neural networks) or new materials, the capabilities of exteroceptive sensors for environment sensing, motion tracking, and data processing are improving.

This section focuses on exteroceptive sensors. Proprioceptive sensors and communication sensors are beyond the scope of this review. The next subsections classify exteroceptive sensor systems into four types: camera alone, lidar alone, camera-lidar, and millimeter-wave radar. These classifications describe the concepts, advantages and disadvantages, challenges faced, and emerging technologies in the field.

2.1.1 Camera

The two main reasons why cameras have become popular sensors in self-driving cars are availability and cost. Cameras typically provide two-dimensional RGB image information with high accuracy and large monitoring ranges (which can vary from a few centimetres to close to 100 metres), and can be competent for classification and processing of environmental datasets and for some of the dynamic location tracking tasks. In addition, a complete commercial production line and relatively mature training techniques contribute to the widespread use of cameras in self-driving cars. But cameras have two major drawbacks. One is that due to the principle of light-based imaging, the usability of the camera is significantly reduced in environments with impaired field of view, such as heavy rain and foggy weather. The second is that the camera is limited in the size of the dataset it can process and is poor at depth estimation for sparse data inputs. In order to solve the problem of depth estimation of monocular camera images, Liu et al. [5] proposed an effective model based on deep convolutional neural network (CNN) framework for depth estimation of individual images. This scheme combines the advantages of CNNs and learning continuous conditional random fields (CRFs) to achieve depth estimation of general scenes without geometric prior and additional image information by means of log-likelihood optimisation. Figure 2 depicts a comparison of the results of depth estimation using this CNN model. However, the limitation of this method is that it does not utilize any geometric information from the image, which can result in lower accuracy when estimating depth. Event cameras have great potential to be used as sensors for autonomous driving due to their ability to record pixel brightness changes and their excellent dynamic vision capabilities. Binas et al. [6] investigated the use of a Dynamic Vision Sensor (DVS) in conjunction with a standard Active Pixel Sensor (APS) image stream. We utilized a Convolutional Neural Network (CNN) to learn and predict the steering angle from both the DVS and APS vision data. This approach effectively improves the dynamic range and effective frame rate.

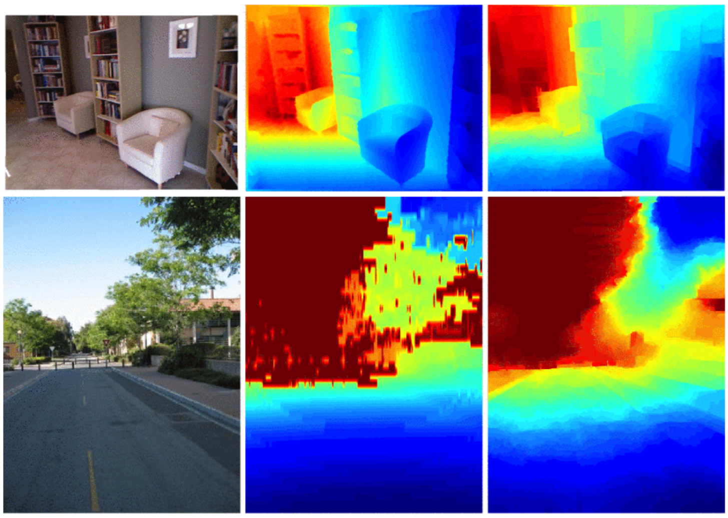

Figure 2. Example of depth estimation results using the proposed deep convolutional neural field model. First row: the NYU v2 dataset; second row: the Make3D dataset. From left to right: input image, ground-truth, author's prediction. [4]

2.1.2 Lidar

Lidar is a radar system that emits infrared light waves to detect the position and velocity of objects, which has two main disadvantages. Firstly, it is greatly affected by the environment. A large number of water droplets in heavy fog or snow will scatter light, which may greatly weaken its detection ability. Secondly, the installation space is large and the lidar is usually several times the size of the camera, occupying a large space, which may affect the aerodynamic performance of the whole vehicle.

To solve these problems, researchers put forward some new lidar technologies. For example, Poulton et al. [6]designed an FMCW LIDAR transmitter and receiver module using silicon photonic technology, which is mainly composed of silicon photonic devices that transmit and receive waveguides and photodetectors. Besides, distance and speed can be measured simultaneously by this device. McManamon et al. [7] have argued that one-dimensional optical arrays work better than two-dimensional arrays for efficient large-angle steering of sensors. The authors have fabricated optical phased arrays with an aperture of 4 cm by 4 cm, in which orthogonally oriented one-dimensional arrays can deflect the beam in two dimensions and can efficiently control the green, red, 1.06 /spl mu/m and 10.6 /spl mu/m radiations with a high degree of accuracy, pointing, and optical power handling.

2.1.3 Millimeter Wave Radar

Millimeter wave radar is a radar working in a millimeter wave band, which has many advantages compared with other commonly used perceptive sensors. Millimeter wave radar has strong environmental adaptability, which can keep a good working state in extreme environments. However, millimeter wave radar also has the shortcoming. Because the transmission power of the signal source is limited, it will lead to an unfavorable transmission path and a large loss.

Driven by deep learning (DL) technology, extensive research on the fusion of radar sensor data processing using deep neural networks (DNN) has emerged. This research has significantly enhanced the performance of radar systems in terms of detection, perception, mapping, and other related aspects. Schumann et al. [8]investigate the optimisation of deep convolutional neural networks (CNNs) and the authors use the PointNet++ classification algorithm for semantic segmentation of radar data, a method that also consists of a two-step pre-processing step, namely clustering and feature generation. The authors ultimately demonstrated that the new method improves the radar environmental data processing and the grouping of different objects to obtain a high level of semantic understanding of the surrounding environment. Chang et al. [9]proposed a new method called Spatial Attention Fusion (SAF) for environmental monitoring using visual sensors embedded in millimeter-wave radar. This method combines the features of both millimeter-wave radar and visual sensors to enhance the information processing capabilities of the sensing system. Information processing capability. The authors also developed a generative model that transforms radar points into radar images, which can be integrated into a deep learning data framework for training deep neural networks (DNNs). Experimental results show that the new fusion method has better detection performance than traditional visual detection frameworks at all scales.

2.1.4 Camera-Lidar Fusion

In recent years, sensor fusion methods with complementary characteristics to enhance perception and reduce cost have become new research, with camera-lidar fusion representing the fusion model of camera and radar. According to different types of cameras, camera-lidar fusions are divided into monocular camera-lidar fusion as well as stereo camera-lidar fusion.

As for the monocular camera-lidar fusion, Ma et al. developed a depth regression network model for depth completion and a framework for self-monitoring and training only by color images and sparse depth image sequences. Park and others [10] proposed a sensor fusion framework for high-precision depth estimation, which consisted of a calibration network and deep fusion network. The calibration network estimates external parameters through disparity input, and the deep fusion network encodes by using complementary characteristics of sparse LiDAR data and dense stereo depth information of the stereo camera. Jaritz et al. [11]used an encoder-decoder scheme to train a dense CNN architecture that can process sparse data with minor variations and sparse training strategies. It can process data using RGBs of different densities, effectively combining dense RGBs with sparse data in a late fusion fashion to accurately process inputs of various densities, as shown in Fig. 3. The authors also demonstrate the robustness of the new method to varying densities of bottom data and different levels of input sparsity. Cheng et al. [12] proposed LidarStereoNet, an unsupervised lidar-stereo fusion network, to address the difficulty of collecting large-scale ground truth depth datasets and the problem of noisy radar point interference. The authors propose a novel "feedback loop" that connects the sparse inputs to the depth-estimated outputs. This approach effectively reduces the impact of noisy radar point interference and enables end-to-end training without relying on ground-truth depth images. In addition, the authors propose stacking and migrating multiple small datasets, merging segmented planar models into network learning, and building special convolutional architectures to further refine depth data processing and 3D geometry modeling.

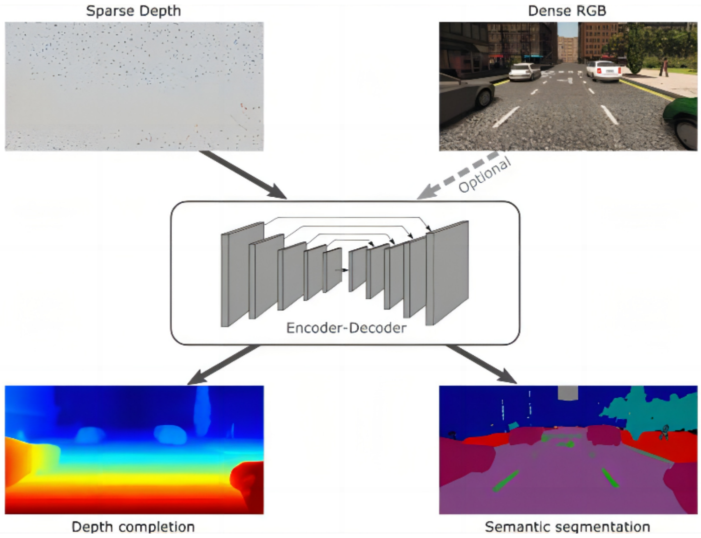

Figure 3. Deep complementation or semantic segmentation done by processing sparse deep input data with or without additional dense RGBs [11]

2.2 Localization System

The localization task is responsible for finding the self-position relative to the map, while the mapping work builds a multi-layer high-definition map for path planning. Therefore, the accuracy of localization and mapping will affect the feasibility and security of path planning.

2.2.1 Global Positioning System and Inertial Measurement Unit (GPS-IMU)

GPS-IMU system is a typical multi-sensor fusion algorithm combined with map matching. When a single IMU is used, error values are accumulated with continuous dead reckoning, which leads to error accumulation over time. Using GPS-IMU systems to integrate the measurement data of other sensors to generate more accurate high-resolution maps may be an effective way to solve its accuracy defects. Montemerlo et al. [13] integrated GPS, IMU, wheel mileage measurement, and lidar data, which generated infrared spectral two-dimensional surface images with 5 cm pixel resolution through an offline relaxation technique. According to experiment results, the relative accuracy of map data generated by the new method is more than one order of magnitude higher than that of the traditional method based on GPS-IMU.

2.2.2 Simultaneous Localization and Mapping (SLAM)

The principle of SLAM [14] is to help simultaneous localization in an unknown environment by repeatedly observing environmental characteristics, and then incrementally building environmental maps according to their position and attitude, so as to realize simultaneous localization and mapping. SLAM technology basically solves the shortcomings of traditional localization systems, which are universal enough to be combined with almost any sensor and estimation technology to realize simultaneous localization and mapping. SLAM can be divided into filter-based SLAM and optimization-based SLAM.

1) Filter-based SLAM

SLAM technique based on an Extended Kalman Filter (EKF) is being studied. EKF algorithm can perform linearization steps on the estimation, which enables it to be accurate in estimating the state vector. Optimization-oriented SLAM usually consists of two subsystems. The first is used to find the corresponding relationship between new observations and maps and determine constraints according to sensor data. The second calculates or refines the vehicle attitude and the map of the given constraints to have a coherent whole. Due to the large amount of computation in global optimization, local optimization, small block layering, and integration techniques are usually adopted. Paz et al. [15]proposed a Divide and Conquer SLAM as a variant of the EKF SLAM technique. The algorithm implements the use of estimates and covariances directly in data association without any computation.

Experiments demonstrate that the computational complexity cost of the new method is an order of magnitude lower than that of the original KF SLAM technique at each step of the computational process. Additionally, the resulting vehicle positioning estimates and mapping are more accurate than those obtained using standard EKF SLAM. However, as the accuracy requirements for generating maps continue to increase, the maps may become so large that the pace of traditional EKF algorithms supports a large SLAM framework. Therefore, spatial scalability and long-term operation have become the focus of research, and local mapping and layering techniques are emerging.

2) Optimization-based SLAM

In addition, incremental motion estimation methods, as a type of local optimisation technique, have been investigated for dealing with sparse image sequences, which have significant advantages in terms of attempted matching and consistency. An incremental method for motion and structure estimation of long images was proposed by Zhang et al [16]. The method uses a threaded operation to connect two consecutive elementary matrices to implement a sliding window over a ternary keyframe image. Experiments demonstrate that the new method exhibits a high degree of consistency and robustness in processing speed-demanding procedures as well as on several local processes overlapping consecutive images, and the processing results are very close to those of global bundle adjustment. However, the technique of obtaining structure from motion data (SfM) in incremental-based SLAM may suffer from poor management of critical motions (e.g., pure rotations), leading to inaccurate motion pose estimation and error accumulation. To address this problem, Michot et al. [17] proposed an online EKF batch algorithm based SLAM that can process and integrate visual as well as inertial data from multi-sensor fusion. The authors analysed three different bundle adjustment weights using the BA cost function and showed that the real-time auto-weighting technique using L-curve weighting had the highest accuracy and the lowest bias in motion estimation.

2.2.3 Localization Based on a Prior Map

At present, the method of using pre-built high-definition maps or a priori maps is widely used. The essence of localization technology, which relies on a pre-existing map, is to correlate the real-time sensor data with the pre-built high-definition map. This allows for precise vehicle positioning and motion prediction. There are two mainstream matching technologies for prior map localization: landmark-based technology and vision-based technology. Landmark-based technology typically relies on cameras or radars to identify landmarks such as telephone poles, curbs, and road signs in urban environments. This approach requires significantly less computational power compared to point cloud-based technology. Vision-based technology is a novel approach that primarily involves matching piror maps with 2D camera data and matching 2D data with high-precision 3D maps. It has a high computational cost but may have low sensor requirements. Wolcott et al. [18] propose a visual localisation system that generates accurate localised data at reduced cost. This system first locates a monocular camera in a 3D a priori map constructed by a LiDAR scanner, followed by a graphics processing unit (GPU) that generates multiple synthetic views, which are directly compared and normalised to the camera's live image (Figure 4). The authors state that this method is simpler and more practical than methods that rely on complex feature sets for comparative stabilisation, and that highly accurate localisation can be achieved with the help of mobile-grade GPUs.

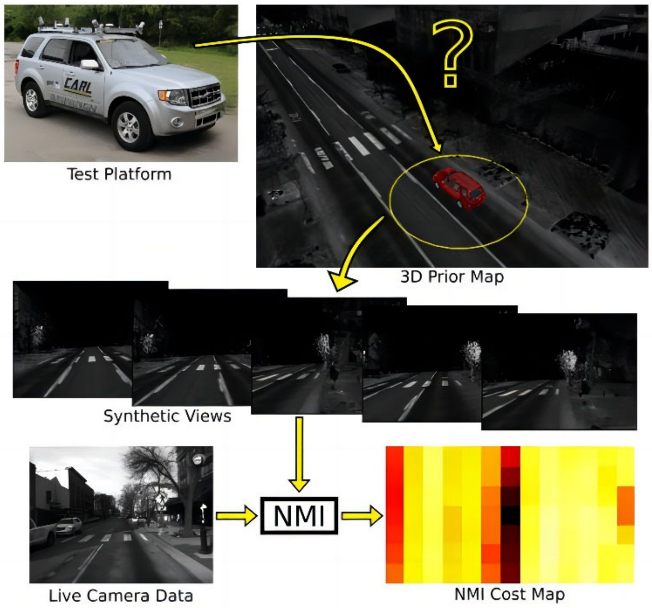

Figure 4. Overview of the image processing flow of the visual localisation system proposed by Wolocott et al. [18]

2.3 Detection System

Detection is an important part of the environment awareness component in an autonomous driving system. Figure 4 illustrates the increasing number of publications on "target detection" over the past two decades. It is usually categorized into object detection and lane detection. For object detection, the mainstream research direction is to address the trade-off between processing speed and accuracy [19]; for lane detection, the main focus is to improve its semantic segmentation and semantic understanding capabilities. With the development of deep learning technology (such as CNN or DNN) [20], the data classification ability in the detection method is more accurate, and the semantic segmentation and understanding are faster and more effective. Additionally, many emerging deep learning network fusion detection techniques have been developed.

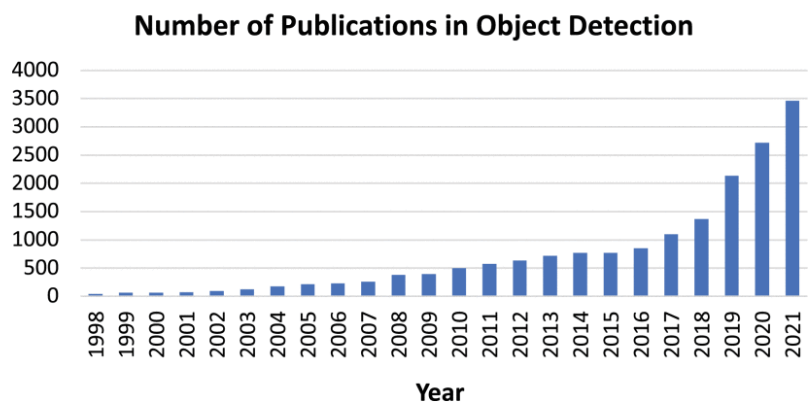

Figure 5. Increasing number of publications on object detection from 1998 to 2021 [21]

2.3.1 Object Detection

Generally speaking, object detection technology is divided into traditional object detection algorithms and deep learning-based detection.Some core technologies in traditional detection algorithms, such as Viola-Jones Detectors and directional gradient histogram description method [4], are also enlightening to the emerging object detection technology based on deep learning. As for the object detection technology based on deep learning, Hsu et al. [22] designed a fast recognition convolutional neural network (Fast R-CNN) for vehicle detection. To address the issues of overfitting and depth degradation resulting from excessive network depth, the author removed the object recognition component from the original R-CNN and only kept its regional proposal network layer. Experiments prove that the new system can quickly and effectively detect object vehicles, even with limited training data. In addition, Redmon et al. [23] proposed the YOLO model for object detection in 2015 for the first time, whose principle is different from the traditional detection method using the classifier. Object detection was defined as two kinds of problems, that is, image bounding box and class probability. A single neural network was once used to predict directly in one evaluation.

2.3.2 Lane Detection

Bounding box technology and other core ideas in previous object detection are not suitable for road detection, because roads belong to continuous surfaces. For the automatic driving system, it is a vital challenge to fully identify the differences and connections of different lanes through semantic segmentation and semantic understanding before choosing the driving road. Traditional lane detection technology is usually based on direct visual analysis. Lee et al. [24] proposed a multi-task road detection system VPGNet that can be optimized end-to-end. A benchmark dataset of lane and road signs is established and used to fully train multiple versions of the multitasking network. Experiments manifest that the new system solves information damage in low illumination images, realizing simultaneous high accuracy (20 fps) and robustness in severe weather conditions. Li et al. [25] proposed an extended deep neural network framework incorporating spatial structure based on deep CNNs. The authors develop a multi-task deep convolutional network that can simultaneously take into account the presence of targets as well as detect the spatial-geometric properties of targets to help the subsequent accurate modelling of lane structures. Second, the authors develop a companion recurrent neuron layer that can directly and automatically detect lane boundaries without prior knowledge. The fusion of the novel hierarchical neural network and recurrent neurons enables accurate modelling of roads and lanes as well as accurate determination of obstacles and boundaries.

2.4 Tracking System

The object tracking system is usually called Detection and Tracking of Moving Objects (DATMO) or Multi-Object Tracking (MOT). Its main function is to locate multiple objects, judge data categories, and predict the next moving course and speed of objects. Compared with single object tracking (SOT), MOT pays more attention to recognizing complex external shapes and the motion prediction of multiple objects. In addition, the main challenges in the design of MOTs are in 1). Frequent masking; 2). Similar target differentiation; 3). Interaction of multiple tracked targets; 4). computational cost of 3D environment map construction.

In order to solve the problems of poor validity of sensor 2D sliced data and excessive space for 3D point cloud data input, Azim et al. [26] designed a framework based on 3D distance sensors and the octree data structure to address DATMO in dynamic environments. The new system builds upon the previous approach of 2D to 3D mapping by utilizing octree mesh mapping to model the vehicle's surrounding environment. It compares the current image with the previously scanned image to identify any inconsistencies and track moving objects in real time. Experiments demonstrate that the new method can efficiently detect and track dynamic objects around the vehicle, achieving an accurate representation of the 3D surroundings while consuming less memory. However, the limitation of this method is that tracking multiple targets simultaneously in a chaotic environment may result in a significant decrease in accuracy. Hwang et al. [27] proposed a new modular DATMO framework based on sensor fusion. Firstly, the image information from the color camera and lidar was segmented in this framework. Object detection was divided into two-dimensional image detection and three-dimensional space detection. Then, two kinds of information were matched and combined by fractional calibration to realize continuous tracking of moving objects. The 2D/3D joint object proposal in this framework greatly improved the efficiency of information processing. Thanks to sensor fusion, the robustness of object tracking in complex environments is enhanced.

3. Assessment and Prediction

The primary function of the assessment and prediction system is to constantly evaluate the overall risk level surrounding the vehicle and forecast the intentions of nearby pedestrians, vehicles, and obstacles using the prediction module. The assessment primarily focuses on the driving environment, while the prediction is based on the data processed by the assessment and the driver's actions. A well-designed assessment and prediction system can effectively avoid accidents and improve safety performance. Usually, the assessment, prediction, and decision planning of an autonomous driving system are interconnected or integrated. A correct assessment of the surrounding driving environment and obstacles is necessary for generating a reasonable global and local trajectory through decision planning. In this chapter, we will discuss the evaluation prediction system, which is primarily divided into three distinct functional systems.

3.1 Uncertainty Assessment

Measuring and estimating the uncertainty of a driving scene is an effective risk assessment. Meanwhile, proper uncertainty quantification is one of the major challenges in the application of deep learning networks in perception and tracking systems [28]. To address this problem, Russell et al. [29] implemented a Bayesian deep learning network to model multivariate uncertainty in regression problems. They used two separate methods for modeling multivariate uncertainty: direct training of the Gaussian maximum likelihood loss function and inverse training for uncertainty prediction via a Kalman filter. It has been experimentally demonstrated that this approach can accurately assess uncertainty while significantly improving the performance of the filter. However, this method assumes that the evaluated uncertainties all follow a multivariate Gaussian distribution. This assumption implies that the accuracy of the evaluation decreases in highly nonlinear systems. Another method of uncertainty assessment is to conduct an overall risk level assessment of the environment. Yurtsever et al. [30] designed a driving risk assessment framework based on deep learning for a lane change, which used a convolution neural network (CNN) and time-distributed long-term and short-term memory network (LSTM) trained on large data sets to classify and assess the risk of lane change videos taken by cameras. It is shown that the new method can only rely on monocular cameras to complete the analysis, advantageous in low cost, high accuracy, and strong implement ability.

3.2 Surrounding Driving Behavior Assessment

Accurately estimating the driving behavior of surrounding vehicles, particularly the intentions of surrounding human drivers, is essential for making reasonable medium- and long-term predictions and behavioral decisions. Given the limited adaptability of traditional prediction models in predicting driving behavior across various driving scenes, Geng et al. [31] proposed a new adaptive ontology model based on the Hidden Markov Algorithm (HMM) to simulate traffic scenarios. The method can specify candidate models, input features, and prior knowledge in various scenarios using rules. It then combines a priori and a posteriori probability to predict the future behavior of the target vehicle. Experiments demonstrate that the new method significantly improves the estimation of lane-changing behaviors in terms of prediction time horizon and long-term prediction accuracy compared to the traditional method. Kumar et al. [32] proposed a new model to evaluate lane change intention, which consists of a support vector machine and a Bayesian filter. The real-world data collected from the lane tracker were passed through a multi-class support vector machine and a Bayesian filter in sequence. The final evaluation of lane changes was determined based on the output results. In the actual road test, it was found that the algorithm not only accurately predicts lane changes an average of 1.3 seconds before they occur, but also accurately distinguishes between left and right lane changes. However, fluctuations in the lane tracker data may lead to a decrease in the accuracy of the prediction results. Future improvements may include enhancing the perceptual stability of the integrated lane tracker. This can be achieved by increasing the variety of input data, such as vehicle lateral distance and cues indicating driver overtaking intentions, as a way to enhance the robustness of the system.

3.3 Driving Style Recognition

Most of the prediction and evaluation of driving behavior is aimed at the common reaction patterns of drivers under natural driving conditions, which ignores their unique driving style. Driving style [33] is not a strictly defined term, but it profoundly influences the different driving behaviours made by drivers in different situations, which is difficult to detect by traditional recognition and prediction models (Figure 5). Doshi et al. [34] suggest that the driver assistance system should be adaptively updated based on the driving style and provide customized behavioral decision-making suggestions to the driver. The authors classify the group of drivers into "aggressive" and "non-aggressive" based on various metrics, including transverse and longitudinal acceleration, as well as additive acceleration. They compare the two groups in terms of predictability, responsiveness, and other factors. Experiments show that "aggressive" drivers are more predictable, while "non-aggressive" drivers are more responsive to ADAS feedback. However, this categorization may still be insufficiently detailed, and future research will be conducted to further categorize the group of "non-aggressive" drivers. In addition, categorizing driving styles based on fuel consumption is another common method. Corti et al. [35] propose a new method for quantitatively estimating driving styles focused on energy consumption, using traditional inertial measurements. Such a system evaluates three quantitative cost functions related to vehicle energy consumption in real-time and can provide visual feedback to the driver through a smartphone application. Tests show that the system adapts to different driving styles and generally promotes energy savings of up to 20-30 percent.

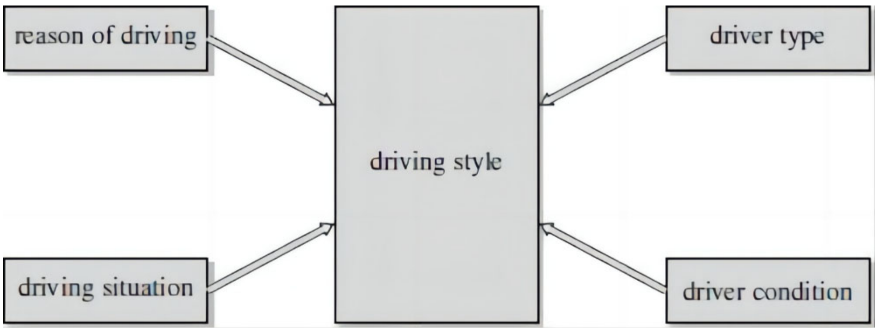

Figure 5. Different influences on driving style [33]

4. Conclusion

In conclusion, this paper examines two modules of autonomous driving: perception system and assessment and prediction system. These modules are considered cognition systems for autonomous driving. It provides a thorough and comprehensive review of this crucial technology in autonomous driving. Starting with various sensor hardware systems, this paper outlines the fundamental operating principles and recent advancements in the positioning, detection, and tracking of self-driving vehicles in the perception section. In describing the assessment and prediction techniques, the paper focuses on uncertainty assessment, prediction of driving behavior, and recognition of driving style. These studies are crucial for understanding cognitive technologies for self-driving cars, although they do not provide comprehensive coverage of all technologies.

Another contribution of this paper includes providing an overview reference for the further development of the technology. The development of self-driving cars, and specifically their cognitive technologies, is currently facing a bottleneck. Uncertainty arising from various factors that need to be taken into account in real-world situations is the main cause of bottlenecks. Therefore, there is a need to further advance current technology in the field of multidisciplinary intersections. This can be achieved by integrating the machine learning techniques of the latest generation of artificial intelligence with various sensory hardware. It is hoped that this overview of self-driving technology will provide the public with a better understanding, and that scientists working on self-driving car technology will be better able to design safe and intelligent future transportation systems.

References

[1] E! 45: Programme For a European Traffic System With Highest Efficiency and Unprecedented Safety, May 2019, [online] Available: https://www.eurekanetwork.org/project/id/45.https://www.eurekanetwork.org/project/id/45.

[2] Thorpe, Charles, et al. "Toward autonomous driving: the cmu navlab. ii. architecture and systems." IEEE expert 6.4 (1991): 44-52.

[3] L. Liu et al., "Computing Systems for Autonomous Driving: State of the Art and Challenges," in IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6469-6486, 15 April15, 2021, doi: 10.1109/JIOT.2020.3043716.

[4] F. Liu, C. Shen, G. Lin and I. Reid, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 10, pp. 2024-2039, 1 Oct. 2016, doi: 10.1109/TPAMI.2015.2505283.

[5] Binas, Jonathan, et al. "DDD17: End-to-end DAVIS driving dataset." arXiv preprint arXiv:1711.01458 (2017).

[6] C. V. Poulton, D. B. Cole, A. Yaacobi and M. R. Watts, "Frequency-modulated continuous-wave LIDAR module in silicon photonics," 2016 Optical Fiber Communications Conference and Exhibition (OFC), Anaheim, CA, USA, 2016, pp. 1-3.

[7] P. F. McManamon et al., "Optical phased array technology," in Proceedings of the IEEE, vol. 84, no. 2, pp. 268-298, Feb. 1996, doi: 10.1109/5.482231.

[8] O. Schumann, M. Hahn, J. Dickmann and C. Wöhler, "Semantic Segmentation on Radar Point Clouds," 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 2018, pp. 2179-2186, doi: 10.23919/ICIF.2018.8455344.

[9] Chang S, Zhang Y, Zhang F, Zhao X, Huang S, Feng Z, Wei Z. Spatial Attention Fusion for Obstacle Detection Using MmWave Radar and Vision Sensor. Sensors. 2020; 20(4):956. https://doi.org/10.3390/s20040956

[10] Kim, Tae-Hyeong, and Tae-Hyoung Park. "Placement optimization of multiple lidar sensors for autonomous vehicles." IEEE Transactions on Intelligent Transportation Systems 21.5 (2019): 2139-2145.

[11] Jaritz, Maximilian, et al. "Sparse and dense data with cnns: Depth completion and semantic segmentation." 2018 International Conference on 3D Vision (3DV). IEEE, 2018.

[12] Cheng, Xuelian, et al. "Noise-aware unsupervised deep lidar-stereo fusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[13] M. Montemerlo, J. Levinson and S. Thrun, Map-Based Precision Vehicle Localization in Urban Environments, Collection of open conferences in research transport (2007). Vol. 2007, 141 URL https://www.scipedia.com/public/Montemerlo_et_al_2007a

[14] G. Bresson, Z. Alsayed, L. Yu and S. Glaser, "Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving," in IEEE Transactions on Intelligent Vehicles, vol. 2, no. 3, pp. 194-220, Sept. 2017, doi: 10.1109/TIV.2017.2749181.

[15] L. M. Paz, J. D. TardÓs and J. Neira, "Divide and Conquer: EKF SLAM in $O(n)$," in IEEE Transactions on Robotics, vol. 24, no. 5, pp. 1107-1120, Oct. 2008, doi: 10.1109/TRO.2008.2004639.

[16] Z. Zhang and Y. Shan, "Incremental motion estimation through modified bundle adjustment," Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 2003, pp. II-343, doi: 10.1109/ICIP.2003.1246687.

[17] Michot, Julien, Adrien Bartoli, and François Gaspard. "Bi-objective bundle adjustment with application to multi-sensor slam." 3DPVT'10 3025 (2010).

[18] R. W. Wolcott and R. M. Eustice, "Visual localization within LIDAR maps for automated urban driving," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 2014, pp. 176-183, doi: 10.1109/IROS.2014.6942558.

[19] McCall, Joel C., and Mohan M. Trivedi. "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation." IEEE transactions on intelligent transportation systems 7.1 (2006): 20-37.

[20] Zou, Zhengxia, et al. "Object detection in 20 years: A survey." Proceedings of the IEEE (2023).

[21] Z. Zou, K. Chen, Z. Shi, Y. Guo and J. Ye, "Object Detection in 20 Years: A Survey," in Proceedings of the IEEE, vol. 111, no. 3, pp. 257-276, March 2023, doi: 10.1109/JPROC.2023.3238524.

[22] S. -C. Hsu, C. -L. Huang and C. -H. Chuang, "Vehicle detection using simplified fast R-CNN," 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 2018, pp. 1-3, doi: 10.1109/IWAIT.2018.8369767.

[23] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[24] Lee, Seokju, et al. "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition." Proceedings of the IEEE international conference on computer vision. 2017.

[25] J. Li, X. Mei, D. Prokhorov and D. Tao, "Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 3, pp. 690-703, March 2017, doi: 10.1109/TNNLS.2016.2522428.

[26] A. Azim and O. Aycard, "Detection, classification and tracking of moving objects in a 3D environment," 2012 IEEE Intelligent Vehicles Symposium, Madrid, Spain, 2012, pp. 802-807, doi: 10.1109/IVS.2012.6232303.

[27] S. Hwang, N. Kim, Y. Choi, S. Lee and I. S. Kweon, "Fast multiple objects detection and tracking fusing color camera and 3D LIDAR for intelligent vehicles," 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Xi'an, China, 2016, pp. 234-239, doi: 10.1109/URAI.2016.7625744.

[28] Russell, Rebecca L., and Christopher Reale. "Multivariate uncertainty in deep learning." IEEE Transactions on Neural Networks and Learning Systems 33.12 (2021): 7937-7943.

[29] Russell, Rebecca L., and Christopher Reale. "Multivariate uncertainty in deep learning." IEEE Transactions on Neural Networks and Learning Systems 33.12 (2021): 7937-7943.

[30] Yurtsever, Ekim, et al. "Risky action recognition in lane change video clips using deep spatiotemporal networks with segmentation mask transfer." 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019.

[31] Kumar, Puneet, et al. "Learning-based approach for online lane change intention prediction." 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2013.

[32] Kumar, Puneet, et al. "Learning-based approach for online lane change intention prediction." 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2013.

[33] Dörr, Dominik, David Grabengiesser, and Frank Gauterin. "Online driving style recognition using fuzzy logic." 17th international IEEE conference on intelligent transportation systems (ITSC). IEEE, 2014.

[34] Doshi, Anup, and Mohan M. Trivedi. "Examining the impact of driving style on the predictability and responsiveness of the driver: Real-world and simulator analysis." 2010 IEEE Intelligent Vehicles Symposium. IEEE, 2010.

[35] Corti, Andrea, et al. "Quantitative driving style estimation for energy-oriented applications in road vehicles." 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2013.