Autonomous Vehicle Navigation Based on Vision and Mapless Strategies

Yunsong Guo*

Swinburne College, Shandong University of Science and Technology, Jinan, China

* Corresponding Author Email: 2015141241023@stu.scu.edu.cn

Abstract. Vision camera recognition system has many roles as face recognition can be used in the criminal system, object recognition can quickly classify items, and behavioral recognition can be placed in the intelligent cockpit to identify and analyze the driver's behavior and make the correct judgment and decision-making. Vision cameras can also be used in automatic automobile driving because the camera will not be distracted and can be installed with multiple cameras to observe simultaneously, so the camera can constantly observe the surrounding situation, which is incomparable to human vision. Mapless traveling is also probable to happen in life, so it is a very common situation. The use of mapless and camera recognition systems is also extensive. In this paper, I will introduce three kinds of autonomous driving. Firstly, I will introduce optical flow, which is made by stacking images frame by frame and calculating the speed and position of objects through the position of each frame; secondly, vision tracking, which is caused by camera shooting and calculating to localize the other objects and make the right decision for the vehicle; and lastly, the no-map strategy, which also combines the camera and radar, and combines the two. Finally, the mapless strategy can also combine camera and radar, combining the advantages of the two to make the data of autonomous driving more accurate so that autonomous driving will be safer for drivers and passengers.

Keywords: Autonomous Driving, Vision Assist, Radar Assist, Data Computing.

1. Introduction

Autonomous driving is a function based on multiple sensors and a large amount of data to control the vehicle to reach its destination safely. Today's autonomous driving still requires human intervention at critical moments. Still, in the future, autonomous driving will take the place of the driver and run the car independently on the road. Autopilot is a vital support function on the road, and the driver's control of the vehicle is based on visual and physical contact. Whether traveling at high or low speeds, vision is always first for the driver's judgment, and because visual judgment can be near or far, there are fewer limitations on distance, and it makes up for the fact that the driver can be distracted. In the process of driving on the road, there are many unknown factors, such as the side of the vehicle overtaking or everyday driving, can ensure that you change lanes or turn around when the passing speed of the opposite lane can be safely passed, safely is always the first of the automatic driving, rather than the passing speed. The map is also traveling for the route to determine a significant focus, but there is a map of the path that someone walked; if you reach an unknown road, the map will no longer be applicable if you can scan the environment through the camera around the calculation of the best path which will make up for the lack of data because of maps and other data led to the route is lost, so mapless strategies at this time will play a more significant role. The route planning from the camera can also be used to deal with unexpected events and road signs, such as temporary road closures, road repairs, or speed limit signs.

Autonomous driving also has some drawbacks; when encountering rain, fog, snow, and other extreme weather, vision may be inaccurate data and other problems, resulting in reduced safety or traffic accidents. Therefore, in these harsh conditions, vision cameras and related computing systems must be upgraded to ensure the best results and efficiency [1]. If the camera and radar can be fused and the data obtained from both can be integrated, it will be a great improvement for the autonomous driving of the vehicle. Therefore, vision and mapless strategies will play a more critical role in the case of no signal or no map.

2. Optical flow

Optical flow is the motion of an object between successive frames of motion over a continuous period of time. It is created between the camera and the object due to the presence of a velocity difference. If you are in a three-dimensional scene, the motion field is the vector motion of all points of the full-view image, where each moving point is a three-dimensional vector. An optical flow is a two-dimensional velocity field as if the object in three dimensions were projected onto a plane. Optical flow is the instantaneous velocity of the motion of a pixel of a moving object in space in the plane of the viewing image. When an object is viewed with the naked eye in motion, a series of motion images are produced simply because the naked eye cannot view the position and velocity of the object at any given moment. Still, the velocity of the object's motion is being recorded at every moment. Still, as the vehicle is moving forward, everything in the vicinity that is stationary or that has a relative velocity is in motion for the car. So all of it produces an optical flow. All of it has a velocity, and all of it is there. You can think of the car as a stationary object, then everything around it is in motion.

Because this is all motion, can calculate and predict the speed, direction, and positional trends of these things in action. Multiple cameras and position itself can photograph the car according to the difference in speed between the positions of the cameras; for example, when the speed of the left and right cameras is the same, it will continue to keep moving forward, and if the speeds of the two cameras are different, it will actively move closer to the slower side of the car to ensure that it is positioned accurately. Centering reflex has inspired the development of several successful navigation systems, which have been implemented to guide a mobile robot through A mobile robot platform equipped with a binocular vision system to estimate optical flow, which, in some manner, mimics the corridor-following behavior for navigation. Corridor-following behavior for navigation purposes [2].

There are two main types of computation regarding optical flow, which are gradient-based and matching methods. Gradient-based methods compute the optical flow directly from the luminance variation of the image, of which the method of Horn and Schunck is an example, where the optical flow is computed by minimizing the energy function—matching-based methods such as Lucas-Kanade. The Lucas-Kanade method is a widely used method for optical flow estimation, where Lucas-Kanade assumes that the optical flow vectors of all pixels are the same over a period of time and then solves for them using the least squares method.

In addition, the MARS (DARPA-Mobile Autonomous Robot Software) program has developed a simple and efficient navigation system based on optical flows [3]. The proposed technique estimates image edge maps by first detecting zero Laplacian Gaussian (LOG) intersections. The edge maps of subsequent frames are then used in a speckle-matching procedure. The edge maps of successive frames are then used in a speckle-matching procedure to calculate the corresponding optical flows.

3. Visual tracking

Any object is made up of lines and angles, so the way to localize the position and speed of other objects based on lines and angles in a so-called moving video is widespread and is gradually being used in visual tracking for autonomous driving.

Visual tracking uses this principle to identify and localize moving objects. Focused our development on a limited number of essential categories within each category. The first set of important categories are road objects. They represent static infrastructure, i.e., roads, lanes, and curbs. They cannot, therefore, need to be segmented. The second category is dynamic objects, including vehicles, pedestrians, and cyclists [4]. Siam et al. provided a detailed overview of semantic segmentation for automated driving applications [5]. The majority of semantic segmentation algorithms follow an encoder/decoder-like architecture. The encoder is a trimmed version of a pre-trained classification network, and the decoder consists of a series of deconvolutions or upscaling layers [6][7]. Cameras are mounted on multiple sides of the vehicle to capture and record the environment around it, taking concise videos or photos of each period of time and uploading the results to a computer that calculates and determines the position of the vehicle, the position of the reference, the speed at which it is traveling, the speed of the other vehicles, and the relative speeds of the other vehicles. When the camera for other objects around the scanning process also makes a certain error, such as the size of the reference, height and the exact position, other vehicles, and the vehicle position difference, these are the most critical impact on the automatic driving path planning, because if the vehicle is entirely dependent on the camera's visual scanning and calculation of the vehicle's automated assisted driving, then this will be the only basis.

Cameras have the following characteristics: high-resolution cameras can capture high-resolution images, and the vehicle can accurately display a variety of road signs, signals, the rest of the vehicles and pedestrians on the road, as well as obstacles, which can maximize the smoothness and safety of traffic. The camera can not only shoot when the vehicle is running, but also can shoot when it is stationary, for any object passing by the car when parking, whether it is a car or a person, can play a role in monitoring, so as to ensure the safety of the vehicle. And because the body is surrounded by cameras, it can have a more complete function and decision-making during autonomous driving. With the development of technology, today's cameras are very good at both day and night, and some of them are also equipped with infrared rays for excellent night imaging, so don't need to worry about the clarity of the camera to ensure the best auto-pilot status at both day and night. Cameras can also record passing environments. If you frequently travel the same stretch of road, such as a parking lot or garage, can set the camera and vehicle to drive to a designated location upon entry automatically, but this feature also has certain drawbacks; for example, these vehicles may be prohibited from entering a specific department of the classified area or approaching certain particular areas. Economically speaking, the cost of the camera is lower, so this is more money for the development of autonomous driving can be invested in the subsequent maintenance and development, which will lead to better autonomous driving.

In addition to the above advantages, cameras have some disadvantages compared to radar detection. In some extreme weather, such as heavy rain, snow, and fog, the imaging effect of the camera will be greatly reduced because in these weather, the field of view will become less apparent or even maybe snowflakes or water blocking the camera, resulting in the inability of safe autonomous driving. In places where black and white meet, such as the moment of entering or exiting a tunnel, the camera may not be able to see the road ahead for a split second, just like the human eye, which is faced with an unknown environment ahead of it and around it. Furthermore, the camera calculates the distance and position of the surrounding objects through vision, which is entirely different from the calculation method of radar, and the error of the camera's calculation method will be larger than that of radar. There is also some room for progress in software adaptation and calculation; if encountered on the road advertising or signage, it is easy to make errors is another situation; in this case, the results are inaccurate or even may lead to dangerous occurrences, so this requires more powerful recognition and calculation capabilities, which will undoubtedly increase the pressure on the designer.

The above shortcomings can be slowly solved with the development of the advantages of the camera. Automatic driving is still very obvious, so with the progress of technology and algorithms, the camera can build a more powerful and more reliable automated driving system to meet the increasing demand for automatic driving.

Tesla's visually assisted autopilot is a good example. The Tesla chip on-board can identify over 250 signs in more than 50 countries [8]. These include everything from turn signs to speed limits. The system can also identify and interpret traffic lights, road markings, and general items like traffic cones. The system has the capability to detect the road surface as well as any debris present. This allows the Tesla car to be aware of not only what kind of road the car is traveling on (highway, countryside, etc.) but also to detect debris and other undesirables, such as potholes on the road (and consequently avoid them). Tesla Visual Autopilot Recognition Backend, the visual sensor recognition can be clearly seen in Figure 1.

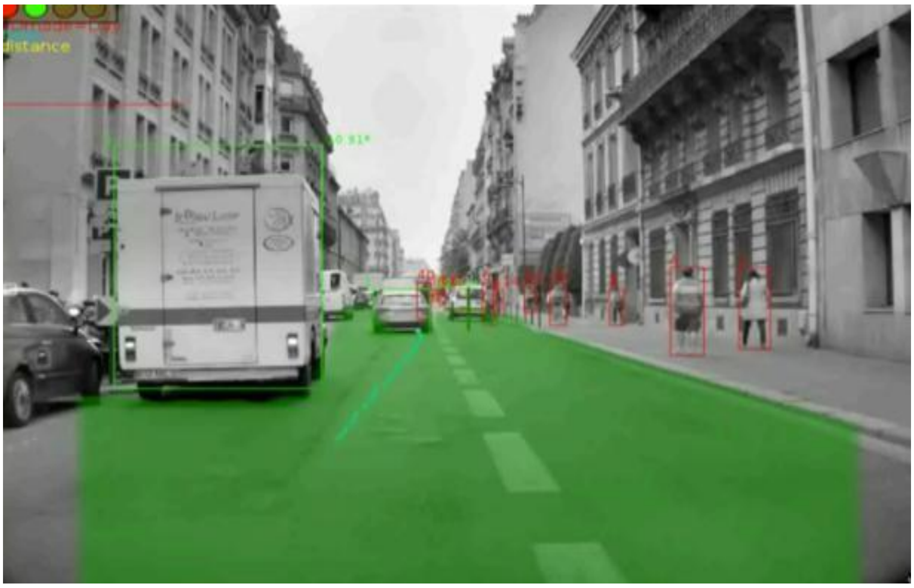

Figure 1. Vehicle and Pedestrian Detection [9]

4. Visual and radar assistance

Better mapless autonomous driving is autonomous driving that combines camera and radar data so that the camera and radar make up for each other's deficiencies making road data more accurate, more helpful and safer for the driver. Behind the front bumpers of the Tesla Model S and Model X are radars with a detection range of several hundred meters that can detect vehicles and moving parts. Behind the front bumpers of the Tesla Model S and Model X are radars with a detection range of several hundred meters that can detect vehicles and moving objects at considerable distances. Unfortunately, the radar cannot detect stationary objects such as driving lanes or inactive people. Camera systems are simple wide-angle cameras mounted on the front or roof of the vehicle that detects various objects, including cars, trailers, bicycles, pedestrians, and road signs [5]. There are multiple ultrasonic sensors and cameras mounted around the body of the Tesla, and the accuracy of the data can only be guaranteed if there is enough of it, as there is no guarantee that the individual data is correct. The most accurate set of data can only be obtained by combining and calculating the data from different sensors in different locations. These are from the vehicle itself to read out the data; also need to call some external data that can be obtained, such as GPS positioning and positioning of some satellites, so that you can determine the exact location of the body and in the field of vision outside the map data.

However, the route may not be able to find out on the map, but can be determined through the satellite to the general routes and road conditions, and only after understanding the data can you Only after knowing these data can know whether the road ahead can be passed, and only after determining whether it can be passed can smoothly use the vehicle to obtain data to ensure automatic driving. You can use GPS and satellite data to know whether the road ahead is smooth or congested, which can be used to choose a better route. After reading these data, it is necessary to integrate the data for calculation and utilization. The chip for Tesla's autopilot is Mobileye's Eye Q3 processor, built on a 40-nanometer process. The Eye Q3 processes data from satellite imagery, radar, ultrasonic sensors, and cameras to control steering and speed systems. Mobileye is the first Tesla electric vehicle to incorporate a digital neural network. Mobileye is the first Tesla electric vehicle to incorporate a digital neural network (DNN) responsible for four essential tasks: path planning, joint object detection, and signal recognition [9]. The final data obtained after processing the data through the processor will be fed back to the car machine and the safest and most appropriate feedback will be made so that the car can reach its destination in the safest and fastest way.

5. Conclusion

To summarize, autonomous driving is definitely a trend for future road driving, but it is ultimately different from a driver, so have to ensure the reliability of autonomous driving to the greatest extent possible. The recognition function of camera is the most basic and the closest to human drivers in automatic driving, and it can look around, record all the surrounding conditions and add calculations to reach the optimal conclusion. In rainy or foggy conditions, camera recognition may not be optimal because these conditions cause the field of view to become unclear, and the camera is just like the human eye if the vision is not clear, then it will not be able to observe the complete data, which is why errors and accidents occur. So if permitted, the addition of radar can better overcome this major drawback of the camera, because radar can be avoided in rainy and foggy conditions can be detected around the object, through the emission of signals and reflected back to the time to determine the position as well as speed.

But the machine will also have the time to make mistakes, such as camera damage or computer crash, so although the driverless seems relatively reliable, but once the failure will be the beginning of a big accident, so the optimization of hardware and software must be the first focus of the development of the current autonomous driving. At the same time, compared to pure camera autonomous driving, camera and radar with more reliable automatic driving, but also means that this will require more cost, so to solve the technical difficulties of the product and be able to mass production is also a very important goal, which can significantly reduce the production cost and thus make the overall cost of the car, so as to reduce the threshold of purchase.

References

[1] Prasath G D S, Poopathi M K R, Sarvesh P, et al. Application of Machine Learning Algorithms in Autonomous Vehicles Navigation System[C]//IOP Conference Series. Materials Science and Engineering. IOP Publishing, 2020, 912(6).

[2] Duchon A P, Warren W H, Kaelbling L P. Ecological robotics[J]. Adaptive Behavior, 1994, 6: 1–30.

[3] Temizer S, Kaelbling L P. Optical Flow Based Local Navigation[R]. MIT, Cambridge, Mass, USA, 2003.

[4] Sistu G, Leang I, Chennupati S, et al. Neurall: Towards a unified visual perception model for automated driving[C]//2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019: 796-803.

[5] Siam M, Elkerdawy S, Jagersand M, et al. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges[C]//2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1-8.

[6] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1520–1528.

[7] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[8] Larsen S A, Koren H, Solberg R. Traffic monitoring using very high-resolution satellite imagery[J]. Photogrammetric Engineering & Remote Sensing, 2009, 75: 859–869.

[9] Ingle S, Phute M. Tesla autopilot: semi-autonomous driving, an uptick for future autonomy[J]. International Research Journal of Engineering and Technology, 2016, 3(9): 369-372.