The Application of Multi-sensor Fusion Navigation in Unmanned Driving

Shichang Cui*

Shandong Experimental High School, Jinan, China

*Corresponding author: shichangcui131@gmail.com

Abstract. With the popularization of autonomous driving technology, multi-sensor fusion technology is becoming increasingly important in unmanned driving. This paper mainly studies two types of sensors, namely visual cameras and LiDARs, as well as two multi-sensor fusion algorithms, namely Kalman filtering and Simultaneous Localization and Mapping (SLAM). Compared with using a separate sensor, the two sensors can have advantages such as a wide measurement range, high precision, and reduced influence of weather factors. The Kalman filter can effectively reduce the error in navigation and positioning. In SLAM technology, by comparing visual SLAM technology with laser SLAM technology and analyzing the SLAM technology that fuses with other sensors, it can be concluded that multi-sensor fusion has the advantages of enhancing robustness in complex environments, optimizing positioning accuracy, and balancing real-time performance and global consistency. Therefore, a more in-depth study of the combination and fusion algorithms among various sensors can enhance real-time performance, accuracy and robustness. In this way, we can achieve true driverless technology sooner.

Keywords: Multi-sensor Fusion Navigation; Unmanned Driving; Application.

1. Introduction

With the development of navigation technology, autonomous driving technology is becoming increasingly common in automobiles. From no automation to full automation, the higher the level of automation, the higher the requirements for the accuracy and comprehensive perception of positioning. This leads to the three most important points in intelligent driving cars being real-time performance, accuracy, and robustness. However, there may be drawbacks such as misjudgment or malfunction in a single sensor. For instance, although millimeter-wave radar has strong penetration, its resolution is low, and it is difficult to distinguish the details of objects. Relying solely on cameras may lead to misjudgment of obstacle types due to target object occlusion, shadows, etc. Therefore, the necessity of multi-sensor fusion is highlighted.

To enhance the accuracy of positioning and the comprehensive perception capability, this paper adopts a method combining two types of sensors: LiDAR and visual cameras. The first sensor is LiDAR. LiDAR can emit laser beams and determine the distance to the target by calculating the flight time of the laser and the phase difference, thereby constructing three-dimensional environmental information. The second one is the visual camera. A visual camera is a device that captures visible light through a lens and converts it into image signals. It can record visual information such as colors, textures, and object shapes in the environment. Due to multi-sensor fusion, specific algorithms are needed to combine them. The first one is the Extended Kalman Filter (EKF). The Extended Kalman Filter is an algorithm that linearizes nonlinear systems. It can be approximated through Taylor expansion for state prediction and update, and is suitable for processing data from nonlinear sensors such as visual cameras. The second one is SLAM (Simultaneous Localization and Mapping) technology. SLAM technology is a technique that perceives the surrounding environment in real time through sensors and builds a map, while determining its own position on the map. In the field of autonomous driving, multi-sensor fusion technology is of great significance. Its development began with visual SLAM technology and laser SLAM technology, and then advanced algorithms such as VINS-Mono (camera and IMU tightly coupled), LIO-SAM (lidar and IMU), and R3LIVE (lidar, camera and IMU) emerged. These have achieved technological breakthroughs from filtering to factor

graph optimization, and from loose coupling to tight coupling, promoting the development of autonomous driving technology.

This article mainly analyzes the advantages and characteristics of two types of sensors: visual cameras and LiDAR. Then, it examines two primary multi-sensor fusion algorithms, namely Kalman filtering and SLAM technology, focusing on their principles, features, and the benefits of multi-sensor fusion.

2. Relevant Theories and Key Technologies

In normal driving, human drivers can use their senses to perceive the surrounding environment, make judgments, and then achieve the purpose of driving through a series of operations. However, in unmanned driving, especially in advanced intelligent driving, the process of perceiving the surrounding environment and making judgments needs to be accomplished by the vehicle itself. Therefore, when achieving unmanned driving technology, sensors such as visual cameras, lidar, millimeter-wave radar, GPS, and IMU play a crucial role. These sensors can perform vital functions such as determining the relative position in the environment, obstacle recognition, map construction, and path planning.

2.1. Visual Cameras

Similar to human vision, visual cameras can obtain high-resolution image data, and the data they output is more intuitive. Therefore, visual cameras are one of the main sources for vehicles to acquire external information. Compared with lidar, although it is greatly affected by environmental lighting, visual cameras still have many unique features. For example, visual cameras are low-cost, have mature related technologies, and have high image resolution. They can achieve functions such as recognizing color and texture information. The common visual camera image has about 2.07 million pixels, and the acquisition frequency is 24FPS, 30FPS, 60FPS, etc. Therefore, visual cameras can be used to recognize traffic lights, traffic signs, etc. At the same time, because of the high frame rate of visual cameras, they also have the characteristic of strong target tracking ability, and can achieve object tracking by observing the changes of objects in multiple frames of images [1].

Visual cameras have several main applications in unmanned driving. Firstly, they can recognize and detect objects. The image data captured by them can identify pedestrians, traffic lights, and other information, thereby understanding the surrounding environment. Xiangcun Kong et al. proposed a camera rotation strategy based on monocular active environmental perception, which can detect the blind spots of unmanned vehicles when the vehicle's yaw angle changes [2]. Secondly, visual cameras have the functions of positioning and mapping. By analyzing image information, they can estimate the changes in their own pose, thereby achieving the purpose of autonomous positioning and navigation in unknown environments. At the same time, visual cameras also have the function of target tracking. By analyzing the time series of images, they can obtain the position, speed and other information of the tracking target [3].

The working principle of a visual camera is to project three-dimensional image information onto a two-dimensional plane. The process to achieve this goal is the transformation of image data among four coordinate systems: the world coordinate system, camera coordinate system, image coordinate system, and the pixel coordinate system. It quantitatively describes imaging by projecting a point A in an image to the pixel coordinate system to obtain its corresponding pixel coordinates. Figure 1 illustrates the spatial positional relationship among the four coordinate systems [1][4].

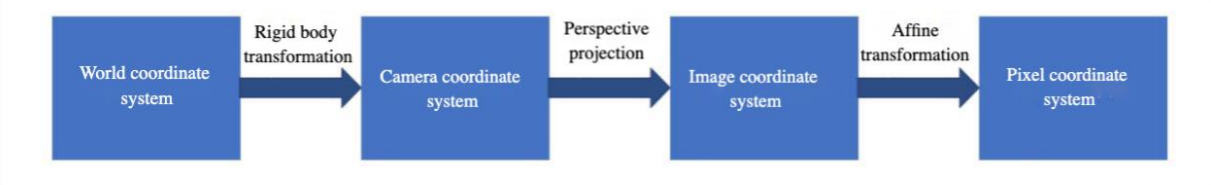


Figure. 1 Coordinate System Transformation in the Imaging Process

2.2. Lidar

The principle of lidar is to emit laser beams into the surroundings. When the laser hits an object, it bounces back and the return signal is captured by the sensor. By measuring the time it takes for the laser to return, the distance between the object and the lidar can be determined. Additionally, by measuring the reflected energy, amplitude, frequency, and phase of the reflected light spectrum of the object, the three-dimensional structure can be accurately extracted [5]. However, due to the shape characteristics of lidar, as the detection distance increases, the density of the point cloud decreases, resulting in a reduction in detection accuracy. Therefore, the effective detection range of various lidars varies. For example, 64-line lidar and 128-line lidar; the more line beams there are, the denser the point cloud information obtained, and the higher the resolution of environmental perception [4].

Based on the principle of lidar, it can be seen that although it has the disadvantages of high cost and limited measurement perspective, lidar also has unique advantages. Firstly, lidar can obtain higher-precision point cloud data through laser scanning. This enables it to provide centimeter-level precision measurement. Combined with a visual camera, it can make up for the lack of environmental distance information. Although the measurement perspective of lidar is limited, its detection distance is farther, and its perception of the environment is less affected by lighting conditions. Combined with a visual camera, it can make up for the defect that the camera is difficult to accurately detect objects due to the influence of lighting. On this basis, Tariq Hussain et al. proposed a system to detect and avoid all possible attacks on the radar system in order to establish a better safety mechanism for the radar system. The success rate was 94.57%, which proved the effectiveness of the system [6]. In addition, lidar can also achieve vehicle positioning. When generating data, lidar generates three-dimensional information. Combining this with point cloud data can obtain a three-dimensional map of the current environment. Thus, during the vehicle's driving process, lidar can match the currently collected point cloud data with the three-dimensional map, thereby achieving the purpose of vehicle positioning [1].

3. Kalman Filtering and Extended Kalman Filtering

Kalman filtering is a method for estimating the state of a system. Due to its efficient operation and stable performance, it is widely used in multiple fields, such as intelligent driving, navigation and positioning, and even in medical health. Therefore, the principle of Kalman filtering is to predict the current state of the target based on its state at the previous moment, and then update the state of the target by combining the current state observation value. To achieve these goals, it mainly consists of two stages: the prediction stage and the update stage. In the prediction stage of Kalman filtering, the state prior estimate value and covariance prior estimate value at the current moment are obtained through the state posterior estimate value and covariance posterior estimate value at the previous moment of the target. In the update stage of Kalman filtering, the state posterior estimate value and covariance posterior estimate value at the current moment are calculated by using the current state observation value, the state prior estimate value and covariance prior estimate value predicted[7]. Based on the characteristics of Kalman filtering, Guihua Miao et al. proposed a framework for forward collision warning by applying Kalman filtering and the YOLO algorithm. In real traffic scenarios, this framework has good performance and can accurately predict and alert to collisions[8].

The Kalman filter is applicable to linear Gaussian systems, and the fundamental characteristic of a linear Gaussian system is that the relationship between the system's state and the observed variables is linear. Therefore, in some nonlinear scenarios, the Kalman filter may not be applicable. In such cases, the Extended Kalman Filter (EKF) algorithm can be used. The EKF is a filter suitable for nonlinear systems. The core idea of the EKF is to first approximate the nonlinear function using a first-order Taylor expansion, that is, to linearize the motion equation and the observation equation. Then, it uses prediction and update steps similar to those of the Kalman filter to predict the state and calculate the Kalman gain and update the posterior probability distribution[9].

When using filtering methods to estimate the state of a rigid body, since the mean and covariance of the state need to be calculated at each state update, the computational load of the extended Kalman filter is proportional to the square of the dimension of the state vector. Therefore, when the scale of the environment being explored by the vehicle is complex, handling large-scale state vectors will pose challenges. This is because large-scale state vectors will lead to a significant increase in computational and storage requirements. This may result in insufficient computing resources, prolonged computation time, and excessive memory consumption, etc[9].

4. Research on SLAM Technology

SLAM technology can be classified into visual SLAM algorithms and laser SLAM algorithms based on different sensors, but the basic principles of the two algorithms are the same. Its algorithm framework is shown in Figure 2. Depending on the sensors, the front end is responsible for estimating the relative motion between two consecutive frames of visual images or point cloud images, extracting features and abstracting models from the data, and generating local maps. Loop closure detection is used to determine whether previously visited locations form a loop and to pass the relevant information to the back end for processing. The back end acquires the odometry pose transformation sent by the front end and the processing results of loop closure detection, corrects the data globally for consistency through nonlinear algorithms, and finally builds a map based on the optimized trajectory and the generated data [9][10].

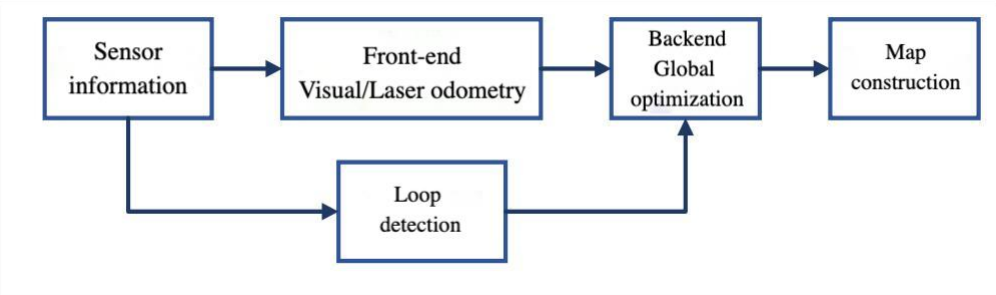


Figure. 2 The Classic SLAM Algorithm Framework

4.1. Camera-based Visual SLAM Algorithm

Visual cameras are important sensors in SLAM systems. They assist in pose estimation and map construction through continuous images. They can provide deep semantic features, and are low-cost, small in size, and widely used [11]. Cameras can be classified into monocular cameras, binocular cameras, and RGB-D cameras based on their construction differences. Visual SLAM can be divided into SLAM systems based on monocular, binocular, and RGB-D cameras according to the type of camera used. If classified by the way of processing image data, it can be divided into feature point methods and direct methods. The feature point method achieves camera pose estimation and mapping by extracting feature points in images for matching and calculation. On this basis, Jiawei Chen et al. introduced Shake_VINS, which improved feature tracking by 48% and reduced jitter offset by 30% when the camera was shaking. This solution improved positioning accuracy by 50% in the case of camera shaking [12]. The direct method does not rely on feature point extraction, but directly uses pixel information to calculate camera motion and position, and builds a semi-dense map. There is also a visual odometry method called "sparse direct method", which can combine feature point methods and direct methods, that is, use feature point methods for tracking, and then estimate the camera's motion and its own position, similar to the direct method [9].

4.2. Laser SLAM Algorithm Based on LiDAR

Lidar can be classified into single-line lidar, multi-line lidar, and solid-state lidar based on their construction principles. According to the point cloud images output by laser SLAM, laser SLAM algorithms can be divided into 2D laser SLAM algorithms and 3D laser SLAM algorithms. Both

types of algorithms are constantly being optimized in terms of real-time performance, accuracy and scene adaptability. Among them, 2D laser SLAM takes filtering or graph optimization as its core. It is widely used in service robots and focuses on indoor small-scale scenes. However, it also has limitations, such as the need to increase the number of particles when the scene is too large or the pose error is too large, resulting in excessive memory usage, and the particle resampling is prone to causing "particle dissipation". 3D laser SLAM takes feature extraction and matching as its core and is suitable for outdoor, large-scale environments. However, it also has limitations such as the lack of loop closure detection and the inability to solve large-scale cumulative errors [6]. To address this issue, Jiaping Chen et al. proposed an efficient LiDAR odometry enhancement strategy. They improved the system performance by enhancing high-quality features throughout the optimization process. For non-ground features, this method can enhance their influence in point-to-plane and point-to-line constraints. For ground features, this method can provide a unified "point-to-plane" constraint form for ground registration by fitting a local stable plane surface instead of discrete point correspondence, thereby improving the consistency of registration constraints. Ultimately, the performance gain is particularly significant in scenarios dominated by unstructured environments [13].

4.3. Multi-sensor Fusion SLAM Technology

Multi-sensor fusion SLAM technology achieves more robust and accurate localization and mapping by integrating the advantages of different sensors, such as the detailed textures from visual cameras, the geometric distances from LiDAR, and the high-frequency motion measurements from Inertial Measurement Units (IMUs), thereby enhancing real-time performance and environmental adaptability. Among them, the combination of visual cameras and IMUs can integrate the high-frequency motion information from IMUs with the visual features of cameras, compensating for the insufficient dynamic response of pure vision and improving global consistency. The integration of LiDAR and IMUs can optimize global pose and enhance real-time performance. There are various fusion methods that can combine LiDAR, visual cameras, and IMUs, and different fusion methods have their own advantages. Some of these fusion methods can improve robustness in complex environments, some can optimize positioning accuracy, and some can balance real-time performance and global consistency [9].

5. Conclusion

In driverless cars, a single sensor has drawbacks such as misjudgment and malfunction. This paper mainly studies two types of sensors, namely visual cameras and LiDAR, as well as two multi-sensor fusion techniques, namely Kalman filtering and SLAM. This paper also reveals the characteristics of these two sensors and the fusion algorithms, as well as the advantages of multi-sensor fusion. For the two sensors, although visual cameras have the advantages of low cost and high image resolution, they are easily affected by environmental lighting. LiDAR, on the other hand, has the advantages of high precision and being unaffected by lighting conditions, but it has a limited measurement angle and a high cost. The fusion of the two sensors can offer advantages such as a wide measurement range, high precision, and reduced influence of weather factors. For the fusion algorithms of the two sensors, Kalman filtering can estimate the system state through the prediction stage and the update stage, thereby reducing the error in navigation and positioning. Extended Kalman filtering can predict nonlinear data on this basis, enhancing its versatility. However, when the data scale is large, extended Kalman filtering may lead to insufficient computing resources or excessive memory consumption. In SLAM technology, visual SLAM can obtain more detailed map information, while laser SLAM can acquire real-time and high-precision map information. The combination of the two with IMU can optimize positioning accuracy, enhance robustness in complex environments, and achieve real-time performance and global consistency. The integration of multiple sensors with fusion algorithms can improve the real-time performance, accuracy, and robustness of navigation in unmanned driving. The

research results of this paper are of reference value to researchers studying multi-sensor fusion navigation and the characteristics of fusion algorithms. Multi-sensor fusion can achieve complementarity among sensors. Therefore, only by conducting more in-depth research on the combination of multiple sensors and fusion algorithms to improve real-time performance, accuracy, and robustness can true unmanned driving be achieved earlier.

References

[1] Wu Yi. Research on Automatic Calibration Technology of LiDAR and Vision System for Autonomous Driving [D]. University of Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2024.

[2] Kong X, Shi S, Yan B, et al. Research on Camera Rotation Strategies for Active Visual Perception in the Self-Driving Vehicles [J]. Actuators, 2024, 13(8): 317.

[3] Yang Yifan. Research on Autonomous Driving Environment Perception Method Based on Multi-Sensor Fusion [D]. Lanzhou University of Technology, 2024.

[4] Zhang Guoxin. Research on 3D Object Detection Algorithm in Autonomous Driving Scenarios [D]. Hebei University of Science and Technology, 2024.

[5] Cheng Xiaoxiao. Research on LiDAR SLAM Optimization Based on Multi-Sensor Fusion and Practical Application of Indoor and Outdoor Unmanned Driving [D]. Shandong University of Technology, 2024.

[6] Hussain T, Khan N M, Yang B, et al. LiDAR point cloud transmission: Adversarial perspectives of spoofing attacks in autonomous driving [J]. Computers & Security, 2025, 157: 104544.

[7] Wei Chao, Wu Xitao, Zhu Gengting, et al. Research on Obstacle Detection and Tracking for Unmanned Vehicles Based on Fusion of Visual Camera and LiDAR [J]. Journal of Mechanical Engineering, 2025, 61(02): 296-309+320.

[8] Miao G, Wang W, Tang J, et al. Vehicle Collision Warning Based on Combination of the YOLO Algorithm and the Kalman Filter in the Driving Assistance System [J]. Journal of Advanced Transportation, 2025, 2025(1): 1188373.

[9] Shen Xi. Research on Multi-Sensor Real-Time Localization and Mapping Technology for Autonomous Driving [D]. North China University of Technology, 2024.

[10] Qin Fuhong. Research on Improvement of SLAM Algorithm Based on Multi-Sensor Data Fusion [D]. University of Electronic Science and Technology of China, 2025.

[11] Zeng Ruiqi, Ji Xinchun, Wei Dongyan, et al. Fusion Application of Visual and Laser Information in SLAM [J]. Journal of Navigation and Positioning, 2025, 13(03): 116-129.

[12] Chen J, Zhang W, Wei D, et al. Optimized visual inertial SLAM for complex indoor dynamic scenes using RGB-D camera [J]. Measurement, 2025, 245: 116615.

[13] Chen J, Jia K, Wei Z. Small but Mighty: A Lightweight Feature Enhancement Strategy for LiDAR Odometry in Challenging Environments [J]. Remote Sensing, 2025, 17(15): 2656.