Active Obstacle Detection System Based on Video Recognition and Lidar Information Fusion

Jiang Yaodong

1Casco Signal Co., Ltd., Innovation and Incubation Center, Shanghai 200070. E-mail: jiangyaodong@casco.com.cn

© The Authors 2020

ABSTRACT

In terms of the requirements for obstacle detection in the rail transit application field, an architecture and implementation method for active obstacle detection system based on the fusion of video recognition and lidar information is proposed. The studies on the video recognition algorithms based on deep learning neural network and lidar for orbit area recognition, pedestrian vehicle recognition, and small foreign object recognition are analyzed, and the necessity of the fusion of video recognition and lidar data and the related key technical points are discussed. Through the tests on domestic metro and tram lines, the feasibility of the scheme is verified, and the technical parameters are optimized, which can effectively reduce the probability of accidents caused by foreign object intrusion.

In the field of rail transit, accidents caused by foreign object intrusion are common. In recent years, there have been serious accidents on metro lines such as vehicle collisions due to drilling through the metro tunnel, falling of the drill bits, and abnormal open accidents of the civil air defense doors. On tram lines, cases of pedestrians at level crossings and vehicle intrusion into the boundary are frequent.

Traditionally, the driver is responsible for looking ahead. When an obstacle appears in front of the driver, he will adopt braking measures timely to avoid the collision accident. However, because the driver is prone to fatigue after driving for a long time, which may cause problems such as slow response and negligence, there is still the possibility of accidents due to that there is no time to deal with the intrusion. With the continuous maturity of autonomous driving technology and the improvement of system reliability, “fully automatic driverless driving” has become the main goal of the metro control system. When there is no longer a driver in the driver’s cabin to keep lookout, there must be a system that can replace the driver’s eyes, timely sense whether there are obstacles in the space boundary in front of the traffic, make accurate identification, respond in time, and control the braking and deceleration of the train so as to prevent collision accidents. How to design and implement such an obstacle detection system is the subject to be studied and solved in this paper.
1 Current situation and problems

At present, in the rail transit industry, many relevant companies and universities have carried out relevant studies on the demand for obstacle detection. The achievements mainly include two categories, i.e., “passive” and “active”. Currently, “passive obstacle detection system” has been applied in some metro lines, where the “detection rod” is installed in front of the wheel set at a certain height from the rail. When the detection rod touches the obstacle ahead, it pushes the associated sensor and timely triggers the train “emergency braking” to prevent the accident from further expanding. The system is limited to passive and contact detection of obstacles in front of the vehicle. When the system detects an obstacle, the vehicle has touched the obstacle and the collision has already happened. It can only play a role in reducing the degree of accident damage but cannot prevent the accident in advance.

Another idea is “active obstacle detection”. It generally adopts sensors such as camera, millimeter wave radar, and lidar for non-contact detection of obstacles within a certain distance in front of the vehicle and makes response in time. At present, in the field of automatic driving obstacle recognition, the main sensors adopted include camera, millimeter wave radar, ultrasonic radar, lidar, etc. The detection range and working characteristics of these sensors are different (see Tab. 1), which need to be selected according to the characteristics of the application scenarios of rail transit.

| Sensor                  | Camera | Lidar | Millimeter wave radar | Ultrasonic radar |
|-------------------------|--------|-------|-----------------------|------------------|
| Detecting distance      | 300 m  | 300 m | About 200 m           | 5 m              |
| Resolution              | High   | High  | Low                   | Low              |
| Cost                    | Moderate | High  | Moderate              | Low              |
| Affected by light       | Great  | Seldom| No                    | No               |
| Influencing weather     | Rain, snow, dense fog | Rain, snow, dense fog | Rain             | No               |

Currently, in the field of autonomous driving, multiple sensors are generally adopted in combination. Camera can image in high definition with good detail resolution. By adapting the lens with different focal lengths (see Fig. 1), it can monitor a wide range or focus on a small range at a distance, and its cost is moderate. The disadvantages are that it is greatly affected by the environmental lighting conditions and is greatly affected by bad weather such as “rain, snow” and “dense fog”.

Lidar has the ability to scan quickly, and generally adopts the wavelength of 905 nm or 1 550 nm. It is safe for human eyes, has good collimation, high angular resolution, ranging accuracy up to centimeter level, and strong anti-interference ability, and can obtain various image information of the target (depth, reflectivity, etc.), as shown in Fig. 2. The detection range of lidar is related to the laser emission power, the size and reflectivity of the target object, etc. The current measured level of lidar is 240 m—300 m, and the cost is relatively high. In recent years, with the localization and mass production, the price is continuously falling. Lidar can also be affected in the conditions such as dense fog.

![Figure 1. Relationship between camera lens and viewing angle.](image-url)
Millimeter wave radar has been widely used in the automotive field, and also has some application cases in the rail transit field. Forward detection radar generally adopts 77 GHz. In recent years, millimeter wave radar has a longer detection range with the cost falling rapidly. Its resolution is not comparable to that of lidar, which is easy to be interfered with by clutter and thus produces false alarms. Millimeter wave radar is not affected by light conditions. However, heavy rain will interfere with its operation.

Ultrasonic radar is mainly used for short-range detection, and its cost is low. It is seldom used in the field of rail transit. The video recognition technology based on the visual imaging of camera has been widely used. In recent years, based on the rapid development of the deep convolutional neural network (DCNN) algorithm, there has been tremendous progress in the accuracy of image recognition, and many deep neural network models with high recognition accuracy have been derived, e.g., mobileNET, ResNET, and DenseNET. At present, some successful applications have been developed for the autonomous driving in the automotive field, e.g., the Autopilot system of Tesla.

However, the neural network models and algorithms studied and trained in the automotive field are not completely applicable to the rail transit field. One of the biggest differences is that rail transit has the characteristics of “fixed driving area” operation. In addition, the areas that need to be protected are mainly the range of “track area boundary”. For targets in other ranges, false alarm should not be given. The protective area in the automotive field is the “forward driving area”, which is generally a fan-shaped area extending to the front and a lane planned by lane lines. The scope is larger, and the concept of “boundary” is relatively loose. Therefore, the accurate identification and definition of “track area boundary” is an important basis for the active obstacle detection in rail transit and a key point to reduce misjudgment.

The industry counterparts have also studied on such topics. However, the results are generally not yet mature, most of which are in the experimental
development stage. In addition, some schemes used expensive sensors, which makes the product implementation difficult. Based on the above factors, a number of explorations and studies on the active obstacle detection technology in the rail transit field are conducted in this paper. In terms of the application characteristics of rail transit, an efficient and feasible active obstacle detection system is designed with the goal of product implementation.

2 System architecture scheme

This scheme is a non-contact active obstacle detection system based on the combination of video recognition and lidar. It has the ability to continuously detect and recognize obstacles within a range of 280 m in front of the vehicle, and can timely give an alarm and trigger the brake for foreign objects in the track area.

This system consists of vehicle-mounted computing host, interface equipment, lidar, far-focus camera, near-focus camera, inertial navigation device, display screen, etc. The system architecture is shown in Fig. 3. One set of this system is set in each of the driver’s cabins at both terminals of the train, and only the vehicle-mounted host in the driver’s cabin at the active terminal is enabled to output alarms and control commands. Related sensors are also installed at both terminals, respectively, which are connected to the local vehicle-mounted host, respectively. In order to enhance the recognition accuracy of distant targets, a tele-focus len camera and a close-focus lens camera are adopted. Lidar and camera complement each other to enhance the accuracy of target recognition. An inertial measurement unit (IMU) is used to measure the running attitude of the vehicle and assist the measurement of the speed and displacement of the vehicle itself.

![Figure 3. System architecture.](image-url)
The close-focus camera mainly covers target recognition within the range of 3 m—70 m, while the tele-focus camera mainly covers target recognition within a range of 20 m—240 m. For metro application scenarios, where the vehicle speed is fast and the line curve radius is large, it is necessary to focus on the recognition of long-distance targets ahead. A narrow viewing angle lidar can be used, covering the obstacle recognition in the range from 25 m to 280 m with a 15º viewing angle. For tram application scenarios, where the speed is relatively slow and the main security threat comes from the side, a wide-view lidar can be selected, covering the obstacle recognition in the range from 5 m to 200 m with a 60º viewing angle.

3. Research process and key technologies

For the identification of “foreign object intrusion boundary”, the critical first step is to accurately identify the “track area”. The second step is to identify the targets near the boundary of the track area. The third step is to determine whether there is a foreign object invading the boundary and determine whether alarm and related trigger actions are required according to the relevant rules.

In order to achieve the above goals, we have tried many methods. Initially, we adopt simple image processing algorithms to perform operations such as “edge search” and “Hough transform” on the image to extract and match its features, by which part of the track area on the image can be identified, as shown in Fig. 5. However, due to the diversity of the environment around the track, such simple machine vision algorithms may have recognition errors in some environment. The adaptability to complex environments is poor, and the effects by the external lighting conditions are very obvious, as shown in Fig. 6.
learning to solve this problem. Deep learning is a computing model composed of multiple processing layers, with which multiple abstract layer representations of data can be obtained through learning. This method significantly improves the effects of image recognition and detection.

Traditional machine learning has great limitations in processing natural data in the original form. It requires highly skilled engineers or experts to design corresponding feature extractors for different application targets, with which the original data (such as each pixel in the image) can be transformed into a suitable intermediate expression form or into a corresponding feature vector, so that the subsequent processing can be continued and the results can be output. The process and effect depend heavily on the skill level of the personnel, which requires repeated debugging. The workload is large, and the generalization ability is poor. If the processing object changes slightly, incorrect results may be obtained.

The advantage of deep learning is that there is no need for either manual design of feature extractors or additional data preprocessing. The machine will automatically learn to obtain transformed features. It is especially suitable for natural data processing with many changing patterns, and shows outstanding performance in generalization ability and robustness. Deep learning is a multi-layer representation learning method, with which the raw data are transformed into higher-level and more abstract expressions through hierarchical modules. When the learning system has enough such simple nonlinear modules, the improvement in depth ensures that more complex functions can be learned and more complex features can be recognized.

*Figure 7.* Characteristics learning and classification of neural network.

We need to use the deep learning training process to allow the neural network algorithm to correctly obtain the true features in the image and to perform correct inference recognition for similar contents outside other training set images. In this system, we use a semantic segmentation model based on Mask R-CNN to identify the orbital area. In order to reduce the computational intensity and improve the real-time performance, we only make a classification for the identification of the orbital area to distinguish the images, i.e., the image area is divided into only “track area” and “non-track area”. We adopt the video recorded on site as the original training material, carry out multiple rounds of labeling, training, and correction processes, and finally achieve the expected results, as shown in Figs. 8 and 9.
For the recognition of targets such as pedestrians, cars, and bicycles, we adopt the Yolo model for multiple rounds of additional training, and also achieve good results, as shown in Figs. 10—14. Near, medium, and long distance targets can all be accurately recognized. In the future, according to the needs of scenes, additional training can be used to improve the recognition accuracy and enable the neural network model to recognize more types of targets. Based on the coordinate relationship between the targets identified on the screen, we mark the targets invading the boundary of the track area, and determine the alarm level according to the running speed of the current vehicle and the distance between the vehicle and the target.

Although the video recognition scheme can complete the recognition of foreign object intrusion in most cases, due to its working principle, video recognition is only sensitive to “trained and classifiable” targets,
while cannot respond correctly to the targets which cannot be clearly classified. The video recognition technology itself is greatly affected by the surrounding light conditions, and the accuracy in ranging is not good. Note that the surrounding light conditions of the actual application scene in the metro tunnel are darker than those in normal scenes. In order to compensate for these shortcomings and further improve the reliability of the system, we use lidar to scan the front continuously with high frequency. Through the clustering processing of laser point cloud data, the obstacles in the front are detected and coordinated with the video recognition system to form more stable and reliable recognition results. Lidar and video recognition can form a good complement, greatly improve the accuracy of recognition, reduce missed judgment and erroneous judgment, and better meet the high reliability requirements of rail transit.

Lidar is the product combining traditional radar technology with modern laser technology. Its basic principle is to use the time-of-flight method (ToF). The detection signal (laser beam) is emitted to the measured target, and then the parameters such as the arrival time and the strength of the reflected or scattered signal are measured to determine the distance, the orientation, the movement state, and the surface optical properties of the target. The laser beam can accurately measure the relative distance between the edge of the object contour and the device in the field of view. This contour information forms a so-called “point cloud” and can be used to draw a 3D environment map with an accuracy of centimeters.

At present, the lidar host used in autonomous driving are mainly two types, i.e., “mechanical scanning” lidar and “solid state” lidar. The mechanical scanning lidar turns the laser beam from “line” to “plane” by continuously rotating the transmitter head, and arranges multiple laser beams in the vertical direction (i.e., 32 or 64 line radar) to form multiple surfaces to achieve the purpose of dynamic 3D scanning. However, it has the shortcomings of “big, heavy, and expensive”. There are no mechanical rotating parts in solid state lidar. Instead, electronic parts are used to rotate the emitted laser beams. At present, what have been produced and commercialized are mainly the micro-electro-mechanical system (MEMS) solid state lidar. In the MEMS, the mechanical mechanisms are miniaturized and electronically designed. The mechanical mechanisms with large volumes originally are integrated on a silicon-based chip through microelectronic process, which is conducive to mass production.

We use high-density MEMS solid-state lidar as the sensor of this system to ensure its long service life and good shock resistance. With an angular resolution of 0.1º, fine point cloud data can be formed for the obstacle perception. For objects with the reflectivity of 50%, the detection distance can be up to 300 m.

In order to make the lidar and the camera data coordinate system coincide, coordinate joint calibration should be carried out. The result of the coordinate calibration is to project the 3D coordinate system \((x_w, y_w, z_w)\), where the lidar is located, onto the 2D discrete coordinate system \((u, v)\), where the image is located. The transformation from 3D to 2D coordinates has undergone processes such as rotation and translation, projection, and coordinate dispersion.
(i) Rotation translation

The lidar and the camera are installed at different positions of the train, e.g., the position directly in front of the window glass. Their optical centers have deviations of $\Delta x$, $\Delta y$, and $\Delta z$ in the 3D coordinate system. Meanwhile, due to the installation error, their optical axes cannot be completely parallel, but there is a certain angle. This angle can be represented by three angles in the 3D space, i.e., $\psi$, $\theta$, and $\varphi$, which rotate around the $z$-, $y$-, and $x$-axes, respectively. Any point in the lidar can be expressed in the camera coordinate system as follows:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} + T$$  (1)

where $R$ is the rotation matrix, and $T$ is the translation matrix.

(ii) Projection

The 3D coordinates in the camera coordinate system are projected onto the 2D image plane by removing the depth information in the $z$-direction and retaining the information in the $x$- and $y$-directions.
\[
\begin{bmatrix}
X_p \\
Y_p \\
1
\end{bmatrix} =
\begin{bmatrix}
f & 0 & 0 & 0 \\
0 & f & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
X_c \\
Y_c \\
1
\end{bmatrix}
\]

(iii) Coordinate dispersion

On the 2D image plane \((x_p, y_p)\), images are imaged by the imaging sensor and discretely converted to the \((u, v)\) coordinates (where \(u\) and \(v\) are integers). Assume that the size of the imaging area is \(W_s\), \(H_s\), the resolution is \(W_s, H_s\), the size corresponding to each pixel is \(dx\) and \(dy\), and the center of the image is located at \((u_0, v_0)\). Then, the image coordinates of any point in the image coordinate system \((u\) and \(v\) are the pixel positions) are

\[
\begin{bmatrix}
u \\
v
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & u_0 \\
0 & 1 & v_0
\end{bmatrix}
\begin{bmatrix}
x_p \\
y_p \\
1
\end{bmatrix}
\]

(iv) Combined formula

The corresponding relationship between the point \(L_{xyz}\) in the lidar and the pixel coordinate point \((u, v)\) in the camera image can be obtained by combining the formulas in the above steps as follows:

\[
\begin{bmatrix}
u \\
v
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & u_0 \\
0 & 1 & v_0
\end{bmatrix}
\begin{bmatrix}
f & 0 & 0 & 0 \\
0 & f & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
R \\
T
\end{bmatrix}
\begin{bmatrix}
x_l \\
y_l \\
z_l \\
1
\end{bmatrix}
\]

In the actual calculation, the calibration board and the calibration software are used to calculate the relevant conversion formula. The calibration board can be imaged in both the lidar and the visible light camera, and the conversion matrix can be obtained by solving the correspondence relationship. Coordinate fusion is achieved through joint calibration, and the laser point clouds are combined with the video images, which can effectively compensate for the perception of unclassified targets and will not be affected by the ambient light.

This research subject has been carried out multiple rounds of experiments on Shanghai Metro Line 12, Shenzhen Longhua Line Tram, Chengdu Metro Line 9, and Shanghai Metro Line 10, and has been optimized for many times based on the test data.

4 Summary and prospect

In recent years, the video recognition algorithm based on the deep neural network model has been continuously innovating, and the accuracy of target recognition has been continuously improving. Rail transit has the characteristics of “fixed driving area”, which further reduces the requirements for model generalization. Through targeted and intensive training on the previously misidentified areas and targets, the recognition model can be effectively corrected, so that the accuracy of video recognition can reach a very high level.

However, it is undeniable that systems based solely on video recognition have low ranging accuracy and are greatly affected by lighting conditions. Meanwhile, since we cannot exhaust all types of foreign objects that may appear in the orbiting area, there is always missed judgment due to some “untrained and unclassified” foreign objects. The detection of obstacles ahead by lidar is only related to the laser point clouds reflected by the obstacles, and there is no need to identify what the obstacles are. Therefore, the obstacle detection by combined lidar and video recognition is an efficient and feasible complementary method. Through joint calibration, the information of both the lidar and the video recognition can be fused to make comprehensive judgment, which can effectively improve the recognition accuracy and reduce the probability of misjudgment and missed judgment.

The unremitting pursuit of “fully automated and unmanned driving” and “smart transportation” in the rail transit field has given birth to the extensive application of artificial intelligence technology in many fields related to rail transit. The development of deep learning and neural network algorithms is changing with each passing day, the accuracy rate is
constantly improving, and the real-time performance of calculations is also constantly improving. The large-scale localization of lidar has greatly reduced the price of lidar devices, making it more acceptable for customers. It is foreseeable that the active obstacle detection system described in this paper can be used as a beneficial supplement to the current communication-based train control (CBTC) system, to equip vehicles with their own “eyes”, especially for “driverless” application scenarios. The system will replace the lookout of drivers, undertake the tasks of obstacle detection and collision avoidance, and become an important active safety protection system.

References

Gowtham, R. (2016). Obstacle Detection and Avoidance Methods for Autonomous Mobile Robot. *International Journal of Science & Research, 5*(1), 1245-1249.

He, K., Gkioxari G, Dollar P, et al. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(2), 386-397.

Hua, C. (2019). Tensorflow and convolutional neural network: from algorithm introduction to project practice. Beijing: Publishing House of Electronics Industry.

Kim, J., & Cho, J. (2019). YOLO-based real-time object detection scheme combining RGB image with LiDAR point cloud. *Journal of Korean Institute of Information Technology, 17*(8), 93-105.

Lin, G., Liu, F., Milan, A., Shen, C., & Reid, I. (2019). Refinenet: Multi-path refinement networks for dense prediction. *IEEE transactions on pattern analysis and machine intelligence, 42*(5), 1228-1242. https://doi.org/10.1109/TPAMI.2019.2893630

Ople, J. J. M., Yeh, P. Y., Sun, S. W., Tsai, I. T., & Hua, K. L. (2020). Multi-Scale Neural Network with Dilated Convolutions for Image Deblurring. *IEEE Access, 8*, 53942-53952. https://doi.org/10.1109/ACCESS.2020.2980996

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE transactions on pattern analysis and machine intelligence, 39*(6), 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

Publisher’s note: Eurasia Academic Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) licence, which permits copy and redistribute the material in any medium or format for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the licence terms. Under the following terms you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorsed you or your use. If you remix, transform, or build upon the material, you may not distribute the modified material.

To view a copy of this license, visit https://creativecommons.org/licenses/by-nd/4.0/.