Development of Automatic Human Detection System for Forklifts using Image Recognition Technology

KENTA NAKAO[*1] TOMOHIRO MATSUMOTO[*1]
SATOSHI IIO[*1] KIICHI SUGIMOTO[*2]
SHOGO MINAMI[*3] KOJI SHIIZAKI[*4]

The operator is the key to safe forklift use. The operator must, at all times, be vigilant and remain aware of the surrounding environment. The operator must consistently look in the direction of travel to ensure that people, vehicles, and objects are not in the path of travel, and ensure that forklifts are being used in a safe and proper manner. Our forklifts include many safety features to enhance the operator's safe operation and use in intended applications. One such safety feature that we recently developed utilizing human detection technology is discussed in more detail below.

1. Introduction

We recently developed human detection technology that has the potential to enhance the forklift operator's ability to detect pedestrians and other persons present in the surrounding environment. This technology utilizes audible and on-monitor visual alerts to notify the operator that pedestrians or other persons are located in the forklift's vicinity. The technology works in "real time" using "deep learning" from images generated by various cameras mounted on the forklift (**Figure 1**). We are currently conducting additional testing under simulated operating conditions to continue to advance the development of the detection system for eventual field use.

This report presents an overview of the current status of our human detection system capabilities, the future prospects for the development and utilization of such technology, and a description of a prototype system that incorporates the technology.

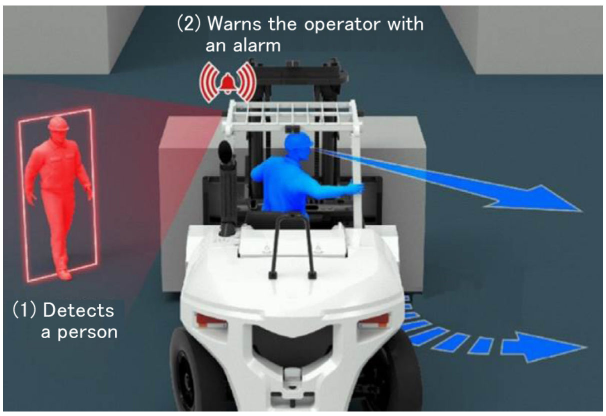


Figure 1 Safe operation support using human detection system

*1 CIS Department, ICT Solution Headquarters
*2 Chief Staff Manager, CIS Department, ICT Solution Headquarters
*3 Component and Advanced Technology Division, Engineering Headquarters, Mitsubishi Logisnext Co., Ltd.
*4 Acting Manager, Component and Advanced Technology Division, Engineering Headquarters, Mitsubishi Logisnext Co., Ltd.

2. Overview of human detection technology

2.1 Human detection using deep learning

In recent years, machine learning methods using multi-layered neural networks, known as "deep learning," have been applied to various fields and have achieved great results. Particularly, in the field of image recognition, convolutional neural networks (CNN) that extract planar feature quantities by adding various convolutions to the network have been actively researched. In object detection that locates a specific object from a given image, several network models[(1)(2)] with high detection performance have become the mainstream. This is a significant improvement in performance compared with conventional image recognition (utilizing engineer-designed feature quantities and identification logic). These new models are created for the purpose of general object detection, including "human" detection. For the application of CNN to forklifts, both processing speed and detection performance for a moving object are required. Therefore, tests were performed using a single shot multibox detector (SSD) with a good balance of performance and high speed.

2.2 Increasing processing speed

Deep learning is composed of multi-layered networks to achieve high performance, and it is essential to use GPUs which specialize in parallel computing for high-speed processing. However, GPUs typically used for deep learning are not suitable for onboard environments use such as on forklifts due to their size and power consumption. Although certain onboard-use GPUs are commercially available featuring a small size and low power consumption, they are inferior in terms of parallel computing ability, so increasing the processing speed is essential to executing deep learning in real time. Therefore, we evaluated SSD based on a method called Mobilenet[(3)], which divides a convolution operation that is repeatedly executed in deep learning to reduce the amount of calculation and increase the processing speed while suppressing performance degradation. As a result, the processing speed was significantly improved, and the processing time, which had been 200 ms per image when processing using an ordinary SSD with an onboard GPU, was improved to 59 ms per image.

2.3 Improvement of human detection performance

Next, we worked on improving the process of human detection. As mentioned above, SSD is a model for general object detection, and learns various daily scenes, which are not always the same as the environments in which forklifts are used (for example, in a factory or port area, etc., as illustrated in **Figure 2**.) For this reason, the background of a scene in which there is a person is different from the existent learning data, and there is some concern about the possibility of "overlooking" (unable to detect a person) and "false detections" (detecting a non-human as a human being). Therefore, to achieve sufficient detection performance, it is necessary to learn networks using images of actual environments in which forklifts are used. Thus, we mounted cameras on test forklifts and operated them in factories to collect image data under various situations. The collected images were organized as correct data for learning by labelling, and as a result of letting SSD relearn by using these data, the incidence of both overlooking and false detections were reduced compared with before the relearning. We believe that optimization in actual environments has advanced.

Figure 2 Difference between learning data for general object detection and applicable actual environment

3. Overview of prototype system

3.1 System configuration

Our human detection system monitors the area surrounding a forklift (as shown in **Figure 3**) by incorporating our newly developed human detention technology into the forklift's on-board GPU. This system is equipped with four short-range monitoring cameras for detecting a person close to the forklift and two long-range monitoring cameras for detecting a person farther away from the forklift (in the then-current direction of travel). Images taken by each camera are immediately input to the on-board GPU, and human detection processing is automatically executed. When a person is detected, an alarm sounds from a speaker installed on the forklift, and an image of the particular camera that detected the person and a warning are displayed on the monitor screen. All of the devices comprising the human detection system consist of components with low power consumption, and, therefore, are completely powered by the forklift's battery (with no additional power source requirements).

Figure 3 Prototype human detection system for forklifts

3.2 Automatic camera switching

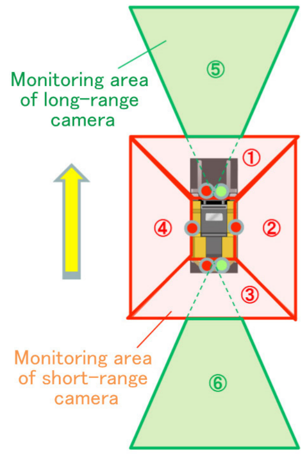
Since forklifts have enhanced, "tight" turning ability, it is necessary to monitor the area surrounding a forklift at all times during travel, including during low speed travel and maneuvering. For this reason, four short-range cameras are utilized in addition to long-range cameras in the front and rear of the forklift.

In order to effectively process all 6 camera images simultaneously in real time, many on-board GPUs would normally be required, resulting in the issues of power consumption and installation space. To overcome this challenge, We determined that by continually acquiring the forklift status through the Controller Area Network (CAN) and then selecting the camera images to be captured, the system was able to process images using a single on-board GPU. For example, as illustrated in **Figure 4**, when the forklift stops or moves slowly, there is a certain time margin for detection, so the four short-range cameras are processed sequentially. When the forklift moves at a higher speed, decreasing the time margin for detection, only one long-range camera (in the direction of travel) is continuously processed. In this way, even with the processing capacity of a single on-board GPU, the system is able to monitor the required areas at the required processing speed.

3.3 Warning to operator

When a person is detected in a camera image, the forklift operator is alerted immediately. This prototype system uses both auditory and visual methods for this alert. Speakers are installed at the four corners of the ceiling of the forklift cabin, and when a person is detected, an alarm sounds from the speaker. The alarm sounds from the speaker closest to the direction in which the person is detected, as shown in **Figure 5**. This, in turn, makes it easier for the operator to visually confirm the person's presence.

In addition, as noted above, when a person is so detected, the camera image is also displayed on a monitor installed on the forklift, and the location of the person is highlighted by framing. When more than one person is detected simultaneously by multiple cameras, the priority of the camera image to be displayed is determined according to the specific forklift use at the time (such as the direction in which the forklift is moving, etc.).



Direction the forklift is moving	Moving speed	camera images to be processed	Processing cycle per camera
Stop	—	①→②→③→④→①...	4fps
Forward	Low	①→②→③→④→①...	4fps
Forward	High	⑤	16fps
Backward	Low	③→④→①→②→③...	4fps
Backward	High	⑥	16fps

Figure 4 Automatic camera switching according to forklift conditions

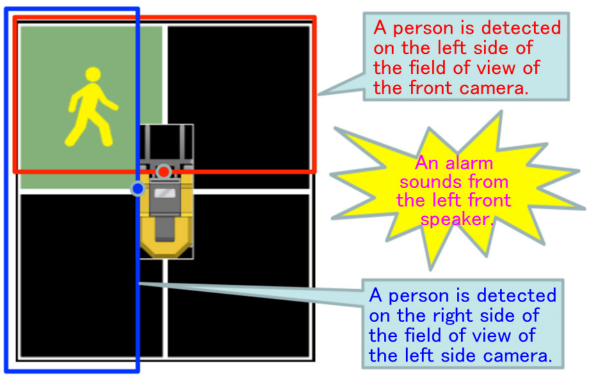


Figure 5 Division of monitoring area and warning to driver

4. Conclusion

We are utilizing new technological developments to enhance the safe operation of forklifts. We currently plan to field-test a prototype system under a variety of environments and conditions to further improve the performance of the system, with the ultimate goal of releasing the system for field use in the near future.

References

(1) Wel, L. et al., SSD: Single Shot MultiBox Detector, ECCV. (2016)
(2) Redmon, J. et al., You only look once: Unified, real-time object detection, CVPR. (2016)
(3) Andrew, G. et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Application, CVPR. (2017)