WILEY | Hindawi

Research Article

Detection for Dangerous Goods Vehicles in Expressway Service Station Based on Surveillance Videos

Kai Huang [ID] [1] **and Qinpei Zhao** [ID] [2]

[1]*China Communications Construction Company Ltd, Beijing, China*
[2]*Tongji University, Shanghai, China*

Correspondence should be addressed to Qinpei Zhao; qinpeizhao@tongji.edu.cn

To improve the safety capabilities of expressway service stations, this study proposes a method for detecting dangerous goods vehicles based on surveillance videos. The information collection devices used in this method are the surveillance cameras that already exist in service stations, which allows for the automatic detection and position recognition of dangerous goods vehicles without changing the installation of the monitoring equipment. The process of this method is as follows. First, we draw an aerial view image of the service station to use as the background model. Then, we use inverse perspective mapping to process each surveillance video and stitch these videos with the background model to build an aerial view surveillance model of the service station. Next, we use a convolutional neural network to detect dangerous goods vehicles from the original images. Finally, we mark the detection result in the aerial view surveillance model and then use that model to monitor the service station in real time. Experiments show that our aerial view surveillance model can achieve the real-time detection of dangerous goods vehicles in the main areas of the service station, thereby effectively reducing the workload of the monitoring personnel.

1. Introduction

An expressway service station is an area that is established along the expressway as a place for passengers to rest and refuel. Service stations usually include a parking area, service area, and gas station [1]. Expressway service stations were initially built for public service, but with the development of national expressways, the increase in the traffic flow has required service stations to improve their management and service capabilities, provide better services, and ensure the safety of both patrons and personnel.

The service station acts as a logistics node on the expressway by providing services for freight vehicles. As a result, many dangerous goods (DG) vehicles enter and leave the service station. According to "GB 6944-2012 Classification and Code of Dangerous Goods, a standard proposed by the Chinese Ministry of Transport [2], "dangerous goods" refer to chemical products with hazardous properties, and the term "DG vehicles" refers to specific vehicles which

transport those dangerous goods by road. When a DG vehicle is involved in an accident at a service station, secondary accidents, such as toxic gas leaks and chemical explosions, can cause significant damage to surrounding facilities and personnel. Therefore, improving the risk control capabilities regarding DG vehicles can significantly improve the safety of service stations.

Service stations need to guide vehicles into the station and monitor the situation in the station and take some response when an accident occurs, or determine the posed risk of an accident [3]. Therefore, surveillance video systems are typically installed in service stations to monitor the situation in the station [4]. Usually, multiple surveillance videos are playing on one screen for surveillance personnel to observe. However, surveillance videos are acquired by cameras set up in different locations within the service station, and there is no continuity between each video, which reduces the efficiency and effectiveness of monitoring management. Because surveillance personnel need to watch

the surveillance videos from different perspectives simultaneously, tracking relevant targets within them can fatigue the personnel; moreover, this can increase the likelihood of them losing track of the target when the target passes between the frames of two surveillance cameras. To improve the efficiency of safety management in the service station, a new method is required to reduce the workload of surveillance personnel.

Computer vision technology is often used to reduce the workload in the field of traffic monitoring [5, 6] and the station monitoring [7]. This technology could recognize and locate some targets in the image. Due to the high recognition success rate and high positioning accuracy, it has many applications in traffic monitoring [8] and industrial measurement [9].

Compared with the vehicular sensor networks [10] and in situ technologies [11], computer vision is more responsive, and it is easy to install and operate [12]. However, computer vision has its limitations in traffic monitoring. The sight of cameras usually on the diagonal to the road and the installation pose between each camera are not the same, so it is difficult to use a unified method to positioning targets in different surveillance videos [13]. There is also a big difference between the surveillance area of service station and traffic monitoring. The service area need to cover a large area; then, it needs more cameras to minimize the blind area. But the surveillance area of traffic monitoring usually is road, which is narrow and long, and it does not need too much cameras. The tasks of the two are also different. On the one hand, the service area needs to guide and adjust the position of DG vehicles and make sure its parking position meets safety requirements; then, it has the requirement of positioning. On the other hand, traffic monitoring only needs to record accident of something else, and it is not positioning requirements (a road segment is sufficient). According to this, there needs a method to combine the information of different cameras, to realize the positioning function of the vehicle in the service station.

With the development of image processing technology, image stitching technology has been widely used in video surveillance applications [14]. Image stitching is a method that is used to stitch multiple images from different perspectives into a single image [15]. By stitching the feeds together from several surveillance cameras, this technique can significantly reduce the workload of surveillance personnel. However, traditional image stitching algorithms are based on matching feature points between images, such as SIFT [16], SURF [17], and ORB [18]. These methods are suitable for some images with large overlapping areas but not for the surveillance video systems typically found in service stations. This is because the traffic surveillance video equipment is more concerned with maximizing the sight area using a limited number of cameras, and therefore, the installation positions of the cameras are more scattered. Consequently, there is a lack of overlapping areas between the different video images. These reasons make it difficult to use traditional image stitching algorithms to stitch service station surveillance videos.

The basic principle of camera imaging is the process of mapping image information using the world coordinate system into the image coordinate system after perspective transformation, which can cause some deformation in the image. This process is called the perspective mapping (PM) of the image. A typical example of PM is that if you take two hypothetical objects of the same size, the one closer to the camera will appear larger than the other in the image.

The most common method for eliminating the perspective effect of an image is inverse perspective mapping (IPM) [19], which is the inverse process of PM. By calculating the transformation matrix of the camera, the IPM maps the information from the image coordinate system onto the world coordinate system. In the field of road traffic, IPM is often used to obtain an aerial view image of the surveillance area from surveillance videos, such as the lane detection method proposed by Oliveira et al. [20] and the urban road marking detection method proposed by Li et al. [21]. In addition, because the aerial view image retains the target's horizontal position information, it is often used for positioning and tracking targets, such as vehicle trajectory extraction methods based on aerial videos [22] and the traffic monitoring based on UAV [23]. Because the installation of surveillance cameras in service stations is similar to their installation in traffic areas, it may be feasible to use IPM to process surveillance videos in a service station.

Additionally, the development of convolutional neural network (CNN) technology has greatly improved the capabilities of vision-based target detection technology [24]. Using the deep-level image information extracted by the convolutional network, CNN detection can quickly extract and classify complex features in an image and then use those features to recognize specific targets. CNN detection also has many applications in traffic detection, such as the vehicle detection method [25, 26]. Because there is a large amount of image information contained in surveillance videos, using CNN detection to detect vehicles in these videos is a way to improve surveillance management efficiency.

In summary, this study proposes a surveillance video processing algorithm based on IPM and CNN detection. This method is best used in the expressway service station, and it is ideal for reducing the workload of surveillance personnel, thereby improving the efficiency of safety management in the service station. The system functions by stitching surveillance videos and automating the detection of dangerous goods vehicles in the service station. The organization of this paper is as follows. Section 2 describes the method of establishing the positioning model based on the surveillance video as well as the detection method for dangerous goods vehicles. Section 3 presents the details of the experiment and the results, and Section 4 presents the conclusion.

2. Materials and Methods

2.1. Surveillance Cameras in the Service Station. The method proposed in this paper uses existing surveillance cameras to capture the image information of the relevant areas. A representation of the common installation of surveillance cameras in a service station is shown in Figure 1, wherein there are several closed-circuit television (CCTV) towers installed in the service station, each of which is equipped with surveillance cameras with different poses, as shown in Figure 2. To ensure that the sight of these cameras can completely capture the main area of the service station, each area is usually captured by several cameras, as shown in Figure 3. There will inevitably be some overlapping areas in the field of view between the cameras, as these overlapping areas provide a prerequisite for ensuring the continuity of the surveillance videos.

2.2. IPM Transformation of Surveillance Video. IPM is applied to remove the perspective effect from a surveillance video and to remap its image into an aerial view image. We used the IPM method based on the homography matrix [27], which uses four corresponding points in each image to obtain the transformation matrix for IPM, and it has good operability and does not require the measurement of the pose parameters of the camera.

Consider that there is a point $P = (X, Y)$ in the image and $Q = (X, Y)$ is the corresponding point in the world coordinate system (supposing that all points are on the same plane in the world coordinate system). Then, the transformation of these points is calculated, as shown in equations (1), (2), and (3). The 3×3 matrix H represents the homography matrix, and this transformation is called homography:

$$\begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix} = H_{3 \times 3} \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{1}$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \tag{2}$$

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \frac{1}{Z_a} \cdot \begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix}. \tag{3}$$

The calculation of the homography matrix is shown in equation (4), where (x_i, y_i) is the point i in the image coordinate system and (x_i, y_i) are the corresponding points of point i in the world coordinate system. The homography matrix H can be calculated by calculating h_{ij} in the equation:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x_1' & -y_1 x_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2 x_2' & -y_2 x_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3 x_3' & -y_3 x_3' \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4 x_4' & -y_4 x_4' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y_1' & -y_1 y_1' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2 y_2' & -y_2 y_2' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3 y_3' & -y_3 y_3' \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4 y_4' & -y_4 y_4' \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = h_{33} \begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \\ y_1' \\ y_2' \\ y_3' \\ y_4' \end{bmatrix}. \tag{4}$$

The transformed image is shown in Figure 4. Because the transformed image has marked image distortions and only a part of the image of a service station (such as the road area and the parking area) needs to be monitored, we cropped the transformed image, leaving only the necessary target areas.

2.3. Aerial View Image Stitching of Target Areas. To generate a service station surveillance model, it is necessary to establish a standard reference system based on the structure of the service station. Because the transformed image has a corresponding area in the service station, splicing these images into a background image that includes structural information of the service station (such as an aerial view image or a structure map) can provide a positioning reference for the detection targets in the image. We used an unmanned aerial vehicle (UAV) to take an aerial view image of the entire service station for use as the background image, which provided a position reference for the surveillance model.

After the IPM transformation of the target areas in each surveillance video, we stitched the aerial view image of the target areas in the background image, as shown in Figure 5. The different colored areas in Figure 5 represent the transformed images taken by different surveillance cameras, and the background model is the service area obtained by aerial photography. Because there is generally no uniform standard for the installation poses of the cameras, it is necessary to manually adjust the target area size during the stitching process to ensure that it coincides with the corresponding area in the background model.

3. Dangerous Goods Vehicle Detection Based on SSD

According to the GB 6944-2012 Classification and Code of Dangerous Goods," dangerous goods are generally defined as chemical products with hazardous properties. Such products should use tank containers for road transportation [28]. Therefore, our proposed method detects DG vehicles in the image by detecting the tank containers, as shown in Figure 6.
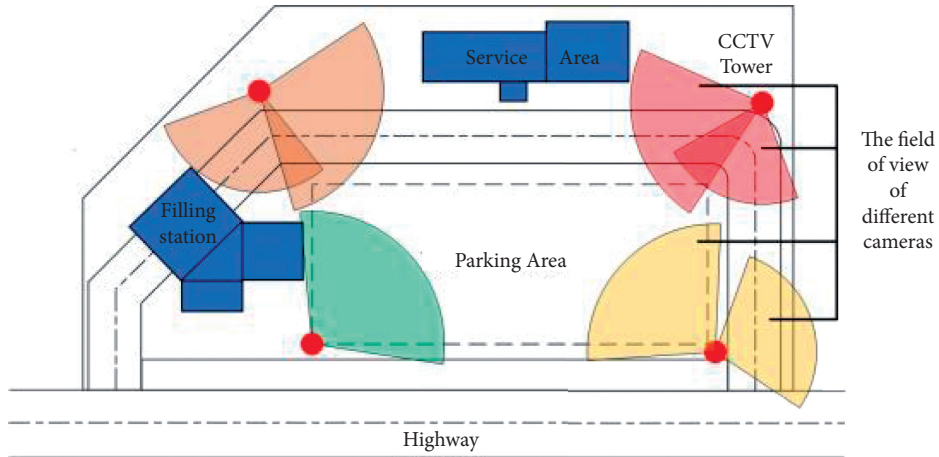
FIGURE 1: Installation of the CCTV towers in the service station.
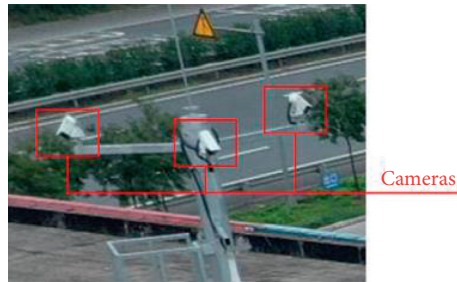

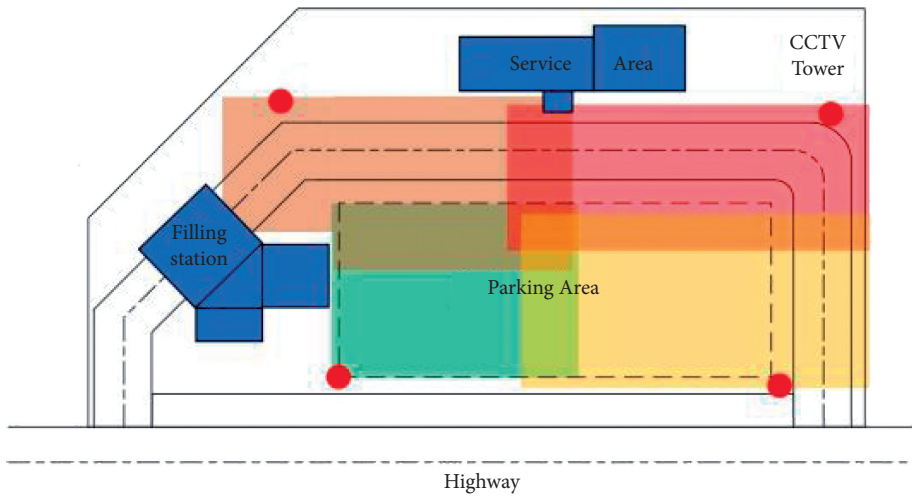FIGURE 2: Installation of the surveillance cameras in the CCTV tower.

FIGURE 3: Overlapping areas of different cameras.

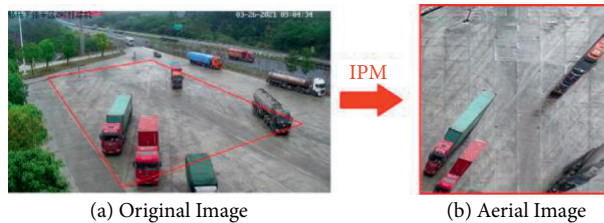(a) Original Image   (b) Aerial Image

FIGURE 4: IPM transformation of the surveillance image. (a) Original image. (b) Aerial image.
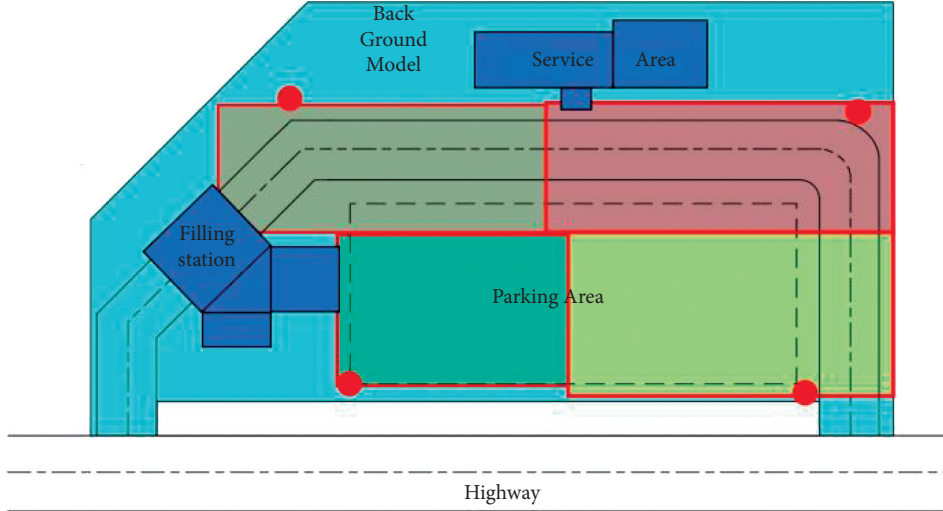
FIGURE 5: Stitched aerial view image of the service station.

FIGURE 6: Tank containers and detection feature area.

The Single Shot MultiBox Detector (SSD) [29] is a one-stage detection CNN model that uses anchor boxes of different scales and ratios to evenly sample the image and then uses the CNN layers to extract the features for classification and regression. This design improves the calculation speed of SSD compared to traditional two-stage methods, such as Fast R-CNN [30]. The loss function of the SSD is combined with the confidence loss and localization loss, as shown in the following equation:

$$L(x, c, l, g) = \frac{1}{N} \left(L_{\text{conf}}(x, c) + a L_{\text{loc}}(x, l, g) \right). \quad (5)$$

To match different objects, SSD sets multiple anchor boxes in each feature layer, as shown in equation (6), where $s_{\max}$ and $s_{\min}$ represent the maximum scale and minimum scale ratios of the anchor boxes to the image, respectively, and m is the number of feature layers:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1), \quad k \in [1, m]. \quad (6)$$

Furthermore, the SSD matches features by setting the aspect ratios. For an anchor box with a specific aspect ratio a_r under the scale s_k, the width and height can be calculated using the following equations:

$$w_k^a = s_k \sqrt{a_r},$$
$$h_k^a = \frac{s_k}{\sqrt{a_r}}. \quad (7)$$

Considering that the detected images are captured by different cameras, these cameras are installed in different positions and poses, and the features in these images have different sizes and ratios; the variable anchor design using SSD can effectively match such features. We set different anchor aspect ratios for the images captured from different cameras. For the image sequence P_i, which is captured by camera I, we selected several aspect ratios a_r to form the ratio sequence R_i to sample in different image sequences, as shown in the following equation:

$$R_i = \{a_1, a_2, \ldots, a_n\}. \quad (8)$$

The detection process for the DG vehicle is shown in Figure 7. Because the image after IPM transformation will have serious image distortion, which reduces the success rate of DG vehicle detection, we designed the vehicle detection process to occur before the IPM transformation process. However, because the image after the IPM transformation is cropped, the detection areas only need to include the cropped target areas, which correspond to approximately 30–60% of the original image. Therefore, for each camera i, the detection area ROI_i is divided to reduce the calculation time. The division of the ROI_i is shown in Figure 8, wherein the size of the ROI_i should be reduced as much as possible, and it should include the entire surveillance area.

It is necessary to note that the coordinates of the SSD detection result are also required for the IPM transformation to conform to the coordinate system of the transformed images. Similar to the IPM homography in the image, assuming that the center point of the detection result is $D_0 = (X_{d0}, Y_{d0})$, the calculation of the transformed point $D_1 = (X_{d1}, Y_{d1})$ is shown in the following equation:

$$\frac{1}{h_{31} h_{32} h_{33}} \cdot \begin{bmatrix} X_{d1} \\ Y_{d1} \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{bmatrix} X_{d0} \\ Y_{d0} \\ 1 \end{bmatrix}. \quad (9)$$
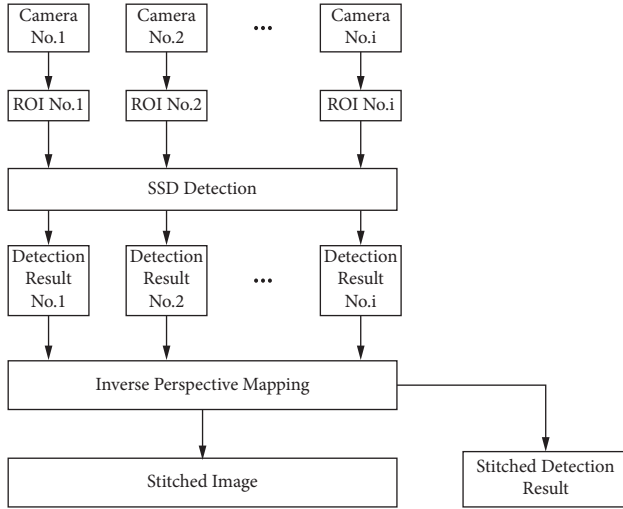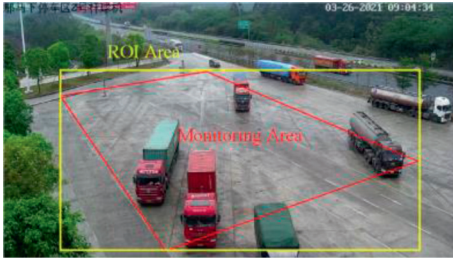
Figure 7: Detection process.

Figure 8: ROI area divide.

4. Experiment and Discussion

4.1. Surveillance Model. To verify the feasibility of the surveillance model establishment method, we used the surveillance data from an expressway service area to build a surveillance model based on Unity, as shown in Figure 9.

The surveillance model is composed of several surveillance video components and a background model. Each surveillance video component is an aerial view image transformed from the surveillance video, and the background model is captured by aerial photography. The different color points in the image in Figure 9 are the CCTV towers, and the color of each box represents the area in which the surveillance video is captured by the CCTV tower of same color. Figure 10 shows the DG vehicle detection results of the surveillance model.

According to equations (10) and (11), we calculated the overlapping area of the parking area covered by the surveillance model, and the overlapping area of the parking and road areas covered by the surveillance model (the area of each part was divided by the traffic index line). The result is that A_{sp} reached 93% and A_{spr} reached 71%:
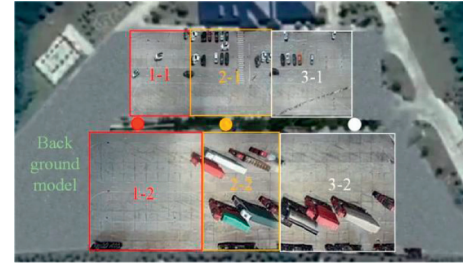

Figure 9: Surveillance model.

Figure 10: Detection result in the surveillance model.

$$A_{sp} = \frac{\text{Area of surveillance model}}{\text{Parking Area}}, \tag{10}$$

$$A_{spr} = \frac{\text{Area of surveillance model}}{\text{Parking Area} + \text{Road Area}}. \tag{11}$$

In this case, we combined 6 surveillance videos into 1 surveillance model. It means that the parking area, which required 6 screens to display, now only needs 1 screen, and the position of the vehicles can be observed more intuitively in the surveillance model. Even if the recognition function is not added, this greatly improves the efficiency of surveillance operations.

However, due to the distortion of images in the monitoring model, the positioning cannot achieve high positioning accuracy. But depending on the traffic index line of the parking area, it is possible to locate the approximate area of vehicles.

4.2. DG Vehicle Detection. Training the SSD required 2,000 images as training samples, which were captured from the surveillance cameras mentioned in Section 4.1. These images contained views of the tank container trucks taken from different angles.

For the test process, we used the equipment configured, as shown in Table 1, to test the performance of SSD detection.

For the evaluation of SSD detection, we used 200 images captured during the daytime and 200 images captured at night. The original resolution of these images was 2560×1440, but because the images were cropped, the test images had several smaller resolutions. We used intersection over union (IOU) as the criterion for detection success, and

TABLE 1: Equipment used for the DG vehicle detection test.

Operating system	Machine learning library	CPU	GPU
Ubuntu 18.04	Pytorch 1.3.0	Intel i7-6700	Nvidia GTX970

TABLE 2: Daytime and nighttime detection results.

Time	AP	Success rate (%)	False alarm rate (%)	Missing alarm rate (%)	Speed (ms)	Average image size
Day	0.892	97.8	0	2.2	20.3	1354 * 965
Night	0.753	84.5	1.5	14		

its calculation is shown in equation (12). Considering that the positioning operation in this method does not require high positioning precision, when the IOU value exceeded 50%, we determined the detection as successful. AP is the main parameter of the object detection algorithm. Its calculation is shown in equation (13). The evaluation results are presented in Table 2:

$$IOU = \frac{\text{Area of overlap}}{\text{Area of union}}, \qquad (12)$$

$$Precision = \frac{N(\text{True Positivers})}{N(\text{Total Objects})}, \qquad (13)$$
$$AP = \frac{\sum \text{Precision}}{N(\text{Total Images})}.$$

Compared with the detection results for daytime samples, the detection result for nighttime samples was greatly reduced. The main reason is the image distortion caused by insufficient light.

4.3. Discussion. The experimental results of the surveillance model demonstrate that, with a reasonable selection of the target areas of each surveillance video, the stitched surveillance model can include the main function areas (e.g., the road and parking areas) of the service station. In addition, multiple surveillance videos were successfully spliced into one video, which greatly reduced the workload of surveillance personnel. However, a significant problem was observed in the surveillance model. When a vehicle passes between two parts of the target area, its appearance is severely deformed, as shown in Figure 11, or the vehicle temporarily disappears from the frame. This is caused by the differences in installation positions and shooting angles between cameras, and it is not an overwhelming concern because the image of the vehicle returns to normal when it completely enters the viewing range of another camera.

The SSD detection experiment showed that SSD detection can effectively detect DG vehicles in the video, in most cases. However, because of limitations in the camera installation positions and poses, it was difficult to detect occluded targets at specific locations. Figure 12 shows an example in which there are four DG vehicles in the image, but only the one marked by the green frame is detected, because it is not occluded, and the undetected vehicles are indicated by the red frame. In this case, the camera captured the image of the parking area, and when several vehicles are parked in the

FIGURE 11: Vehicle image deformations' sample.

FIGURE 12: Detection failure sample.

parking area, some vehicles were occluded by other vehicles. Those occluded vehicles will be difficult to be detected by the SSD detector, as shown in the left of Figure 12. Additionally, the head of a truck can occlude the features of a tank container, as shown in the right of Figure 12.

5. Conclusion

This paper proposes a method based on image stitching and SSD detection for the automated detection of DG vehicles in an expressway service station. The detection data are based on existing surveillance camera data from a service station. Our experiments demonstrated that using surveillance cameras installed in a typical manner, the stitched surveillance model can cover the main traffic areas in the service station, thereby greatly reducing the workload of surveillance personnel. SSD detection also reduced the difficulty of DG vehicle recognition and positioning and improved the efficiency of monitoring work.

The experiment also presented some problems with this method, and these problems need to be solved in future research. Because of limitations in the locations and poses of the surveillance cameras, this method cannot produce a

complete real-time aerial view surveillance model of the service station. Instead, it can only accommodate the monitoring function for a part of the service station. Additionally, DG vehicles are typically rather large; it is therefore easy for them to occlude other DG vehicles, which causes failure in detection. However, because vehicles are not occluded for the entirety of the time when they are in the service station, in future work, we plan to add a target tracking algorithm to link a vehicle's information between different frames to predict the position of the occluded DG vehicles.

Data Availability

The data is not open-sourced as the surveillance videos are not allowed to distribute freely.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Z. S. Z. Dekun, "Planning and design of expressway service station," *New Architecture*, vol. 2, 2004.

[2] GB 6944-2012, Classification and code of dangerous goods.

[3] L. Pan, "Evaluation on the adaptability of dangerous goods vehicle in expressway service station," *Guangdong Highway Communications*, vol. 2014, no. 4, pp. 101–106, 2014.

[4] Z. Fu, *Application Research of Intelligent Monitoring System in the Entrances and Exits of Highway Service Area*, Changan University, Xi'an, 2015.

[5] H. Idrees, M. Shah, and R. Surette, "Enhancing camera surveillance using computer vision: a research note," *Policing: International Journal*, vol. 41, no. 2, pp. 292–307, 2018.

[6] G. T. S. Ho, Y. P. Tsang, C. H. Wu, W. H. Wong, and K. L. Choy, "A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities," *Sensors*, vol. 19, no. 8, p. 1796, 2019.

[7] W. Xiaokui, "Prospects for security monitoring of expressway toll station," *China ITS Journal*, vol. 2011, no. 7, pp. 100-101, 2011.

[8] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 175–187, 2006.

[9] C. Mi, Y. Huang, C. Fu, Z. Zhang, and O. Postolache, "Vision-based measurement: actualities and developing trends in automated container terminals," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 4, pp. 65–76, 2021.

[10] S. P. Biswas, P. Roy, N. Patra, A. Mukherjee, and N. Dey, "Intelligent traffic monitoring system," in *Proceedings of the Second International Conference on Computer and Communication Technologies*, pp. 535–545, Springer, New Delhi, 2016.

[11] M. A. Kafi, Y. Challal, D. Djenouri, M. Doudou, A. Bouabdallah, and N. Badache, "A study of wireless sensor networks for urban traffic monitoring: applications and architectures," *Procedia computer science*, vol. 19, pp. 617–626, 2013.

[12] N. K. Jain, R. K. Saini, and P. Mittal, "A review on traffic monitoring system techniques," *Advances in Intelligent Systems and Computing*, Springer, vol. 742, pp. 569–577, 2019.

[13] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, "A survey of vision-based traffic monitoring of road intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2681–2698, 2016.

[14] X. Feng, L. I. Wei, T. Wei, Y. Zhang, and L. Cao, *Calibration and Stitching Methods of Around View Monitor System of Articulated Multi-Carriage Road Vehicle for Intelligent Transportation*, SAE Technical Paper, 2019.

[15] C. C. Lin, S. U. Pankanti, K. Natesan Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1155–1163, Boston, MA, USA, June 2015.

[16] W. Burger and M. J. Burge, "Scale-invariant feature transform (SIFT)," in *Digital Image Processing: Texts in Computer Science*, pp. 609–664, Springer, London, UK, 2016.

[17] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[18] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the 2011 International Conference on Computer Vision*, pp. 2564–2571, IEEE, Barcelona, Spain, November 2011.

[19] S. Tuohy, D. O'Cualain, E. Jones, and M. Glavin, "Distance determination for an automobile environment using inverse perspective mapping in OpenCV," in *Proceedings of the IET Irish Signals and Systems Conference (ISSC 2010)*, Cork, Ireland, June 2010.

[20] M. Oliveira, V. Santos, and A. D. Sappa, "Multimodal inverse perspective mapping," *Information Fusion*, vol. 24, pp. 108–121, 2015.

[21] H. Li, M. Feng, and X. Wang, "Inverse perspective mapping based urban road markings detection," in *Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, pp. 1178–1182, IEEE, Hangzhou, China, October 2012.

[22] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, "High-resolution vehicle trajectory extraction and denoising from aerial videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3190–3202, 2020.

[23] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "A survey of unmanned aerial vehicles (UAVs) for traffic monitoring," in *Proceedings of the 2013 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 221–234, IEEE, Atlanta, GA, USA, May 2013.

[24] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 650–657, IEEE, Washington, DC, USA, May 2017.

[25] Q. Fan, L. Brown, and J. Smith, "A closer look at Faster R-CNN for vehicle detection," in *Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 124–129, IEEE, Gothenburg, Sweden, June 2016.

[26] N. Seenouvong, U. Watchareeruetai, C. Nuthong, K. Khongsomboon, and N. Ohnishi, "A computer vision based vehicle detection and counting system," in *Proceedings of the 2016*

8th International Conference on Knowledge and Smart Technology (KST), pp. 224–227, IEEE, Chiang Mai, Thailand, February 2016.

[27] G. Bradski, *Learning OpenCV*, O'REILLY, Massachusetts, USA, 2008.

[28] I. A. Karimi, M. Sharafali, and H. Mahalingam, "Scheduling tank container movements for chemical logistics," *AICHE Journal*, vol. 51, no. 1, pp. 178–197, 2005.

[29] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer, Cham, Switzerland, 2016.

[30] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, February 2015.