AI-Empowered Speed Extraction via Port-Like Videos for Vehicular Trajectory Analysis

Xinqiang Chen, *Member, IEEE*, Zichuang Wang, Qiaozhi Hua, Wen-Long Shang, *Member, IEEE*, Qiang Luo, and Keping Yu, *Member, IEEE*

***Abstract*— Automated container terminal (ACT) is considered as port industry development direction, and accurate kinematic data (speed, volume, etc.) is essential for enhancing ACT operation efficiency and safety. Port surveillance videos provide much useful spatial-temporal information with advantages of easy obtainable, large spatial coverage, etc. In that way, it is of great importance to analyze automated guided vehicle (AGV) trajectory movement from port surveillance videos. Motivated by the newly emerging computer vision and artificial intelligence (AI) techniques, we propose an ensemble framework for extracting vehicle speeds from port-like surveillance videos for the purpose of analyzing AGV moving trajectory. Firstly, the framework exploits vehicle position in each image via a feature-enhanced scale-aware descriptor. Secondly, we match vehicle position and trajectory data from the previous step output via Kalman filter and Hungarian algorithm, and thus we obtain the vehicular imaging trajectory in a frame-by-frame manner. Thirdly, we estimate the vehicular moving speed in real-world via the help of perspective projection theory. The experimental results suggest that our proposed framework can obtain accurate vehicle kinematic data under typical port traffic scenarios considering that the average measurement error of root mean square deviation is 0.675 km/h, the mean absolute deviation is 0.542 km/h, and the Pearson correlation coefficient is 0.9349. The research findings suggest that cutting-edge AI and computer vision techniques can accurately extract on-site vehicular trajectory related data from port videos, and thus help port traffic participants make more reasonable management decisions.**

***Index Terms*— AI-empowered techniques, vehicular trajectory analysis, speed data extraction, feature-enhanced scale-aware detector, automated container terminal.**

Manuscript received 20 November 2021; revised 28 February 2022 and 15 March 2022; accepted 6 April 2022. Date of publication 27 April 2022; date of current version 29 March 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1600602, in part by the National Natural Science Foundation of China under Grant 52102397, in part by the China Postdoctoral Science Foundation under Grant 2021M700790, in part by the Beijing Natural Science Foundation under Grant L211027, in part by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research under Grant JP18K18044 and Grant JP21K17736, and in part by the Hubei Natural Science Foundation under Grant 2021CFB156. The Associate Editor for this article was J. C.-W. Lin. *(Corresponding authors: Qiaozhi Hua; Wen-Long Shang.)*

Xinqiang Chen and Zichuang Wang are with the Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China (e-mail: xqchen@shmtu.edu.cn; wangzichuang@stu.shmtu.edu.cn).

Qiaozhi Hua is with the Computer School, Hubei University of Arts and Science, Xiangyang 441053, China (e-mail: 11722@hbuas.edu.cn).

Wen-Long Shang is with the Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100021, China, also with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China, and also with the Centre for Transport Studies, Imperial College London, London SW7 2BX, U.K. (e-mail: shangwl_imperial@bjut.edu.cn).

Qiang Luo is with the School of Civil Engineering, Guangzhou University, Guangzhou 510006, China (e-mail: luoq_yan@gzhu.edu.cn).

Keping Yu is with the Graduate School of Science and Engineering, Hosei University, Tokyo 184-8584, Japan (e-mail: keping.yu@ieee.org).

Digital Object Identifier 10.1109/TITS.2022.3167650

I. Introduction

AUTOMATED container terminal (ACT), which is considered as inevitable trend for future port upgrade, serves as a type of linking node of freight transmission between different regions and countries [1]–[4]. Automated guided vehicle (AGV) serves as one of the main elements of horizontal transport in the ACT, and accurate AGV navigation and positioning are quite important for improving ACT operation efficiency [5]–[7]. Currently, AGV transfers the containers from ship (yard) to yard (ship) navigated by magnetic tape, which requires densely deployment of the tapes in port roadways [8]. The ACT regulations intend to optimize efficiency for both equipment management system (ECS) and terminal operating system (TOS) in the manner of identifying optimal moving trajectory, better AGV loading/unloading strategy, less time cost, etc. [9], [10]. It is noted that ground smoothness is important for the tape installation, whilst the roadway renovation and tape maintenance is quite expensive [11], [12]. Additional attentions have been paid to further develop cost-effective yet high-accurate AGV navigational techniques to enhance port productivity efficiency [13], [14]. More specifically, many studies are conducted to obtain optimal AGV routing trajectories via the help of radar, Beidou, millimeter-wave radar, etc. Both ECS and TOS implement AGV trajectory optimization by controlling and adjusting the motion parameters (speed, direction, acceleration, etc.) [15], [16]. But, the performance of the above measuring facilities and techniques may be strongly degraded caused by steel structure electro-magnetic interference and corrosive environmental conditions in port [17]–[19].

The ACT surveillance video provides rich kinematic AGV spatial temporal information (e.g., AGV trajectory, on-site traffic state) free from the electro-magnetic interference [20]–[23]. In that manner, exploiting AGV moving parameters for the purpose of routing trajectory optimization attracts significant attentions in the ACT management community [24]–[26]. Lin *et al.*, proposed an efficient yet accurate AGV detection model via an ensemble Kalman filter and convolution neural

Fig. 1. Schematic diagram of the proposed framework.

network from port-based surveillance videos [27]. Yang *et al.*, introduced a hierarchical trajectory planning model to fulfill AGV fleet controlling task via the support of improved A∗ model and time window from bird eye view shot videos [28]. Ren *et al.*, implemented hybrid-intelligent real-time AGV trajectory control via a deep neural network [29]. Similar studies can be found in [30], [31]. Previous studies suggest that projection displacement difference phenomenon (i.e., the scale error caused by projection imaging) may corrupt AGV temporal-spatial data extraction accuracy [32], [33]. To address the issue, many researchers attempt to establish the mapping relationships from AGV imaging spatial-temporal information into real-world movement data based on the three-dimensional reconstruction or camera calibration algorithm. But, the acquisition of camera parameters and dimension information (e.g., focal length, rotation matrix, principal point) are one of the major sticking points. We may fail to obtain camera settings due to unpredictable reasons (i.e., focal length may be self-adjusted for obtaining optimal photographic performance) [34]–[37].

Previous studies related with vehicle speed estimation require additional parameter data (such as focal length, camera height, etc.), which cannot be easily applied to real-world applications. The study proposed a novel vehicle speed measurement scheme via the help of computer vision techniques, which are further enhanced by both of the Kalman filter and perspective projection models. The study aims to provide high-fidelity trajectory relevant data for fulfilling the task of AGV unmanned driving, automatic navigation and positioning, conflict-free path planning, etc. The main contributions for the study are summarized as follows: (1) a deep learning based port vehicle detector is developed in the manner of reconstructing the convolutional neural network architecture. More specifically, we replace the network backbone in Faster Region-based Convolutional Neural Network (R-CNN) detection algorithm with ResNet-101 residual network to enhance vehicle feature extraction. The image feature hierarchical structure is further introduced into Feature Pyramid Network (FPN) to achieve multi-scale detection; (2) we propose an efficient AGV vehicle speed exploitation framework for the purpose of trajectory analysis. Besides, we further implement cross-frame data association via Kalman filter and Hungarian algorithm (i.e., vehicle consecutive identity (ID) association); (3) we verify the proposed model performance with different port-like surveillance videos shot at typical challenges, which involves with complicated port environment, restricted visibility condition, and vehicle imaging occlusion.

II. Methodology

A. Overall Framework

The proposed framework for vehicular speed extraction includes three processing steps: vehicle detection, data association and vehicle speed estimation, as shown in Fig. 1. In the first step, we use a deep learning-based vehicle detector to extract target vehicles in the region of interest (RoI) from successive frames in the port-like surveillance video. After obtaining the vehicular position, the second step is to associate these vehicle detection data to trajectory using Kalman filter and Hungarian algorithm, and assign a unique tracking ID to each vehicle. The third step is to convert coordinates from image projection coordinate system into corrected geographic

coordinate system, and then estimate the vehicle speed based on perspective transformation. Simultaneously, a data quality control procedure runs throughout the above three steps to eliminate errors in vehicle detection, tracking and coordinate conversion.

B. Vehicle Detection With FE-SA Detector

There are various factors disturbing the vehicle detection in port-like surveillance videos, such as complicated background, visibility condition, and vehicle imaging occlusion. Moreover, the small-target detection (i.e., the vehicles in low-resolution) becomes one of the toughest challenges. In our study, a feature-enhanced scale-aware (FE-SA) vehicle detector based on improved Faster R-CNN is developed to implement the vehicle identification task in port-like surveillance videos. The main advantage of the proposed detector integrates both of residual learning and multi-scale feature fusion mechanism into deep convolutional neural network (DCNN) to achieve the robust and accurate vehicle detection performance.

The proposed vehicle detection procedure contains three stages. In the first stage, the vehicle relevant image features (e.g., color, texture, shape, etc.) in port-like surveillance videos are extracted through DCNN deep learning model. The second stage is to traverse the feature map obtained above and then receive candidate detection results by preliminary screening operation. The third stage implements the data quality control procedure and obtains the category and position information of target vehicles by bounding-box regression and classification.

Note that the detector is designed to fulfill the task of vehicle detection in specific scenarios of the port-like surveillance video. The DCNN model is supposed to learn rich features and semantic information from port images. The advantages of the proposed detector are illustrated as follows:

1) Feature Enhancement: Firstly, deep residual learning is introduced into the vehicle detector. More specifically, we replace the backbone network with ResNet (Residual Network). Meanwhile, a highly modularized split-transform-merge structure is also built in the residual unit. Specifically, it takes the feature map extracted from the port image as input, which are split into several low-dimensional embedding by 1×1 kernel, transformed by 3×3 kernel, and merged by concatenation. After the above two steps, the feature representational capacity of DCNN can be significantly improved (i.e., through acquiring deeper layers and wider scale), thereby enhancing the extracted image features. More details can be found in [38].

2) Scale Awareness: Another visual task in the port-like surveillance video is the challenge caused by scale variations. Vehicles appear at different scales, depending on their distance to the camera. Thus, the network is endowed with the multi-scale perception ability to improve the detection performance for the small target. Specifically, the hierarchical structure of feature maps in the DCNN is leveraged to build a feature pyramid structure. It realizes that the extracted feature maps at different scales are all with strong semantic information, thereby optimizing the multi-scale detection problem.

Vehicle detection output is closely related to the subsequent steps. But there are still errors caused by irregular fluctuations, such as low-quality source data and external noise. The data quality control procedure is performed to correct possible errors in the program. We use an adaptive weighted filter to implement data quality control, and its core idea lies in optimal fitting based on the polynomial least square method. Specifically, there is relatively small difference of the vehicle position in continuous frames (i.e., in extremely short intervals); the adaptive filter is used to suppress the signal fluctuation to achieve more effective and smoother data denoising.

Fitting the vehicular position in successive frames by higher order polynomial, which are shown as Eq. (1):

$$\Psi_{2l+1} = T_{2l+1} \times A_l + \varepsilon_{2l+1} \tag{1}$$

where Ψ_{2l+1} denotes the vehicle position measurement of $2l+1$ frame around the current frame. A_l denotes the weighted parameter for the adaptive filter. ε_{2l+1} denotes the cumulative error of polynomial fitting.

The weighted parameter of the adaptive filter is solved by performing the least square fitting for the given high order polynomial, which are shown as Eq. (2):

$$\phi_{2l+1} = T_{2l+1} \times A_l^* \tag{2}$$

where ϕ_{2l+1} denotes the filtered value of vehicle position of $2l+1$ frames around the current frame. A_l^* denotes the optimal weighted parameter obtained by least square fitting.

The data quality control procedure is implemented through calculating the relationship matrix between measurement data and filter data of vehicle position, as shown in Eq. (3):

$$\Delta = \phi_{2l+1} \times \Psi_{2l+1}^T \tag{3}$$

where Δ denotes the relationship matrix between measurement values and filtered values.

Extracted traffic parameters of vehicles, including position, tracking ID, speed, etc., may arise unplanned errors by inevitable interference. Thus, the adaptive filter is also used to eliminate noise and outliers in subsequent steps. The data quality control procedure has been described here, and it will not be repeated in subsequent sections.

C. Cross-Frame Data Association

After vehicle detection, the vehicular position (i.e., detection bounding-box) in consecutive frames can be accurately and completely obtained. Note that the detection results in each frame are independent, so a tracker is requisite for data association. Kalman filter is used to estimate the motion state of the vehicle bounding-box. More specifically, it correlates cross-frame detection results from the same vehicle to form the trajectory relevant data. We model the vehicle status as a seven-dimensional vector η as Eq. (4), which describes the vehicular position and the variation information in each dimension [39], [40].

$$\eta = [x, y, r, \tau, \dot{x}, \dot{y}, \dot{r}]^T \tag{4}$$

where x and y are the center coordinates of the vehicle bounding-box, r and τ are the area and aspect ratio, while $(\dot{x}, \dot{y}, \dot{r})$ denote the corresponding variation rate. Note that the aspect ratio is assumed to be constant.

The process of vehicle motion estimation includes two steps: prediction and update. The vehicle state in the current frame is predicted based on the vehicle detector results in the previous frame, which are shown as Eq. (5) and Eq. (6):

$$\hat{\eta}_t^- = \Phi \hat{\eta}_{t-1} + \Gamma \omega \tag{5}$$

$$P_t^- = \Phi P_{t-1} \Phi^T + \Gamma Q \Gamma^T \tag{6}$$

where $\hat{\eta}_t^-$ and P_t^- are the vehicle state vector and its covariance matrix in the current frame (at time t), Φ is the state transition matrix, $\hat{\eta}_{t-1}$ and P_{t-1} are the state vector and covariance matrix in the previous frame (at time t-1), ω and Q are the process noise and its covariance matrix in the frame sequences, Γ is the noise-driven matrix.

The gain matrix for Kalman filter is shown in Eq. (7), which is updated for predicting optimal vehicle state via the help of Eq. (8) and Eq. (9).

$$G_t = P_t^- \Theta^T \left(\Theta P_t^- \Theta^T + R \right)^{-1} \tag{7}$$

$$\hat{\eta}_t = \hat{\eta}_t^- + G_t \left(\gamma_t - \Theta \hat{\eta}_t^- \right) \tag{8}$$

$$P_t = P_t^- - G_t \Theta P_t^- \tag{9}$$

where G_t is the Kalman filter gain, γ_t is the vehicle detection result as observation vector, Θ is the observation matrix, R is the covariance matrix of observation noise.

In the second step, combined with the historical vehicle detection results, Hungarian algorithm is used to optimize the assignment problem for the predicted values given by Kalman filter. Each vehicle will be assigned a unique tracking ID. It realizes cross-frame data association and generates the trajectory relevant data for vehicles.

Considering the contextual information of the vehicle position in an extreme short interval, that is, the detection bounding-box has a high-overlap area between consecutive frames, we introduce the Intersection over Union (IoU) distance as cost matrix in Hungarian algorithm to achieve the optimal allocation for vehicle ID. The IoU distance is defined as Eq. (10):

$$IoU\left(\alpha, \beta\right) = \frac{Det\left(\alpha\right) \cap Pre\left(\beta\right)}{Det\left(\alpha\right) \cup Pre\left(\beta\right)} \tag{10}$$

where $IoU\left(\alpha, \beta\right)$ denotes the intersection and union ratio of two rectangular areas, $Det\left(\alpha\right)$ denotes the bounding-box generated by the vehicle detector in the previous frame, $Pre\left(\beta\right)$ is predicted by the Kalman filter in the current frame.

D. Speed Estimation Based on Perspective Projection

Through vehicle detection and cross-frame data association, we can obtain the vehicular motion state in the port-like surveillance video. Select the center coordinates of the detected bounding-box for quantitative analysis to simplify the vehicle model. Based on the perspective projection, the mapping relationship of vehicle position (i.e., bounding-box coordinates) between the projection imaging coordinate system and the corrected coordinate system is established as Eq. (11). Note that the raw vehicle position in video are captured and recorded by an uncalibrated camera, while the corrected coordinate system obtained after perspective transformation can reflect the actual length in the port scene.

$$Hz = \begin{bmatrix} h_a & h_b & h_c \\ h_d & h_e & h_f \\ h_g & h_h & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \grave{Z} \tag{11}$$

where $z = [x, y, 1]^T$ denotes the raw vehicle position in projection imaging coordinate system, and $\grave{Z} = [u, v, 1]^T$ denotes the corrected position in geographic coordinate system. H is the homography matrix for perspective transformation. Note that the coordinates in space dimension are all set to one considering the vehicle motion is only in a planar dimension.

To solve the homography matrix for perspective projection, we select four pixel-points in the raw image, and then input the corresponding geographical coordinates of the four pixel-points as prior knowledge. Plug the above four pairs of homologous pixels into Eq. (11) to calculate the factors for coordinate conversion, as shown in Eq. (12):

$$\begin{bmatrix} h_a \\ h_b \\ h_c \\ h_d \\ h_e \\ h_f \\ h_g \\ h_h \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1 x_1 & -u_1 y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1 x_1 & -v_1 y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -u_2 x_2 & -u_2 y_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -v_2 x_2 & -v_2 y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -u_3 x_3 & -u_3 y_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -v_3 x_3 & -v_3 y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -u_4 x_4 & -u_4 y_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -v_4 x_4 & -v_4 y_4 \end{bmatrix} \times \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \end{bmatrix} \tag{12}$$

where $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ denotes four pixel coordinates in raw image, $\{(u_1, v_1), (u, v_2), (u_3, v_3), (u_4, v_4)\}$ denotes the corresponding geographical coordinates in the port scene, $\left[h_a, h_b, h_c, h_d, h_e, h_f, h_g, h_h\right]^T$ denotes the coordinate conversion factors in homography matrix. Note that performing the Min-Max Normalization is required to map the result value between [0 - 1] when calculating coordinates.

Through the above steps, the vehicular position in image domain is converted into a corrected coordinate system with the geographical space scale. Thus, the vehicle movement distance can be expressed as the change of bounding-box coordinates in consecutive frames. According to the Euclidean distance, the vehicle moving displacement between two frames can be expressed as Eq. (13):

$$\mathcal{D}_t = \sqrt{(u_t - u_{t-1})^2 + (v_t - v_{t-1})^2} \tag{13}$$

where u_t and v_t denote the horizontal and vertical coordinates of vehicle position in the current frame. u_{t-1} and v_{t-1} denote the coordinates in the previous frame.

Then the vehicle speed can be estimated as Eq. (14):

$$\mathcal{V}_t = \mu \mathcal{D}_t \cdot {}^{N_{FPS}}\!/_{N_t} \tag{14}$$

where N_{FPS} denotes the frame per second (FPS) in videos, N_t is the current frame number when the target vehicle is detected, μ denotes the conversion factor for velocity. In this study, μ is set to 3.6 to achieve a unit conversion from m/s to km/h.

Further, we perform the following pre-processing stage on the input images to achieve a more robust response to the vehicle speed estimation in port-like surveillance videos:

(1) as the imaging region is far away from the camera, the distortion becomes prominent. According to the distance from the camera, the selected lane is divided into multiple sub-regions to optimize the distortion error caused by perspective transformation. Specifically, we divide the selected lane into several rectangular areas and solve the corresponding homography matrix respectively.

(2) the speed extraction program is only implemented to the vehicles locating at the region of interest. We cover every frame of the original video with a mask. More specifically, we use the Iverson bracket indicator function (see Eq. (15)) to suppress detection interference from vehicles in other lanes or the region of no interest.

$$\mathrm{I}_{targ} = \lambda\left[\mathrm{I} \in \Omega\right]\mathrm{I}_{orig} \tag{15}$$

where I_{targ} denotes the target pixel value of the selected lane region, and I_{orig} denotes the pixels in the original image. The indicator function $\lambda\left[\mathrm{I} \in \Omega\right]$ is set to 1 when pixel I belongs to the selected lane region Ω and 0 otherwise.

III. Experiment Design

A. Data Description

We collected the typical yet representative port scenarios in port-like surveillance videos from different time periods. The detailed information of the selected videos is shown in table I. In video #1, it includes 454 frames collected in typical port scenes during the daytime. The videos are all shot at 1280×720 resolution and 25 frames per second (fps). In video #2, it includes 377 frames shot at night, in which some image details are destroyed caused by the restricted visibility. Video #3 also includes 599 frames in the daytime, but it contains the phenomenon of vehicle imaging occlusion caused by the lane change. In video #1 and video #3, since the image profile is smaller when the vehicle is far from the camera, there is a challenge of vehicle small-target detection for a period of time at the beginning of videos. And due to the camera installation angle (i.e., non-bird's-eye view) and the low resolution of captured images, all the collected videos have a certain degree of distortion. In addition, the categories of detected vehicles contain bicycles and motorcycles.

TABLE I
INFORMATION OF PORT SURVEILLANCE VIDEOS

Information	video #1	video #2	video #3
time period	daytime	night	daytime
duration	18 s	15 s	24 s
frame rate	25 fps	25 fps	25 fps
resolution	1280×720	1280×720	1280×720
vehicle occlusion	/*	/	√
small target detection	√	/	√
image distortion	√	√	√

*(/ denotes the situation exist in the video clip, while √ denote the situation is involved in the video)

B. Parameter Sensitivity Analysis

In this section, we further conducted the parameter tuning and sensitivity analysis on the proposed framework, in which four leading parameters were investigated, including the confidence threshold of vehicle detection ρ, trigger condition of vehicle detection ξ, maximum standby time of vehicle tracking ID φ, IoU distance threshold θ. The optimal parameters set in different port scenarios are summarized in table II, where the representative meanings of threshold ξ and φ are as follows:

(1) a unique tracking ID needs to be created accordingly when a new vehicle emerge into the ROI. For avoiding false alarm of the detector, detection results of the new vehicle must last certain frames (threshold ξ). It means that the new tracking ID needs to go through a trial period.

(2) it's also necessary to destroy its ID and terminate the tracking trajectory when a vehicle disappear from the ROI. We set the threshold φ to control the ID survival time after losing the detection match, so as to prevent ID-switch caused by missing detection.

First of all, the confidence threshold ρ was set as 0.6 in the daytime, while it was lowered to 0.5 in order to detect as many vehicles as possible in poor visibility. Likewise, in video #2, we set a lower value of ξ to make it more sensitive to candidate vehicle detection. For video #3 with vehicle occlusion, we appropriately raised the threshold φ to ensure that the tracking ID of the sheltered vehicle will not be discarded, so that it can be re-matched with the original ID after the detection was restored. Lastly, considering the variations of bounding-box area in different port scenarios, we set the IoU distance threshold values θ as 0.65, 0.5 and 0.6 respectively.

C. Performance Evaluation Indicators

To verify the performance of the framework proposed in this study, the ground truth data of vehicle speed were manually calibrated by graduate students when collecting videos. Specifically, we recorded the path length of the target vehicle at every sampling interval (a second), and then calculated it to ensure the accuracy of the ground-truth speed for each vehicle.

Three indicators were taken account to evaluate the fitness of the vehicle speed estimation framework, which are the Root Mean Squared Deviation (RMSE), the Mean Absolute Deviation (MAE), and the Pearson correlation coefficient (Pearson's r). For each vehicle in the port surveillance video, the evaluation indicators were calculated according to Eq. (16) to (20), as shown at the bottom of the next page. A smaller value of RMSE, MAE or larger Pearson's r indicates that the extracted vehicle speed is closer to the ground-truth

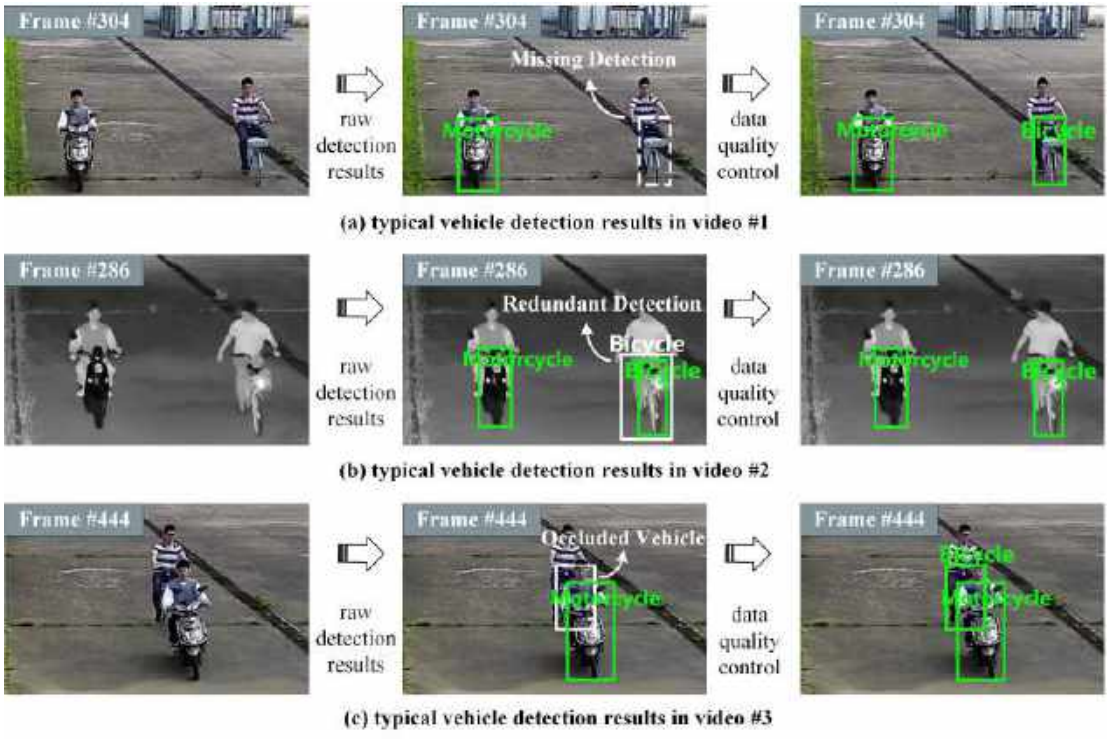


Fig. 2. Results of vehicle detection by the FE-SA detector.

value, and vice versa, where t is the time of the vehicle movement. $V_{smth}(t)$ and $V_{Grud}(t)$ are the smoothed value and the ground-truth value of vehicle speed data. $\overline{V_{smth}}$ and $\overline{V_{Grud}}$ are the mean value of $V_{smth}(t)$ and $V_{Grud}(t)$. n is the number of sampling point in the video.

IV. Experiment Results

A. Framework Procedure Outputs

The outputs of each step in the methodological framework were described to show the overall working process. Due to the limitations from the different port scenarios in surveillance videos, the detection outliers were usually manifested as missing detections (i.e., detector didn't generate a correct bounding box.) and redundant detections (i.e., multiple boxes correspond to the same vehicle). The vehicle detection results are shown in Fig. 2. It can be seen that the developed FE-SA detector can brilliantly tackle the detection defects of vehicle identification and classification.

The vehicle position in frame sequences records a relatively trivial variation. Thus, in the data quality control procedure, missing detections can be filled by interpolating the position information based on polynomial fitting. As for redundant detections, we retained the target bounding-box with the largest IoU distance and eliminated redundant options.

Vehicle imaging occlusion phenomenon caused by the lane change behavior is shown in figure 2(c). The rear vehicle cannot be observed in images when the camera shooting angle

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|V_{Grud}(t) - V_{smth}(t)|} \tag{16}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|V_{Grud}(t) - V_{smth}(t)| \tag{17}$$

$$Pearson's\ r = \frac{\sum_{t=1}^{n}\left[\left(V_{smth}(t) - \overline{V_{smth}}\right) \times \left(V_{Grud}(t) - \overline{V_{Grud}}\right)\right]}{\sqrt{\sum_{t=1}^{n}\left(V_{smth}(t) - \overline{V_{smth}}\right)^2} \times \sqrt{\sum_{t=1}^{n}\left(V_{Grud}(t) - \overline{V_{Grud}}\right)^2}} \tag{18}$$

$$\overline{V_{smth}} = \frac{1}{n}\sum_{t=1}^{n}V_{smth}(t) \tag{19}$$

$$\overline{V_{Grud}} = \frac{1}{n}\sum_{t=1}^{n}V_{Grud}(t) \tag{20}$$

TABLE II
PARAMETERS SETTING IN DIFFERENT PORT SURVEILLANCE VIDEOS

Parameter	Meaning	video #1	video #2	video #3
ρ	confidence threshold of vehicle detection	0.60	0.50	0.60
ξ	trigger condition of vehicle detection	10	3	10
φ	maximum standby time of vehicle tracking ID	20	20	80
θ	IoU distance threshold	0.65	0.50	0.60

TABLE III
RESULTS OF VEHICLE DETECTION WITH DIFFERENT MODELS

	our method			Faster R-CNN			Grid R-CNN			YOLO V5		
	accuracy	missing rate	redundant rate	accuracy	missing rate	redundant rate	accuracy	missing rate	redundant rate	accuracy	missing rate	redundant rate
video #1	99.99%	0.01%	0.00%	81.01%	12.74%	6.25%	90.40%	7.23%	2.37%	74.59%	16.06%	9.35%
video #2	99.99%	0.01%	0.00%	73.21%	14.29%	12.50%	91.18%	7.95%	0.87%	80.85%	5.01%	14.14%
video #3	99.99%	0.01%	0.00%	67.99%	24.80%	7.21%	89.88%	8.99%	1.13%	61.99%	31.04%	6.97%
Average	99.99%	0.01%	0.00%	74.07%	17.28%	8.65%	90.49%	8.06%	1.46%	72.48%	17.37%	10.15%

and vehicles were in same plane. Vehicle imaging occlusion may mislead the detector to fail to identify objects as vehicles. We employed the vehicular spatio-temporal data before and after occlusion scenario to fit the vehicle movement for occlusion frames. After the data quality control, the vehicles of interest in each frame could be properly recognized by vehicle detector.

The vehicle bounding boxes in successive frames is one of the most important information in the framework, which greatly affects the quality of subsequent steps (i.e., vehicle ID association and speed estimation). We trained the detection network on the MS COCO dataset [41], and compared it with the mainstream detection algorithms to preliminarily validate the performance of vehicle detector.

We also introduce three statistical indicators to evaluate the performance of different vehicle detectors in the study. More specifically, the accuracy (see Eq. (21)), missing rate (see Eq. (22)) and redundant rate (see Eq. (23)) are employed to quantify performance of various detectors. The formula used for calculating the three indicators are shown as follows:

$$\text{accuracy} = {N_{true}}/{N_{total}} \tag{21}$$

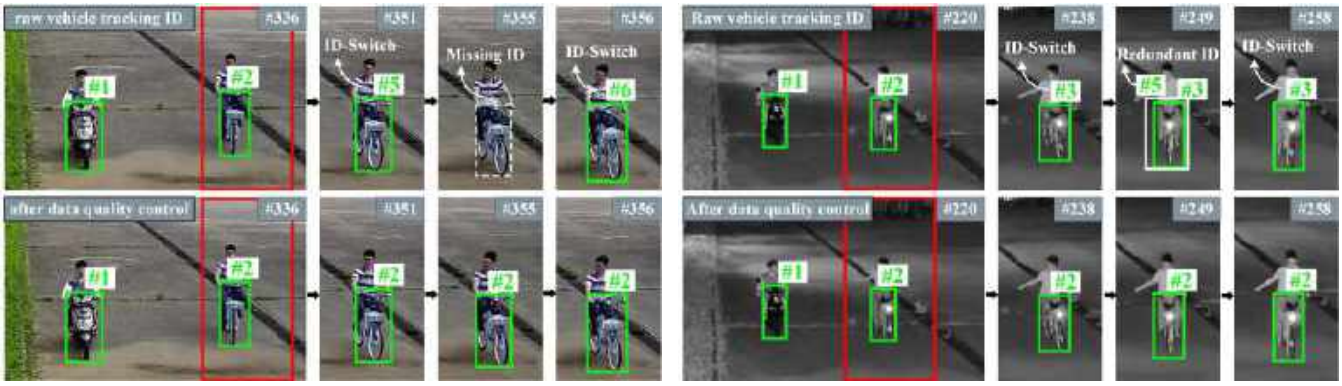
$$\text{missing rate} = {N_{miss}}/{N_{total}} \tag{22}$$

$$\text{redundant rate} = {N_{redun}}/{N_{total}} \tag{23}$$

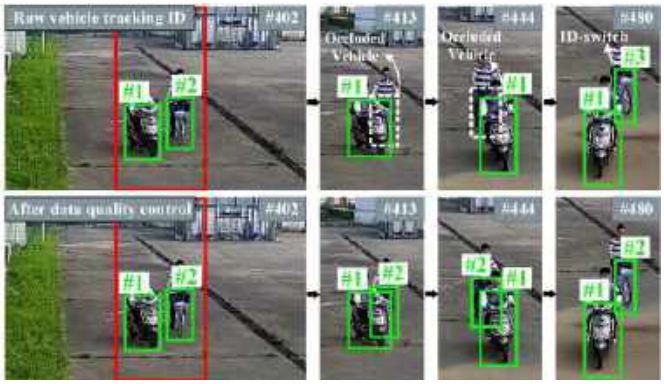
where N_{total} denotes the total number of vehicles in all frames in a port-like surveillance video. The symbol N_{true} is the number of vehicles which are correctly identified by the detector. Besides, the N_{miss} denotes the miss-detected vehicle number, whilst N_{redun} demonstrates the number of vehicles which are detected into different vehicles.

As shown in table III, experimental results show that the FE-SA vehicle detector can achieve superior performance in various port scenarios, basically realizing the vehicles detection in each frame in port surveillance video. Compared with Faster R-CNN, Grid R-CNN, and YOLOv5 detection algorithm [42], [43], the proposed method has a significant accuracy improvement, and also reduces the occurrence of missing detections and redundant detections. Note that there were extremely slight missing detections in FE-SA detector, it occurred at the last moment when the vehicle drove out of the surveillance screen. At this time, due to the obvious fluctuation of bounding boxes caused by the incomplete imaging region of vehicles, this part of the outliers will be discarded in subsequent steps.

In the data association, each target vehicle was assigned a unique tracking ID. As shown in Fig. 3, there are two main reasons for the ID-Switch (i.e., the ID of the same vehicle has changed.) The occurrence of vehicle detection outliers (missing detections and redundant detections) was a primary cause of ID-Switch. For example, when there were several frames of missing detection, the re-detected vehicle would be viewed as a new target and reassign to another tracking ID; when redundant detections occurred, it may result in the same vehicle being assigned multiple tracking IDs. These interferences above could be well resolved after quality control for vehicle detection results. For another reason, in case of vehicle imaging occlusion caused by the congestion traffic flow or the lane switch, the tracking ID of the nearby vehicles may be exchanged and confused since detection bounding boxes are highly overlapped. Consequently, we selected the bounding box with the optimal IoU distance for matching detection to suppress the occurrence of ID-Switch.



(a) typical vehicle data association results in video #1 (b) typical vehicle data association results in video #2



(c) typical vehicle data association results in video #3

Fig. 3. Results of vehicle tracking ID with the data association model.

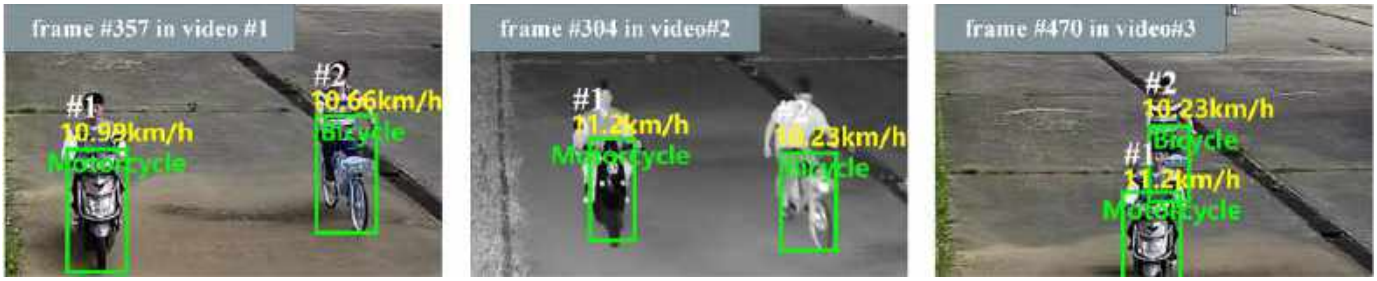


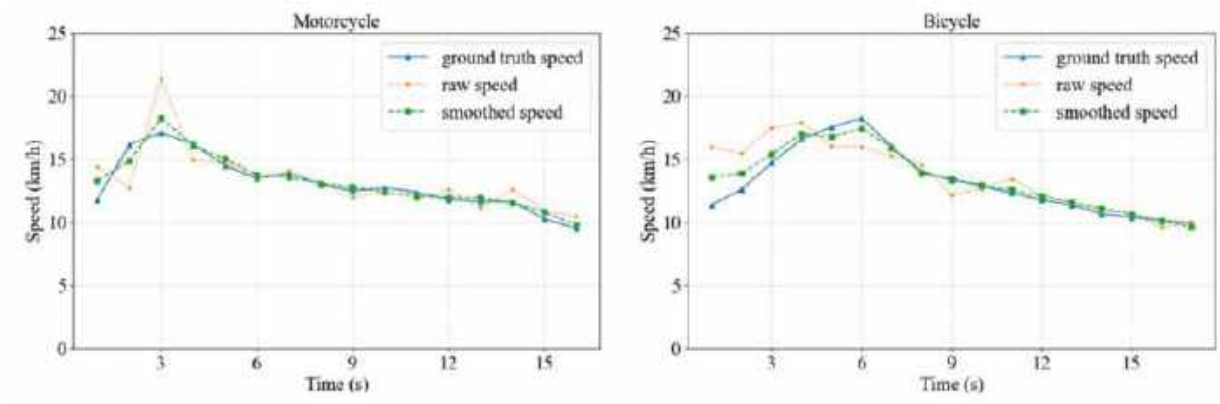
Fig. 4. Typical frames of results of the vehicle speed estimation.

According to the vehicle detected boxes and the trajectory relevant data from association, we can further extract the vehicle speeds from port-like surveillance videos. The result diagram of vehicle speed estimation is shown in Fig. 4. Since vehicle speed is determined via the displacement of the bounding-box center, it might easily lead to an accumulation of errors caused by the abnormal fluctuation in detection results. In data quality control process, the adaptive filter is used to remove the outliers and smooth the noise in the raw vehicular speed data.

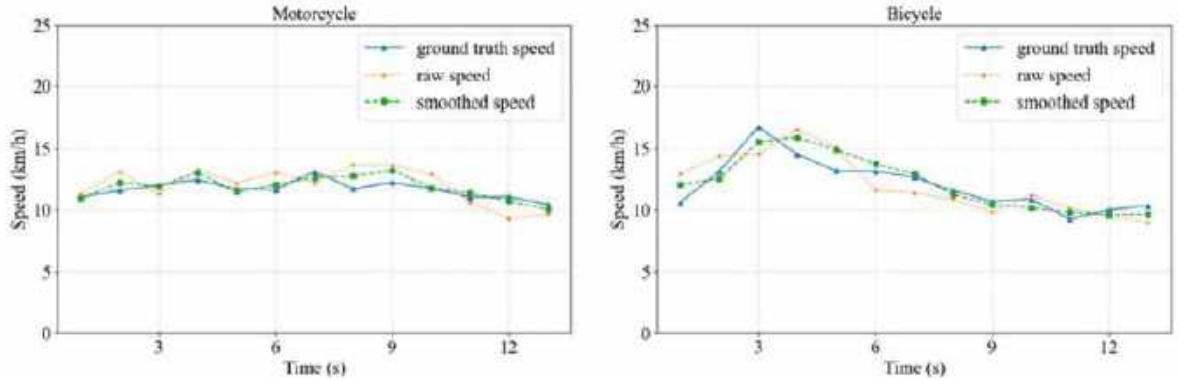
B. Results of Vehicle Speed Extraction

We performed the complete program on three captured videos, and estimated the speeds of vehicles in port-like surveillance videos with a sampling interval of one second. Note that the data obtained by the framework were the instantaneous speed of the vehicle in each frame, thus the speeds of every 25 frames (i.e., one second in the video) would be taken as a group of experimental data and calculated their average values. The raw data and denoising data of the vehicle speed were compared with the ground-truth value, and the results are summarized in Fig. 5. Experimental results show that the proposed method can effectively extract vehicle speed data from port surveillance videos, and the speed is more closer to the ground-truth value after data quality control.

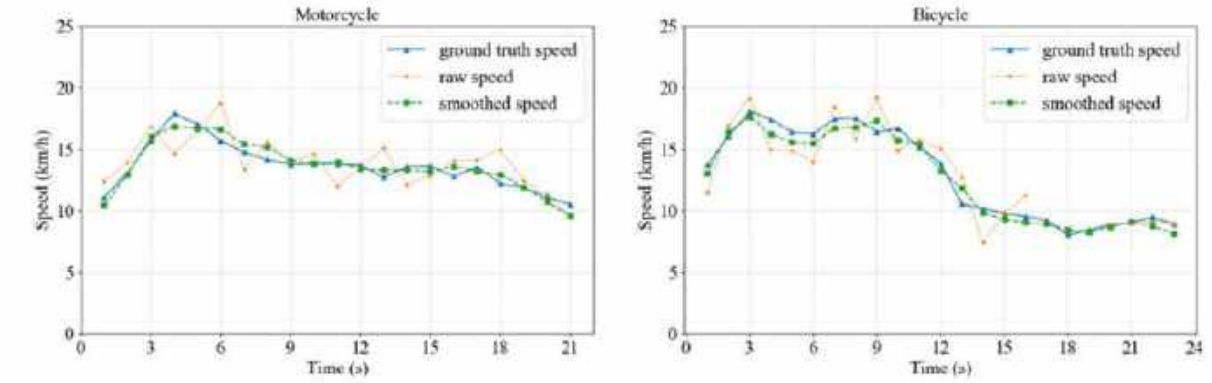
At the beginning of video #1 and video #3, there was a relatively distinct error in the raw data of vehicle speed. The main reason may be that there was a period of small-target detection challenges as the vehicles were far away from the surveillance camera. Therefore, the detected bounding-box



(a) speeds of all detected vehicles in video #1



(b) speeds of all detected vehicles in video #2



(c) speeds of all detected vehicles in video #3

Fig. 5. Results of vehicle speed extracted with the proposed framework.

may not match the actual imaging region of the target vehicle, which could lead to the errors from vehicle position deviation. Meanwhile in video #3, the cause for the significant fluctuation in raw speed data is that the area of bounding box changed in a wider range when lane switch. More specifically, when vehicle switched the lane, the vehicle imaging region in the video changed from the front end of the vehicle (i.e., a long and narrow detection box) to the body of the vehicle (i.e., a short and wide detection box). It can be clearly seen that both two kinds of experimental errors were well suppressed after quality control.

Without the small-target detection challenge, the raw vehicular speed data has a more stable performance in video #2, which indicates that the proposed method can accurately reflect the variation tendency of the vehicle speed even in poor visibility. We further quantified the performances of experimental data by calculating the evaluation indicators for each vehicle in videos, and summarized the results in table IV. It is found that in video #2 the accuracy was slightly inferior than that in the other two port scenarios, which was mainly because the image features and details were partially destroyed by the restricted visibility. In summary, the experimental results show that the proposed method can accurately extract the vehicle speed in various port scenarios. (A smaller RMSE and MAE or a larger Pearson's r indicate that the results are closer to the true value.) And after data quality control, the precision of the smooth data is significantly improved compared with the raw data of vehicle speed. The average measurement error of

TABLE IV
STATISTICAL PERFORMANCE OF VEHICLE SPEED ESTIMATION AT DIFFERENT PORT SCENARIOS

	RMSE *(km/h)*		MAE *(km/h)*		Pearson's r	
	Raw vs. Ground Truth	Smooth vs. Ground Truth	Raw vs. Ground Truth	Smooth vs. Ground Truth	Raw vs. Ground Truth	Smooth vs. Ground Truth
video #1 port in daytime						
Motorcycle	1.628	0.632	1.103	0.467	0.7587	0.9549
Bicycle	1.714	0.723	1.219	0.485	0.7909	0.9641
Average	1.671	**0.678**	1.161	**0.476**	0.7748	**0.9595**
video #2 port at night						
Motorcycle	1.165	0.531	1.047	0.445	0.6320	0.8389
Bicycle	1.455	0.913	1.312	0.782	0.7790	0.9104
Average	1.310	**0.722**	1.180	**0.614**	0.7055	**0.8747**
video #3 vehicle occlusion						
Motorcycle	1.556	0.584	1.284	0.490	0.6824	0.9542
Bicycle	1.616	0.669	1.327	0.583	0.8980	0.9869
Average	1.586	**0.627**	1.306	**0.537**	0.7902	**0.9706**

RMSE and MAE are both lower than 1km/h, and the average Pearson correlation coefficient is 0.9349.

V. CONCLUSION

AGV trajectory related temporal-spatial information is quite important for ACT management efficiency and productivity. With the help of cutting-edge AI models and computer vision techniques, we proposed a novel framework to extract vehicle speed from port-like surveillance videos captured by monocular camera. The model can obtain satisfied performance without the support of scene reconstruction and camera calibration. The framework introduced an improved RCNN deep learning model to accurately detect vehicle from videos, which were further mapped into real-world kinematic information with the help of perspective projection transformation model.

The proposed framework started by automatically extracting vehicle imaging locations with the FE-SA vehicle detector. Then, we associated vehicle trajectory data in consecutive images via a context-aware tracker. After that, we restored each frame into bird-eye view images and further estimated vehicular speeds via perspective projection transformation operation. Note that data quality control procedure was integrated in each step to correct out anomaly vehicle trajectory relevant data samples. In this way, the proposed framework obtained both static and kinematic vehicle motion information (e.g., position, vehicle category, speed) for trajectory based analysis. We verified the proposed framework performance on three typical port-like surveillance videos. Our proposed framework can extract accurate vehicular speed data under varied port-like traffic scenarios considering that the statistical indicators of MAE and RMSE were quite small while the Pearson's r was large.

We can expand our work in the following potential directions. First, the port-like videos were shot with cameras installed at fixed positions. We can further verify the proposed framework performance on port videos taken by AGV-borne cameras. Second, current videos only involved with low traffic volume situations. Thus, model performance verification under large traffic volume situation is considered as another interesting potential direction. Last but not least, we can further improve model performance by enhancing the vehicle tracking accuracy against challenges of vehicle occlusion, camera vibration, adverse weather condition, etc.

REFERENCES

[1] Z. H. Munim, O. Duru, and A. K. Y. Ng, "Transhipment port's competitiveness forecasting using analytic network process modelling," *Transp. Policy*, pp. 1–13, Jul. 2021, doi: 10.1016/j.tranpol.2021.07.015.
[2] B. Wu, Y. Tang, X. Yan, and C. G. Soares, "Bayesian network modelling for safety management of electric vehicles transported in RoPax ships," *Rel. Eng. Syst. Saf.*, vol. 209, May 2021, Art. no. 107466.
[3] X. Chen, J. Ling, S. Wang, Y. Yang, L. Luo, and Y. Yan, "Ship detection from coastal surveillance videos via an ensemble Canny-Gaussian-morphology framework," *J. Navigat.*, vol. 74, no. 6, pp. 1252–1266, Nov. 2021.
[4] H. Bi, W.-L. Shang, Y. Chen, K. Wang, Q. Yu, and Y. Sui, "GIS aided sustainable urban road management with a unifying queueing and neural network model," *Appl. Energy*, vol. 291, Jun. 2021, Art. no. 116818.
[5] M. Zhong, Y. Yang, Y. Dessouky, and O. Postolache, "Multi-AGV scheduling for conflict-free path planning in automated container terminals," *Comput. Ind. Eng.*, vol. 142, Apr. 2020, Art. no. 106371.
[6] X. Wu, C. Sun, T. Zou, L. Li, L. Wang, and H. Liu, "SVM-based image partitioning for vision recognition of AGV guide paths under complex illumination conditions," *Robot. Comput.-Integr. Manuf.*, vol. 61, Feb. 2020, Art. no. 101856.
[7] N. Ma, C. Zhou, and A. Stephen, "Simulation model and performance evaluation of battery-powered AGV systems in automated container terminals," *Simul. Model. Pract. Theory*, vol. 106, Jan. 2021, Art. no. 102146.

[8] N. Singh, Q.-V. Dang, A. Akcay, I. Adan, and T. Martagan, "A matheuristic for AGV scheduling with battery constraints," *Eur. J. Oper. Res.*, vol. 298, no. 3, pp. 855–873, May 2022.

[9] J. C.-W. Lin, S. Ren, P. Fournier-Viger, and T.-P. Hong, "EHAUPM: Efficient high average-utility pattern mining with tighter upper bounds," *IEEE Access*, vol. 5, pp. 12927–12940, 2017.

[10] Z. Yang, W.-L. Shang, H. Zhang, H. Garg, and C. Han, "Assessing the green distribution transformer manufacturing process using a cloud-based q-rung orthopair fuzzy multi-criteria framework," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118687.

[11] G. Demesure, M. Defoort, A. Bekrar, D. Trentesaux, and M. Djemaï, "Decentralized motion planning and scheduling of AGVs in an FMS," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1744–1752, Apr. 2018.

[12] G. Zeng *et al.*, "Percolation-based health management of complex traffic systems," *Frontiers Eng. Manage.*, vol. 8, no. 4, pp. 557–571, Dec. 2021.

[13] F. Ding, K. Yu, Z. Gu, X. Li, and Y. Shi, "Perceptual enhancement for autonomous vehicles: Restoring visually degraded images for context prediction via adversarial training," *IEEE Trans. Intell. Transp. Syst.*, early access, Oct. 26, 2021, doi: 10.1109/TITS.2021.3120075.

[14] N. Gray, S. McDonagh, R. O'Shea, B. Smyth, and J. D. Murphy, "Decarbonising ships, planes and trucks: An analysis of suitable low-carbon fuels for the maritime, aviation and haulage sectors," *Adv. Appl. Energy*, vol. 1, Feb. 2021, Art. no. 100008.

[15] Z. Rozsa and T. Sziranyi, "Obstacle prediction for automated guided vehicles based on point clouds measured by a tilted LIDAR sensor," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2708–2720, Aug. 2018.

[16] S. Sindi and R. Woodman, "Implementing commercial autonomous road haulage in freight operations: An industry perspective," *Transp. Res. A, Policy Pract.*, vol. 152, pp. 235–253, Oct. 2021.

[17] J. C.-W. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, and M. Aloqaily, "Privacy-preserving multiobjective sanitization model in 6G IoT environments," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5340–5349, Apr. 2021.

[18] W.-L. Shang *et al.*, "Benchmark analysis for robustness of multi-scale urban road networks under global disruptions," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 16, 2022, doi: 10.1109/TITS.2022.3149969.

[19] W.-L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: A big-data analysis," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116429.

[20] H. Tang, Z. Dang, Y. Sun, Y. Liu, and S. Pu, "Simultaneous intrinsic and extrinsic calibration of a visual-odometric sensor system," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6856–6868, Mar. 2021.

[21] L. Yue, H. Fan, and M. Ma, "Optimizing configuration and scheduling of double 40 ft dual-trolley quay cranes and AGVs for improving container terminal services," *J. Cleaner Prod.*, vol. 292, Apr. 2021, Art. no. 126019.

[22] K. Yu, L. Lin, M. Alazab, L. Tan, and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4337–4347, Jul. 2021.

[23] W.-L. Shang, Y. Chen, and W. Y. Ochieng, "Resilience analysis of transport networks by combining variable message signs with agent-based day-to-day dynamic learning," *IEEE Access*, vol. 8, pp. 104458–104468, 2020.

[24] B. Liu, Y. Zhang, G. Zhang, and P. Zheng, "Edge-cloud orchestration driven industrial smart product-service systems solution design based on CPS and IIoT," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100984.

[25] F. Ding, G. Zhu, Y. Li, X. Zhang, P. K. Atrey, and S. Lyu, "Anti-forensics for face swapping videos via adversarial training," *IEEE Trans. Multimedia*, early access, Jul. 26, 2021, doi: 10.1109/TMM.2021.3098422.

[26] W.-L. Shang, Y. Chen, H. Bi, H. Zhang, C. Ma, and W. Y. Ochieng, "Statistical characteristics and community analysis of urban road networks," *Complexity*, vol. 2020, pp. 1–21, Sep. 2020.

[27] W. Lin, X. Ren, J. Hu, Y. He, Z. Li, and M. Tong, "Fast, robust and accurate posture detection algorithm based on Kalman filter and SSD for AGV," *Neurocomputing*, vol. 316, pp. 306–312, Nov. 2018.

[28] Q. Yang, Y. Lian, and W. Xie, "Hierarchical planning for multiple AGVs in warehouse based on global vision," *Simul. Model. Pract. Theory*, vol. 104, Nov. 2020, Art. no. 102124.

[29] Z. Ren, J. Lai, Z. Wu, and S. Xie, "Deep neural networks-based real-time optimal navigation for an automatic guided vehicle with static and dynamic obstacles," *Neurocomputing*, vol. 443, pp. 329–344, Jul. 2021.

[30] X. Chen, S. He, Y. Zhang, L. Tong, P. Shang, and X. Zhou, "Yard crane and AGV scheduling in automated container terminal: A multi-robot task allocation framework," *Transp. Res. C, Emerg. Technol.*, vol. 114, pp. 241–271, May 2020.

[31] J. C. W. Lin, S. Ren, and P. Fournier-Viger, "MEMU: More efficient algorithm to mine high average-utility patterns with multiple minimum average-utility thresholds," *IEEE Access*, vol. 6, pp. 7593–7609, 2018.

[32] H. Yu, Q. Wan, Z. Mu, Y. Du, and L. Liang, "Novel nano-scale absolute linear displacement measurement based on grating projection imaging," *Measurement*, vol. 182, Sep. 2021, Art. no. 109738.

[33] M. Famouri, Z. Azimifar, and A. Wong, "A novel motion plane-based approach to vehicle speed estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1237–1246, Apr. 2019.

[34] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, "High-resolution vehicle trajectory extraction and denoising from aerial videos," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 3190–3202, May 2021.

[35] Y. Gao, J. Li, Z. Xu, Z. Liu, X. Zhao, and J. Chen, "A novel image-based convolutional neural network approach for traffic congestion estimation," *Expert Syst. Appl.*, vol. 180, Oct. 2021, Art. no. 115037.

[36] W.-L. Shang, Y. Chen, X. Li, and W. Y. Ochieng, "Resilience analysis of urban road networks based on adaptive signal controls: Day-to-day traffic dynamics with deep reinforcement learning," *Complexity*, vol. 2020, pp. 1–19, Nov. 2020.

[37] W.-L. Shang, Y. Chen, C. Song, and W. Y. Ochieng, "Robustness analysis of urban road networks from topological and operational perspectives," *Math. Problems Eng.*, vol. 2020, pp. 1–12, Aug. 2020.

[38] L. Yang, W. Chen, H. Wang, and Y. Chen, "Deep learning seismic random noise attenuation via improved residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 7968–7981, 2021.

[39] S. Wei, L. Zhang, and H. Liu, "Integrated Kalman filter of accurate ranging and tracking with wideband radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8395–8411, Dec. 2020.

[40] M. B. Khalkhali, A. Vahedian, and H. S. Yazdi, "Situation assessment-augmented interactive Kalman filter for multi-vehicle tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3766–3776, Apr. 2022.

[41] X. Li *et al.*, "COCO-CN for cross-lingual image tagging, captioning, and retrieval," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2347–2360, Sep. 2019.

[42] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[43] X. Lu, B. Li, Y. Yue, Q. Li, and J. Yan, "Grid R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7355–7364.

Xinqiang Chen (Member, IEEE) received the Ph.D. degree in traffic information engineering and controlling from Shanghai Maritime University, China, 2018. From September 2015 to September 2016, he was a Visiting Student at the Smart Transportation Applications and Research Laboratory, University of Washington, USA. His research interests include transportation image processing and smart ship.

Zichuang Wang received the bachelor's degree from Guangzhou Maritime University in 2019. He is currently pursuing the M.S. degree with the Institute of Logistics Science and Engineering, Shanghai Maritime University, China. His research interests include traffic image processing and intelligent transportation systems.

Qiaozhi Hua received the B.E. degree in electrical communication from the Wuhan University of Science and Technology in 2011 and the M.S. and Ph.D. degrees from Waseda University in 2015 and 2019, respectively. He is currently a Lecturer with the Computer School, Hubei University of Arts and Science, Hubei, China. His research interests include game theory and wireless communications.

Wen-Long Shang (Member, IEEE) received the Ph.D. degree from the Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London. He is currently a Lecturer with the College of Metropolitan Transportation, Beijing University of Technology; and an Honorary Senior Research Fellow of Imperial College London. He has already published nearly 40 articles in peer-reviewed journals and conferences, obtained two software copyrights, and finished one chapters of a book. His research interests include but not limit to green transportation, traffic big data, intelligent transport systems, emergency management, safety science, and smart city. He serves as a lead guest editor and a member of editorial board for several international academic journals. He is a reviewer of more than 40 journals and conferences.

Qiang Luo received the M.S. and Ph.D. degrees in traffic information engineering and control from the School of Civil and Transportation, South China University of Technology, Guangzhou, China, in 2011 and 2014, respectively. His research interests include traffic data analysis, complex traffic networks, and intelligent transportation systems.

Keping Yu (Member, IEEE) received the M.E. and Ph.D. degrees from the Graduate School of Global Information and Telecommunication Studies, Waseda University, Tokyo, Japan, in 2012 and 2016, respectively.

He was a Research Associate, a Junior Researcher, and a Researcher with the Global Information and Telecommunication Institute, Waseda University, from 2015 to 2019, from 2019 to 2020, and from 2020 to 2022, respectively. He is currently an Associate Professor with the Graduate School of Science and Engineering, Hosei University, Tokyo. He has authored more than 100 publications, including papers in prestigious journal/conferences, such as the IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazine*, *IEEE Network* magazine, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, IEEE TRANSACTIONS ON RELIABILITY, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, *IEEE Consumer Electronics Magazine*, *IEEE Internet of Things Magazine*, ICC, and GLOBECOM. His research interests include smart grids, information-centric networking, the Internet of Things, artificial intelligence, blockchain, and information security. He has hosted and participated in more than ten projects. He is involved in many standardization activities organized by ITU-T and ICNRG of IRTF. He has contributed to ITU-T Standards Y.3071 and Supplement 35. He has received the IEEE Outstanding Leadership Award from IEEE BigDataSE 2021, the Best Paper Award from *IEEE Consumer Electronics Magazine* Award 2022 (First Place Winner), IEEE ICFTIC 2021, ITU Kaleidoscope 2020, and the Student Presentation Award from *JSST* 2014. He has served as the General Co-Chair and the Publicity Co-Chair for the IEEE VTC2020-Spring First EBTSRA Workshop, the General Co-Chair for IEEE ICCC2020 Second EBTSRA Workshop, the General Co-Chair for IEEE TrustCom2021 Third EBTSRA Workshop, and the Session Chair for IEEE ICCC2020, ITU Kaleidoscope 2016. He is an Associate Editor of IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, *Journal of Intelligent Manufacturing*, and *Journal of Circuits, Systems and Computers*. He has been a Guest Editor for IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS.