

ROAD USER LOCALIZATION FOR AUTONOMOUS VEHICLE INFRASTRUCTURE BY LEVERAGING SURVEILLANCE VIDEOS

Linjun Lu
Fei Dai

Wadsworth Department of Civil and Environmental Engineering
West Virginia University
1306 Evansdale Drive
Morgantown, WV 26506, USA

ABSTRACT

The emergence of autonomous vehicles (AVs) provides a sustainable solution to reshape the current transportation system and help mitigate the negative environmental impacts from transportation activities. However, the AVs may undergo unreliable and insufficient perception by the onboard sensors due to occlusion issues and complex traffic conditions, especially in crowded urban intersections. This study proposed a vision-based method for automated detection and localization of road users in traffic scenarios by leveraging surveillance videos. In this method, the traffic scenario is surveilled in a large visual range and the locations of all surveilled road users are determined. The field experiment was conducted to evaluate the performance of the proposed method. The experiment results demonstrated the promising accuracy of the proposed method for road user location estimation and its potential to provide the AVs with a full-participant perception in complex traffic scenarios and assist them to make the right driving decisions.

1 INTRODUCTION

Over the last three decades, greenhouse gas emissions from the transportation sector have been continuously increasing worldwide as primarily the result of the increased demand for the movement of people and goods. According to the recent report released by the U.S. Environmental Protection Agency, the greenhouse gas emissions from the transportation sector were responsible for 27.1% of total U.S. greenhouse gas emissions in 2020, having made it the leading contributor of U.S. greenhouse gas emissions (EPA 2022). In European countries, the transportation sector generated 31.3% of total EU-28 greenhouse gas emissions in 2019, which was the sixth consecutive year that transportation-related emissions had increased (Ghazouani et al. 2021). It was also estimated that transportation emissions need to fall by around two-thirds by 2050 in order to meet the net-zero greenhouse gas emission target as set out in the Transport White Paper 2011. In China, although the transportation sector only accounted for 9% of the country's total greenhouse gas emissions in 2019, it was still the second-largest contributor (following the U.S.) to global transportation emissions to date. Meanwhile, it is projected that the amounts of transportation emissions will continue to increase in the coming years on the account of the country's current low vehicle ownership level (Yin et al. 2015). The growing transportation emissions worldwide will contribute to the higher concentration of greenhouse gas trap in the atmosphere, which will accelerate global warming and natural climate change and consequently pose a significant threat to various species on the earth (Massar et al. 2021). Therefore, viable measures necessitate being taken to change the current transportation situation.

The emergence of autonomous vehicles (AVs) provides a sustainable solution to reshape the current transportation system and help mitigate the negative environmental impacts from transportation activities (Kopelias et al. 2020; Taiebat et al. 2018). By incorporating advanced vehicular automation, the AVs are

expected to operate in an automated eco-driving manner with smoother speed and fewer stop-to-go movements, thus capable of improving fuel efficiency and reducing transportation emissions thereof. In addition, the introduction of AVs also brings benefits in relation to improved safety, increased traffic capacity, reduced travel time, optimized traffic flow, and less traffic congestion, all of which will in turn contribute to less transportation emission indirectly. Greenblatt and Saxena (2015) carried out simulation experiments to estimate the emissions of shared electric AVs. The experiment results indicated that right-sizing electric AVs incorporated with a low-carbon electricity network could reduce greenhouse gas emissions by about 90% in comparison with the current conventionally driven vehicles. Stogios et al. (2019) developed an integrated traffic microsimulation and emission model to explore the effects of AVs on greenhouse gas emissions. It revealed that the emission intensities could be reduced by up to 26% on the expressway when the AVs operate in an aggressive driving setting (i.e., drive with low headway). Hwang and Song (2020) investigated the impacts of different AV penetration rates on road capacity and transportation emissions. The simulation results disclosed that the increase in AV penetration rate can significantly improve road capacity and reduce transportation emissions by up to 30%.

The enabler of autonomous driving primarily relies on the onboard sensors of AVs to perceive the surrounding environment and infrastructures and coordinate with other road users to make optimal real-time driving decisions (Van Brummelen et al. 2018). However, the AVs may undergo unreliable and insufficient perception due to occlusion issues and complex traffic conditions, especially in crowded urban intersections (Kloeker et al. 2020). This will result in false driving decisions by autonomous algorithms and pose serious safety risks to AVs and passengers. Although these issues can be partially solved by exploiting the vehicle-to-vehicle (V2V) communication technology, the safety risks remain because of the concurrently moving non-AVs and pedestrians, to which the communication units are not applied. Therefore, it will be of significance to provide the AVs with a full-participant perception in complex traffic scenarios. To this end, the authors are devoted to developing a vision-based method that allows for automatically detecting and localization of all road users using surveillance cameras. Multiple surveillance cameras are utilized to cover the entire area of desired traffic scene with each camera monitoring a certain region. By fusing and merging the visual information from different surveillance cameras, the real-time locations of all surveilled road users can be estimated and the relative location relationship between the AVs and other road users can be determined thereof. The perception results of entire traffic scenarios from the surveillance cameras are timely sent to the passing AVs through vehicle-to-infrastructure (V2I) to assist them to make the right driving decisions. The proposed method is promising and provides a measure in the infrastructure manner to help improve the reliability and safety of the AVs. In the meanwhile, the collected traffic data can be stored and utilized for simulation and further development of highly automated driving functions. In the scope of this study, the vision-based method for road user detection and localization was elaborated and its performance was preliminarily evaluated by one single camera for the surveillance of a traffic intersection.

2 METHODOLOGY

The proposed vision-based method contains three main modules: road user instance segmentation, 3D bounding box construction, and road user localization. First, the deep learning model – mask region-based convolutional neural network (Mask R-CNN) is used to detect and segment the road-user instances (silhouettes) in the images. Then, the 3D bounding box is constructed for each identified road user by leveraging the detected silhouette and three orthogonal vanishing points in the traffic scene. Constructing the 3D bounding box for objects can overcome the difficulty of accurate object location estimation in 2D images resulting from the absence of depth information, which is a shared hurdle in existing vision-based methods for road user location estimation. This way, the road user's location can be identified and represented by the four bottom vertexes (or better, the bottom surface) of its 3D bounding box. Finally, the road user's physical location is determined by mapping the four bottom vertexes from the image plane to the road plane. In the following subsections, each module is explained in detail.

2.1 Road-user Instance Segmentation Using Mask R-CNN

In this study, Mask R-CNN is applied to detect the instances of road users in the images, whose performance has been well evaluated and documented in many studies. Mask R-CNN is a two-stage framework: the first stage is responsible for feature extraction and generating object proposals, and the second stage is responsible for classifying the proposals to generate bounding boxes and masks in parallel (He et al. 2017). The architecture of Mask R-CNN is depicted in Figure 1. In the first stage, the feature pyramid network (FPN) integrates a bottom-up pathway and a top-down pathway to construct an in-network feature pyramid for each input image. By leveraging lateral connections (element-wise addition), multiple-scale feature maps are independently produced by merging the low-resolution but semantically strong features from the top-down and the high-resolution but semantically weak features from the bottom-up pathway. Therefore, this architecture enables the detection of objects across a large range of scales, particularly suitable for this study, that is, the large-size road users near and the small-size road users far from the camera. The extracted feature maps are then handed to the region proposal network (RPN), in which a series of object proposals are generated by pixel-wisely sliding a set of predefined anchors with aspect ratios of 0.5, 1, and 2 on the top of each feature map. After that, each generated proposal is mapped to a lower-dimensional feature and passed through two siblings fully connection layers, one pertaining to object-ness scores for classification and the other one to bounding box offset regression.

In the second stage, the region of interests align (ROIAlign) layer is implemented to cast the fixed-size (7×7) ROIs based on the proposals from the RPN and the shared feature maps from the FPN. In comparison to the ROI Pool layer adopted in Faster R-CNN, the ROIAlign layer subdivides the ROIs in a quantization-free manner by the use of bilinear interpolation; thus, it can faithfully preserve the exact per-pixel spatial locations of the proposals. It has been proven that the ROIAlign layer is capable of improving the accuracy of per-pixel segmentation masks by 10%-50% (He et al. 2017), thus allowing for obtaining more fine-grained road users' silhouettes that are greatly favorable to the proposed method. The generated ROIs are fed into the last network head embracing three branches for classification, bounding box regression, and instance segmentation. To achieve per-pixel segmentation without competition among different classes, the mask branch predicts a binary mask for each class independently. It is worth noting that the detection branch (classification and bounding box regression) and mask branch (instance segmentation) run in parallel to each other. Therefore, the classification of particular classes does not depend on the mask predictions.

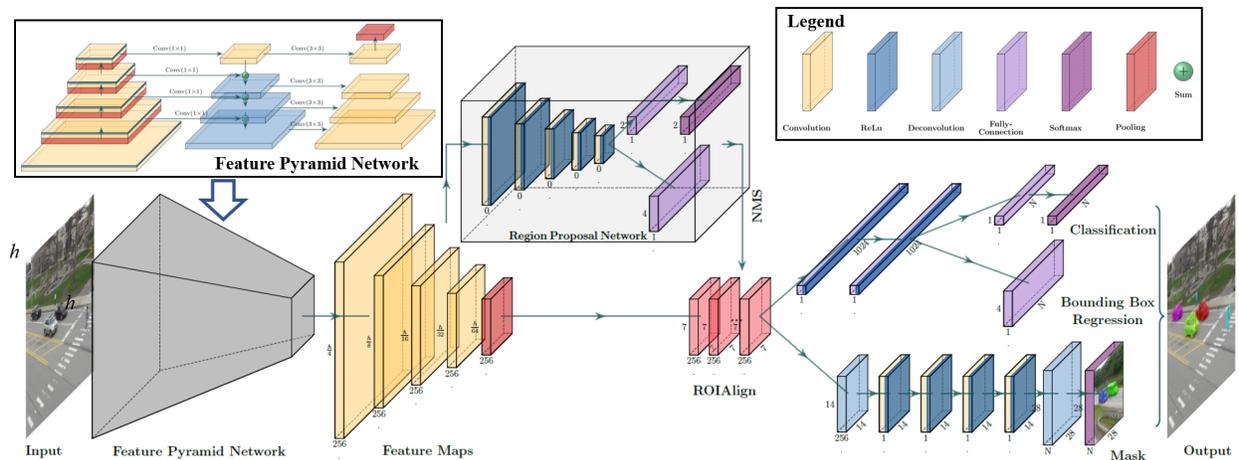


Figure 1: Architecture of Mask R-CNN.

2.2 3D Bounding Box Construction for Road Users

The pipeline of 3D bounding box construction was developed in the authors' previous work (Lu and Dai 2022), which consists of two steps: vanishing point estimation and 3D bounding box construction. In order to make this manuscript self-contained, the proposed pipeline is briefly illustrated herein. For details, interested readers can refer to Lu and Dai (2022).

As for the vanishing point estimation, a random sample consensus (RANSAC)-based method was proposed. Different from the traditional vanishing point estimation methods, the proposed method skips the step of line segment cluster and directly performs on the detected edgelets in the images, thus avoiding causing significant loss of line information along certain directions and introducing an unfavorable system basis. Also, it eliminates the need for trial-and-error parameter tuning thus allowing for the easy implementation of the proposed method across a variety of traffic monitoring scenes with much less human input. The edgelets distributed on the building outlines, road surface (i.e., those from road markers and lanes), and vehicle appearances can be utilized for vanishing point estimation. Herein, the three orthogonal vanishing points are defined as follows. The first vanishing point v_1 is in the direction of the dominant traffic flow, the second vanishing point v_2 is in the direction perpendicular to v_1 and parallel to the road surface, and the third vanishing point v_3 is in the direction perpendicular to the road surface. Note that v_1 , v_2 , and v_3 are vectors defined in homogenous coordinates. Figure 2 shows an example of identified concurrent edgelets along three orthogonal vanishing points, in which the edgelets along different directions are annotated by different colors.

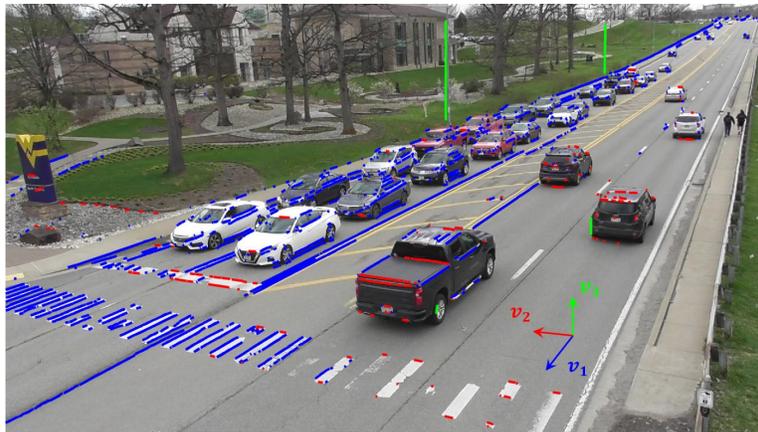


Figure 2: An example of edgelet detection and vanishing point estimation in traffic scene.

Upon identification of the three orthogonal vanishing points in the traffic scene, the 3D bounding box can be constructed for each road user in the surveilled view. Taking an on-road truck for illustration, Figure 3 depicts the schematic pipeline of 3D bounding box construction. First, five lines (L1~5) are generated by constructing the lines passing through three vanishing points and tangent to the truck's silhouette (see Figure 3a). based on this, four intersection points (B1~4) can be obtained, which correspond to the four vertices of the truck's 3D bounding box. Following that, line L6 passing through the point B1 and the vanishing point v_2 is constructed. Together with line L7 passing through the point B3 and vanishing point v_3 , the vertex B5 can be subsequently determined, which lies at the intersection of lines L6 and L7 (see Figure 3b). Likewise, vertex B6 is constructed as the intersection of line L8 passing through the vertex B2 and the vanishing point v_1 and the line L9 passing through the vertex B4 and vanishing point v_2 . With the known vertex B6, the vertex B7 can be obtained from the intersection of the line L1 and the line L11 passing through the vertex B6 and the vanishing point v_3 (see Figure 3c). Similarly, the final vertex B8 can be defined by the intersection of line L4 and line L10 passing through the vertex B5 and the vanishing point v_1 . Once all eight vertexes are obtained, the truck's 3D bounding box can be enclosed eventually by

connecting these vertexes recursively (see Figure 3d). The same 3D bounding box construction strategy can be also applied to other road users, such as cars, buses, vans, pedestrians, motorcycles, and cyclists. Figure 4 shows some examples of 3D bounding box construction for different road users.

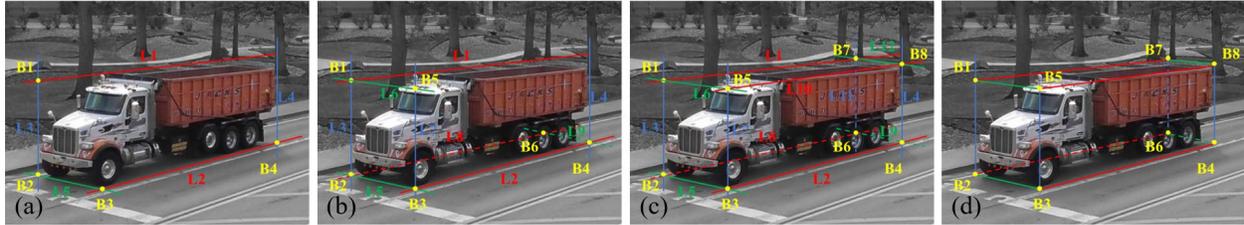


Figure 3: Pipeline of constructing 3D bounding box.

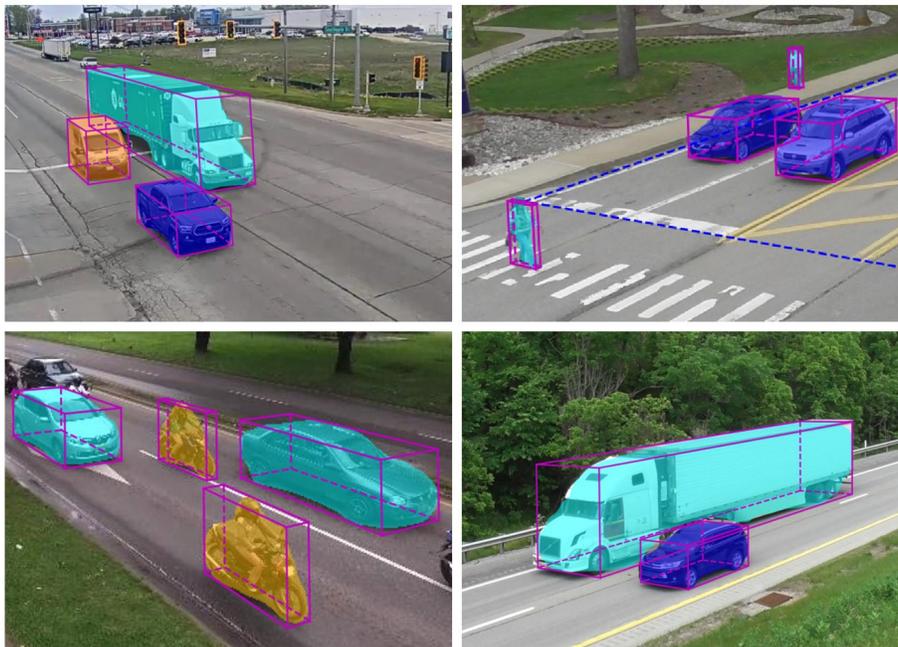


Figure 4: Examples of 3D bounding box construction for road users.

2.3 Localization of Surveilled Road Users

Once the road user's 3D bounding box has been constructed, its location on the road can be represented by the four bottom vertexes (or better, bottom surface) of its 3D bounding box along with converting them into the physical coordinates in the real world. Assuming that the road shape is approximately flat, which stands in most traffic situations, the projective relationship between the road plane and image plane can be described as (Lu and Dai 2021)

$$\gamma \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix}. \quad (1)$$

Or more concisely

$$\gamma \mathbf{m}_i = \mathbf{H} \mathbf{M}_i. \quad (2)$$

where $\mathbf{m}_i = (x_i, y_i, 1)^T$ and $\mathbf{M}_i = (X_i, Y_i, 1)^T$ ($i = 1, 2, \dots, n$) are the homogeneous coordinates of i th pair of point correspondences on the image plane and road plane, respectively. γ is the non-zero scale factor. \mathbf{H} stands for the homography matrix, depicting the projective transformation relationship between the image plane and road plane. Since \mathbf{H} is meaningfully defined up the scale factor γ , there are only eight degrees of freedom in it and thus \mathbf{H} can be determined from at least four image-to-world point correspondences in general condition with the help of the direct linear transformation method (Lu and Dai 2021).

The reference points used for homography estimation are shown in Figure 5, whose image coordinates and physical coordinates were manually measured on the image plane and road plane, respectively. By exploiting the obtained homography matrix, each road user's location on the road can be determined by mapping the bottom vertexes of its 3D bounding box from the image plane to the road plane. An example is shown in Figure 6.

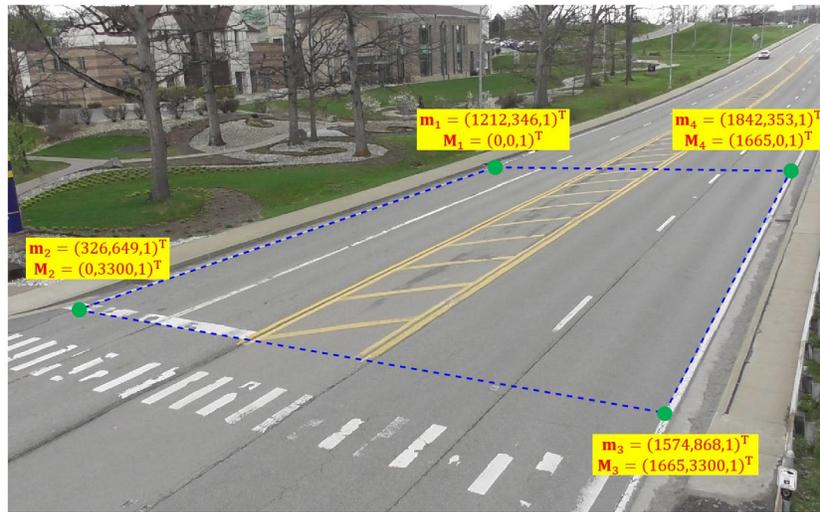


Figure 5: Reference points for homography estimation.

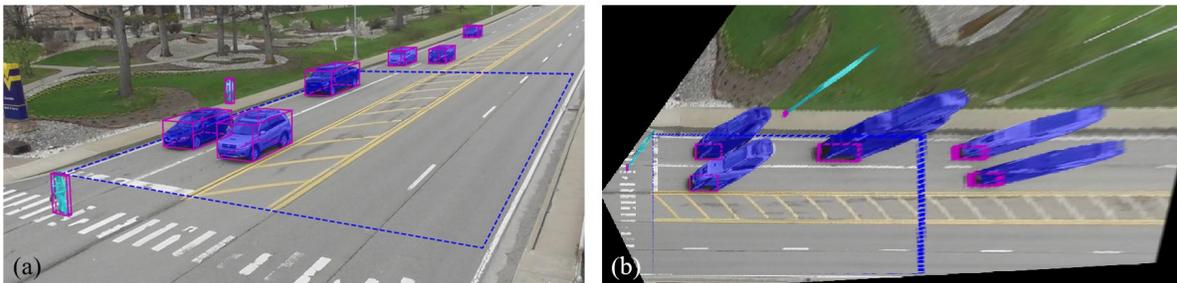


Figure 6: Road users' location on road by applying projective mapping: (a) before projective mapping and (b) after projective mapping.

3 DATASET CONSTRUCTION AND MODEL TRAINING

An urban traffic dataset was established by the authors for Mask R-CNN model training, which contains 684 images at a resolution of 1920×1080 pixels. The traffic images were recorded under different traffic scenarios and surveillance view angles. The labeling work was done by the VGG Image Annotator, which is an image annotation software developed by the Visual Geometry Group (VGG) (Dutta and Zisserman 2019). The road user instances are grouped into seven categories, i.e., car, van, bus, truck, pedestrian,

cyclist, and motorcycle. Some collected urban traffic image samples as well as their annotations are displayed in Figure 7.

The model training task was conducted on the platform of Amazon Web Services (AWS) with a single graphics processing unit (GPU, NVIDIA Tesla V100). The residual neural network ResNet-101 was employed as the backbone of FPN in the Mask R-CNN model, on the account of the fact that it outperforms other deep convolutional architectures in tackling the gradient vanishing issues (He et al. 2016). The pre-training strategy was used to achieve better model performance and reduce the training time. To this end, the Mask R-CNN model was firstly trained from scratch on the COCO dataset (Lin et al. 2014), and then the weights and biases in the network head were finetuned on the established urban traffic dataset. The established traffic dataset was split into a ratio of 7:3 (479:205) for model training and validation. The hyperparameters, that is, learning rate, learning momentum, weight decay, and batch size were set as 0.001, 0.9, 0.0001, and 2, respectively.



Figure 7: Collected urban traffic image samples and their annotations.

4 PERFORMANCE VALIDATION

4.1 Experimental setup

The field experiment was carried out on the campus of West Virginia University to preliminarily evaluate the performance of the proposed method for road users' localization. One commercial camera (Panasonic HC-W580K) was employed to surveil the traffic scene in a full-HD resolution (1920×1080 pixels) with a sampling rate of 30 fps. The camera was installed on one pole at an elevated height of 6 m above the road surface. The surveilled traffic scene is shown in Figure 8(a). In the meanwhile, one unmanned aerial vehicle (UAV, Phantom 4 Pro+ V2.0) equipped with a high-resolution camera was utilized to surveil the same traffic scene also in a bird's eye view, whose clock was synchronized with that of the Panasonic camera by the Internet time server. By setting the optical axis of the UAV perpendicular to the road surface, the road users' locations can be easily and accurately identified in the images recorded from the UAV via the 2D bounding boxes, as illustrated in Figure 8(b). Therefore, the road users' locations estimated from the UAV served as the ground truths for comparison purposes. The resolution of the onboard camera is 3840×2160 pixels and the sampling rate was set to 30 fps. By leveraging the same reference points as set in Figure 5, the projective relationship between the ground plane and the image plane of the onboard camera can be estimated through Equation (2). By doing so, the road users' locations can be converted from the pixel unit to the physical unit. However, due to the self-motion of the UAV, the corresponding projective relationship differs for different frames. Therefore, the homography matrix needs to be estimated for each frame individually.

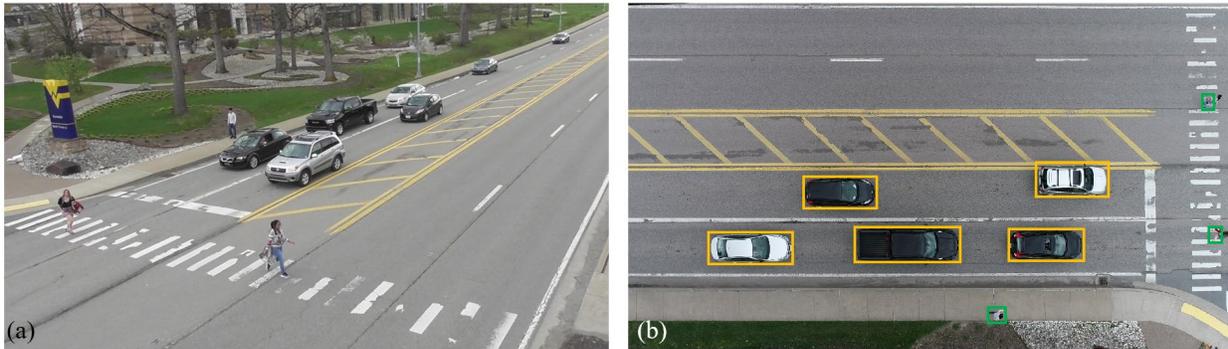


Figure 8: Dataset collection in field experiment: (a) dataset collected from surveillance camera and (b) dataset collected from UAV.

4.2 Experiment results

Two video datasets were recorded from the surveillance camera and UAV in the field experiment, respectively. The video duration of either dataset was 10 minutes. During data postprocessing, both videos were converted into a sequence of images with a resampling rate of 2 s, resulting in 300 images for each video dataset. The image dataset from the surveillance camera was fed into the trained Mask R-CNN model to test its performance for road user detection and instance segmentation. It should be pointed out that no frame in the field video dataset was included in the model training process. Namely, the trained Mask R-CNN model has never “seen” these testing images before. The testing result is depicted in Figure 9 in the form of a confusion matrix, in which the cells in the first row except for the first one correspond to the false positives such that the textures in the image background (BG) were falsely predicted as road users, and the cells in the first column except for the first one correspond to the false negatives such that the actual road users were not successfully detected. The road users regarding cyclists and motorcycles did not appear in the field experiments, thus the performance of the trained model for these two types of road users was not validated. It can be seen that the overall precision and recall for the recognition of each type of road user were above 87.5% and 82.4%, respectively.

Road User Classification Using Mask RCNN

True Class	Confusion Matrix						Recall	
	BG	Car	Bus	Van	Truck	Pedestrian		
BG	300						100.0%	
Car	22	767				10	96.0%	4.0%
Bus			35	2	1		92.1%	7.9%
Van		2	1	14			82.4%	17.6%
Truck	6	4			123		92.5%	7.5%
Pedestrian	9					340	97.4%	2.6%

		Precision					
		BG	Car	Bus	Van	Truck	Pedestrian
Predicted Class	BG	89.0%	99.2%	97.2%	87.5%	91.8%	100.0%
	Car	11.0%	0.8%	2.8%	12.5%	8.2%	

Figure 9: Confusion matrix of road user recognition results.

Table 1 lists the statistics of localization errors of different road users by the proposed method with reference to the ground truths obtained from the UAV. The localization error of each object was computed by averaging the differences between the measured locations of four bottom vertexes and their actual values. The apparently false 3D bounding boxes of road users resulting from occlusion issues were excluded for performance validation. It is worth pointing out that this issue can be easily tackled by setting other surveillance cameras at different locations and surveillance view angles. The results showed that the mean error maximum error of road user localization errors were 12.3 and 20.5, 20.8 and 30.7, 17.2 and 26.8, 23.7 and 35.4, 8.0 and 19.8 cm, respective. Additionally, the standard deviation of localization error for these five types of road users were 4.6, 6.8, 6.7, 7.3, and 3.1 cm, respectively. It is evident that the localization error of pedestrians is much less than those of other types of road users, it is mainly attributed to its less occupation of road spacing that makes it more easily and accurately identified on the road surface.

Table 1: Statistics of road user localization errors (cm).

	Mean Error	Standard Deviation	Maximum Error
Car	12.3	4.6	20.5
Bus	20.8	6.8	30.7
Van	17.2	6.7	26.8
Truck	23.7	7.3	35.4
Pedestrian	8.0	3.1	19.8

5 DISCUSSIONS

Once the road users' locations are identified from the proposed method, their dynamic characteristics, such as trajectory, velocity, acceleration, etc., can be easily obtained by tracking their locations over a sequence of frames. This valuable information can then be sent to AVs in real time through V2I technology so that the AVs can have a full-participant understanding of the entire traffic scenario to help them make the right driving decisions. Furthermore, all types of data collected can be processed and stored in a central database and used for traffic and transportation infrastructure simulations as well as the development of highly automated driving functions. By utilizing the collected dataset, the responses of road users to others can also be thoroughly studied and as an input for simulation. Collectively, virtual traffic scenarios can be established to test and optimize the driving algorithms for AVs in response to the motions of surrounding road users. In addition, since the proposed method is able to achieve a full-participant perception of the surveilled traffic scenes, the collected dataset can be used to conduct the simulation experiments and develop the macroscopic models to optimize the route planning algorithms for AVs, which can help them travel through complicated traffic situations, such as crowded intersections, more smoothly and effectively.

6 CONCLUSIONS AND FUTURE WORKS

In this study, a vision-based method was proposed for the automated detection and localization of road users in urban traffic scenarios. The field testing was conducted to preliminary evaluate the performance of the proposed method. The experiment results demonstrated the promising accuracy of the proposed method for road user localization and its potential to provide the AVs with a full-participant perception in complex traffic scenarios and assist them to make the right driving decisions.

On top of the current study, there are still some works worth future investigation. First, more comprehensive field experiments need to be carried out to evaluate and improve the performance of the proposed method under different traffic conditions. Second, in order to recover the entire traffic scenario and address the occlusion issues, multiple surveillance cameras need to be installed at different locations and with different surveillance view angles. Therefore, some efforts should be made to develop guidance for camera location choice and surveillance view angle optimization. Based upon this, the information fusion and integration from different cameras also deserve further investigation in the future.

ACKNOWLEDGMENTS

This work was sponsored by a grant from the Center for Integrated Asset Management for Multimodal Transportation Infrastructure Systems (CIAMTIS), a US Department of Transportation University Transportation Center, under federal grant number 69A3551847103. The authors are grateful for the support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the CIAMTIS.

REFERENCES

- Dutta, A., and A. Zisserman. 2019. "The VIA Annotation Software for Images, Audio and Video". In *Proceedings of the 27th ACM International Conference on Multimedia*, 2276–2279. New York, NY: Association for Computing Machinery.
- United States Environmental Protection Agency. 2022. "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2020".
- Ghazouani, A., M. B. Jebli, and U. Shahzad. 2021. "Impacts of Environmental Taxes and Technologies on Greenhouse Gas Emissions: Contextual Evidence from Leading Emitter European Countries". *Environmental Science and Pollution Research* 28(18):22758–22767.
- Greenblatt, J. B., and S. Saxena. 2015. "Autonomous Taxis Could Greatly Reduce Greenhouse-Gas Emissions of US Light-Duty Vehicles". *Nature Climate Change* 5(9):860–863.
- He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. "Mask R-CNN". In *Proceedings of the 2017 IEEE International Conference on Computer Vision*, 2961–2969. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition". In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Hwang, H., and C. K. Song. 2020. "Changes in Air Pollutant Emissions from Road Vehicles Due to Autonomous Driving Technology: A Conceptual Modeling Approach". *Environmental Engineering Research* 25(3):366–373.
- Kloeker, L., A. Kloeker, F. Thomsen, A. Erraji, and L. Eckstein. 2020. "Traffic Detection Using Modular Infrastructure Sensors as A Data Basis for Highly Automated and Connected Driving". In *Proceedings of the 29th Aachen Colloquium Sustainable Mobility*, 1835–1844. Aachen: Institute for Automotive Engineering, RWTH Aachen University.
- Kopelias, P., E. Demiridi, K. Vogiatzis, A. Skabardonis, and V. Zafropoulou. 2020. "Connected & Autonomous Vehicles—Environmental Impacts—A Review". *Science of The Total Environment* 712:135237.
- Lin, T. Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. 2014. "Microsoft COCO: Common Objects in Context". In *Proceedings of the 2014 European Conference on Computer Vision*, 740–755. Heidelberg, New York, Dordrecht, London: Springer Cham.
- Lu, L., and F. Dai. 2021. "A Unified Normalization Method for Homography Estimation Using Combined Point and Line Correspondences". *Computer-Aided Civil and Infrastructure Engineering* 37(8):1010–1026.
- Lu, L., and F. Dai. 2022. "Automated Visual Surveying of Vehicle Heights to Help Measure the Risk of Overheight Collisions Using Deep Learning and View Geometry". *Computer-Aided Civil and Infrastructure Engineering*. <https://onlinelibrary.wiley.com/doi/abs/10.1111/micc.12842>, accessed 12th April 2022.
- Massar, M., I. Reza, S. M. Rahman, S. M. H. Abdullah, A. Jamal, and F. S. Al-Ismael. 2021. "Impacts of Autonomous Vehicles on Greenhouse Gas Emissions—Positive or Negative?". *Environmental Science and Engineering* 18(11):5567.
- Stogios, C., D. Kasraian, M. J. Roorda, and M. Hatzopoulou. 2019. "Simulating Impacts of Automated Driving Behavior and Traffic Conditions on Vehicle Emissions". *Transportation Research Part D: Transport and Environment* 76:176–192.
- Taiebat, M., A. L. Brown, H. R. Safford, S. Qu, and M. Xu. 2018. "A Review on Energy, Environmental, and Sustainability Implications of Connected and Automated Vehicles". *Environmental Science & Technology* 52(20):11449–11465.
- Van Brummelen, J., M. O'Brien, D. Gruyer, and H. Najjaran. 2018. "Autonomous Vehicle Perception: The Technology of Today and Tomorrow". *Transportation Research Part C: Emerging Technologies* 89:384–406.
- Yin, X., W. Chen, J. Eom, L. E. Clarke, S. H. Kim, P. L. Yu, S. Patel, and G. P. Kyle. 2015. "China's Transportation Energy Consumption and CO₂ Emissions from A Global Perspective". *Energy Policy* 82:233–248.

AUTHOR BIOGRAPHIES

LINJUN LU is a PhD student in the Wadsworth Department of Civil and Environmental Engineering at West Virginia University. His research interests include computer vision and machine learning for automated construction, infrastructure operations, and structure health monitoring. His email address is l10074@mix.wvu.edu.

FEI DAI is an Associate Professor in the Wadsworth Department of Civil and Environmental Engineering at West Virginia University. His research interest focuses on applying advanced information and sensing technologies to improving performance of

Lu and Dai

current practice in civil and construction engineering. Particularly, he is interested in applied photogrammetry in 3D modeling, quantity surveying, and augmented reality for construction engineering applications; discrete-event simulation, evolutionary optimization, and visualization to develop simplified yet practical approaches for construction project and operations planning. His email address is fei.dai@mail.wvu.edu and his homepage is <https://feidai.faculty.wvu.edu>.