Smart Curb Digital Twin: Inventorying Curb Environments Using Computer Vision and Street Imagery

Haiyan Hao and Yan Wang

***Abstract*—Digital Twin (DT) offers a novel framework to track, model, analyze, and anticipate complex urban processes and support data-driven decision-making. However, a premise of developing DT applications is to inventory physical urban built environment digitally, which are often lacking for small- and medium-sized cities due to limited resources. Particularly, few digital inventories have been built for urban curb environments, which have been increasingly challenged by new vehicle technologies and emerging mobility services. We propose a data-driven framework to inventory curb facilities across types and locations using computer vision (CV) and Google Street View (GSV) imagery. Specifically, we used a state-of-the-art semantic segmentation model, i.e., DeepLab V3, pre-trained on the CityScapes dataset, to detect curb facilities of interest from GSV images. We then used the Inverse Perspective Mapping (IPM) to estimate the spatial location for each detected facility and used spatial processing to aggregate and filter estimation results. We demonstrated the framework for inventorying curbs in the Innovation District in the City of Gainesville, FL. The preliminary research contributes to Smart Curb Digital Twin for more safe, accessible, and productive curb environments.**

***Index Terms*—Curb environment, computer vision, digital twin, street imagery, smart cities.**

I. INTRODUCTION

CURBS are spaces that separate vehicular and pedestrian flows. Curbs locate important urban assets used by travelers to switch transportation means, residents to access curbside properties, drivers to park vehicles, municipalities to place public facilities, and so forth. Under the burgeoning smart city initiatives, recent advances in vehicular technologies and mobility services, such as shared micro-mobility and electric vehicle (EV), have complicated curb environments with different private and public sectors placing their facilities (e.g., EV charging stations and bike racks) [1], [2], [3]. These new facilities have turned curbs contested spaces for more intensive and diverse activities that can be difficult to manage [4]. Consequently, cities are in increasing need of up-to-date strategies to design, regulate, and manage urban curbs and ensure the efficiency and equity of different curb uses [5].

Manuscript received 22 July 2022; revised 1 September 2022; accepted 1 September 2022. Date of publication 30 November 2022; date of current version 28 June 2023. This work was supported by the National Science Foundation (NSF) through SCC-PG: SmartCurb: Building Smart Urban Curb Environments under Award 2124858. *(Corresponding author: Yan Wang.)*

The authors are with the Department of Urban and Regional Planning, University of Florida, Gainesville, FL 32611 USA (e-mail: hhao@ufl.edu; yanw@ufl.edu).

Digital Object Identifier 10.1109/JRFID.2022.3225733

Digital twin (DT) empowers urban and transportation practitioners by creating virtual replicas for the urban built environment and enabling the tracking, modeling, analyzing, and prediction of complex urban processes [6]. Researchers across disciplines have employed the DT framework for energy monitoring, building information management, disaster management, construction management, and so forth [6], [7]. Curb management can benefit from DT in the following aspects. First, DT can present urban stakeholders with the spatial layout of curb environments, which informs the decision-making of placements of new curb facilities such as EV charging stations and micro-mobility racks. Second, DT can model human-infrastructure interaction and simulate curb-use conditions across scenarios, which enables the evaluation of the performance of curb designs and regulations [8]. Third, DT supports the development of data-driven management systems which can track the real-time operation and usage conditions of different curbs, leading to more coordinated and efficient curb uses among different stakeholders.

However, a salient barrier to developing DT applications for urban curbs is the lack of the digitalization of curb environments. A few public institutions and private companies, e.g., SharedStreets and Coords, have recognized the challenges and started to inventory curb environments for metropolitans [9]. However, the data collection mainly relied on hired surveyors or crowdsourcing, which can be labor-intensive and expensive for small and medium-sized cities with increasing populations. Some recent research in landscape and urban planning has shown the potential of using street view imagery in auditing urban environments. For example, a few studies explored approaches to automatically identify curb facilities, including roadway signs [10], trees [11], and sidewalks [12], from street view images leveraging either conventional image processing techniques or computer vision (CV) algorithms. However, these approaches mostly focused on a single object and may require additional data (e.g., LiDAR and satellite images) for accurate locating [13].

Given the presented needs and challenges, we present a low-cost, fast, and automated data inventory framework for Smart Curb Digital Twin. The framework extracts various discrete and continuous curbside furniture, e.g., poles and sidewalks, from Google Street View (GSV) imagery and processes them into geospatial inventory that can be regularly updated with the updating of GSV. We demonstrated the application of this framework with GSV imagery collected from the Innovation

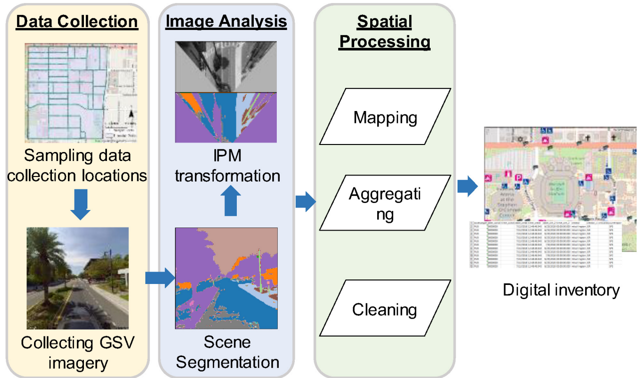

Fig. 1. Framework design.

District in the City of Gainesville, FL, which includes diverse land-use patches characterized by different curb environments.

II. Framework Design

The proposed framework is shown in Fig. 1. It mainly includes two steps, i.e., 1) detecting and locating curb furniture from individual GSV images and 2) aggregating detection results from all GSV images for the digital inventory, which will be introduced in the following subsections.

A. Locating Discrete and Continuous Objects From Individual GSV Images

We show the pipeline for the first step in Fig. 2. For each obtained GSV image, we first performed semantic segmentation to detect the presence of curb facilities of interest. We compared different pre-trained models in Detectron2, a platform including various off-the-shelf CV models [14]. We selected the DeepLab V3 model pre-trained on the CityScapes dataset for semantic segmentation. DeepLab is a scene segmentation architecture that uses atrous convolution to improve the performance of the model in segmenting objects at multiple scales [15]. Its recent version, i.e., DeepLab V3, shows state-of-the-art performance in various scene segmentation benchmarks. CityScapes consists of more than 20,000 annotated urban scene images collected from 50 cities and across different seasons [16]. The combination of DeepLab V3 and CityScapes is hence capable to digitalize urban built environment.

This pre-trained model can identify several curb facilities of interest, e.g., traffic signs, traffic lighting, poles, and sidewalk. We distinguished two types of objects at curb spaces, i.e., discrete (e.g., poles) and continuous (e.g., sidewalks), and adopted different processing procedures (Fig. 2). Specifically, for discrete objects, we performed a set of image processing, including removing small holes (i.e., incontinuity), removing small objects (i.e., noises), filtering, and labeling, to clean and segment identified objects of the same category on the same image. We then draw bounding boxes for individual objects and estimate their distances with Inverse Perspective Mapping (IPM).

IPM is a technique that computes the top-view perspective of images from other perspectives with known camera specifications (i.e., focal length, resolution, and pixel size) and installation parameters (i.e., camera height and angles) [17] (Fig. 3). IPM can be used to estimate the distance of the object from the camera with the camera projection matrix, which project objects from the 3-D world system to 2-D camera system, as in:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{h}\begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{00} & r_{01} & r_{02} & t_0 \\ r_{10} & r_{11} & r_{12} & t_1 \\ r_{20} & r_{21} & r_{22} & t_2 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where, X, Y, and Z are the 3-dimension location of the object in the world coordinate system; u and v are object locations in the image plane, measured with lateral and vertical pixels; u_0 and v_0 are the location of the camera center in the image plane, measured with lateral and vertical pixels; f_x and f_y are the focal lengths of the camera, measured with lateral and vertical pixels; h is camera height from the ground (Fig. 3), measured in meters; $\boldsymbol{r}$ is the rotation matrix determined by the camera's yaw angle and pitch (Fig. 3); and $\boldsymbol{t}$ is transformation vector denoting the location differences between origins for the world coordinate system and camera coordinate system.$\boldsymbol{t}$ is the zero vector when we set the origin of the world coordinate system as the origin for the camera coordinate system (Fig. 3)

For the simplest case, when the camera's yaw angle and pitch were set to zero, i.e., $\boldsymbol{r}$ equals the zero matrix, and $\boldsymbol{t}$ is also set as the zero vector. Then the lateral (i.e., x axis in Fig. 3) and vertical (i.e., y axis in Fig. 3) offsets of the detected object in relation to the camera can be calculated with (2):

$$\begin{cases} X = \frac{h(u-u_0)}{f_u} \\ Y = \frac{h(v-v_0)}{f_v} \end{cases} \tag{2}$$

In this research, we calculated the lateral and vertical offset of discrete objects with their middle point of the lower bounds of bounding boxes. For continuous objects such as sidewalks, we firstly performed the IPM to convert the inclined GSV image to top-down views, and then performed similar cleaning and segmenting procedures as the discrete objects as shown in Fig. 2. In this way, the bounding box can tightly fit the identified continuous object. We then calculated the lateral and vertical offsets for the start and end points of the closest bound for continuous objects. As the accuracy of IPM estimation reduces with the increasing distance, we removed all discrete objects with offsets further than 12 meters and clipped the length of continuous objects to not exceed 15 meters.

B. Aggregate Detection Results for Curb Inventory

The step described in Section II-A converts each GSV image into a set of discrete and continuous objects associated with the lateral and vertical offsets with respect to the camera. For each GSV image, we then mapped the spatial location of the detected objects with their relative offsets, the location (i.e., longitude and latitude) of the camera, and the heading angle of the camera. As we used a small spatial interval (i.e., 10 meters)

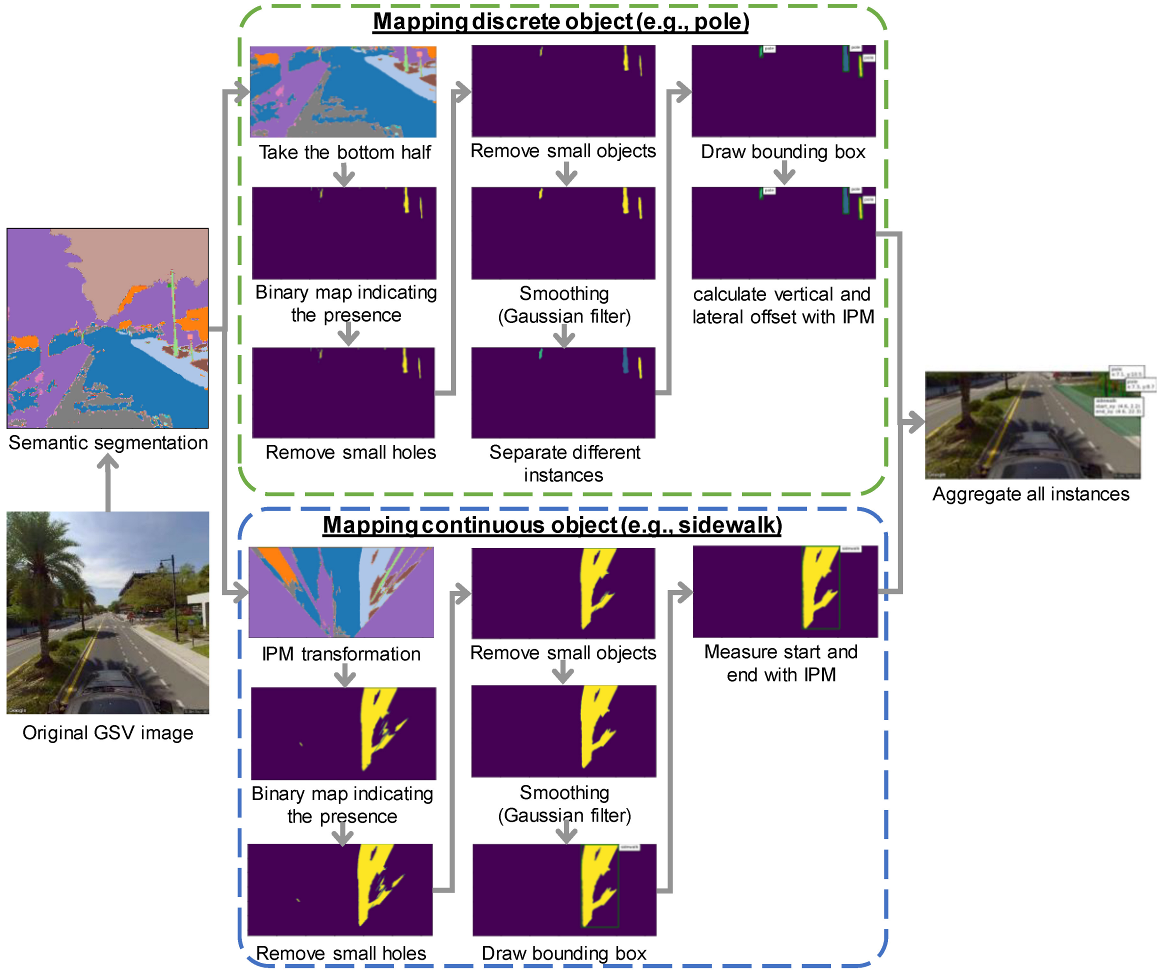

Fig. 2. The pipeline of mapping discrete and continuous objects from Google Street View imagery.

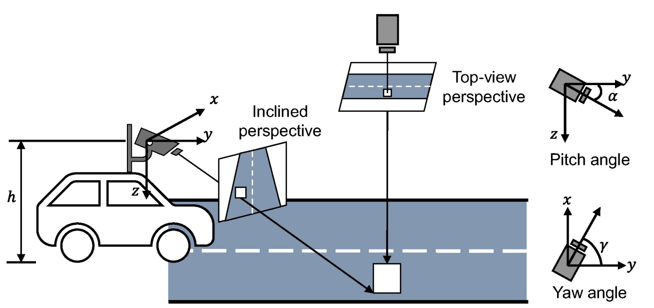

Fig. 3. Illustration of inverse perspective mapping for depth estimation.

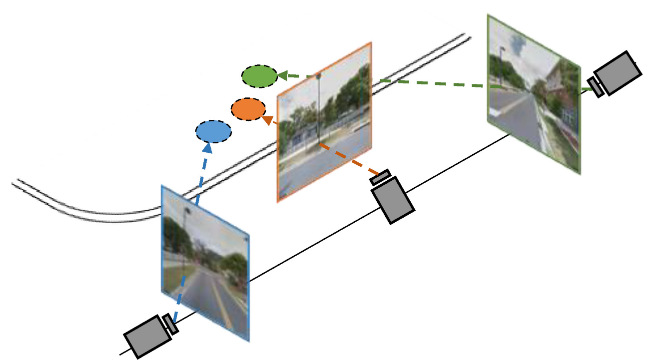

Fig. 4. Aggregating detection results from different GSV images.

to sample GSV images, one object will be captured by multiple GSV images from different angles and, hence, are associated with multiple estimated locations as Fig. 4 shows. For discrete objects, we used spatial clustering to group location estimations for the same object. We then counted the occurrence for each object and removed the ones that only occurred once, which means that object was only found in a single GSV image and may be misclassifications made by the CV algorithm. We used the averaging coordinates of the different location estimations as the final spatial location for discrete objects. For continuous objects that are processed into line features, we combined the segments that are located within one meter.

III. CASE APPLICATION

We demonstrated the application of this framework with data collected from the Innovation District in the City of Gainesville (Fig. 5). The Innovation District is located between the city's downtown and the University of Florida. The district consists of diverse land use patches, e.g., residential, commercial, official, and mixed-use, that are associated with disparate curb environments. The district also has a high share of micro-mobility trips that transport students between campus and downtown, which results in more frequent curb uses and demands novel curb management strategies.

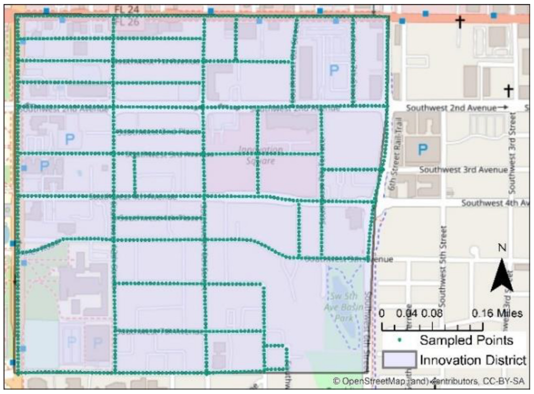

Fig. 5. The map for sampled coordinates for GSV image query.

Fig. 6. Mapping result for (a) discrete objects and (b) continuous objects.

Specifically, we sampled coordinates for data collection along roadways in Innovation District with the "Generate Points Along Lines" function in ArcGIS Desktop 10.7. We set the distance between nearby sampling points to be 10 meters and identified 1,110 sampling coordinates in the study region (Fig. 5). We then used the coordinates of the sampled points to query the closest available street-view imagery with Google Street View Application Programming Interface (API). For each coordinate, we queried four street-view images corresponding to the four heading angles of 0°, 90°, 180°, and 270°. Note that 0° represents north. Each image is of 640 * 640 pixels resolution, taken at the pitch of 0°, and with a 90° field of view. We also acquired the actual coordinates of GSV images referring to the associated metadata. Additionally, we back-calculated the camera projective transformation matrix, including the focal length and camera height, by calibrating (1) with a few field measurements conducted at the pilot site.

Fig. 6 shows the mapping results for discrete objects (i.e., poles and traffic signs) and continuous objects (i.e., sidewalks) identified for the study region. For discrete objects, the 4,440 GSV images initially yield 1,943 detected objects (with repetitions). 188 different objects were identified to be captured by more than one GSV image and were displayed in Fig. 6a. The identified poles are generally street lighting, utility poles, and posts for non-traffic signs. As Gainesville does not archive data for poles and traffic signs, we manually measured the disparities between the real and estimated object locations with Google Map's "Measure Distance" function. It was shown that the algorithm can achieve 76.5% precision for object detection and 0 - 4m accuracy for spatial mapping (Fig. 6). False positive cases include advertisement bands, road barriers, and bollards.

For continuous objects, the algorithm initially yielded 773 line segments (with repetition) for sidewalks, which are displayed in Fig. 6b together with the actual sidewalk data collected by the city. It was shown that the pre-trained CV model can better identify sidewalks located near commercial or office blocks than residential blocks. The algorithm misclassifies medians as sidewalks at some locations.

IV. Discussion and Conclusion

The emerging disruptive technologies will make urban environments more complex and dynamic. Digital Twin (DT) has great potential to contribute to more smart, efficient, and sustainable cities by empowering decision-makers to make data-driven decisions. The presented data inventory framework for Smart Curb Digital Twin provides a fast and economic solution to digitalize curb environments and, therefore, assists local governments in developing DT applications for the planning and management of urban curbs. The framework is applied in a real-world test site, i.e., the Innovation District in Gainesville, FL. Based on the experiments, we identified the following improvements that can be considered in future work.

First, we used an off-the-shelf model, i.e., DeepLab V3 pre-trained on the CityScapes dataset, for the detection of curb objects. Though CityScapes is among one of the few datasets for urban scene understanding, it was not specifically designed for the curb environment. Only a few categories of curb facilities are annotated in this dataset. Future studies may tune the model with street-view images specifically annotated for curb environments. It is also recommended to include up-to-date categories of curb facilities in the annotation, such as EV charging stations and shared micro-mobility parking stations. Second, we used a conventional technique, IPM, to estimate the location of identified curb facilities. One challenge we met is the unknown specification and installation parameters of GSV cameras, which is critical for accurate location estimation with IPM. To address this, we did field measurements and back-calculated the parameters in this study. Note that the acquired GSV images can be from different years and collected by different GSV vehicles with different cameras specification and/or installation parameters. Also, some GSV images are captured during rainy and cloudy days and are of low image quality. These limitations can be easily overcome when cities deploy their own data collection vehicles and operate on sunny days. Additionally, many cities also started to operate autonomous shuttles with sensors frequently scanning surrounding environments including curbs. Future studies may also consider leveraging these data to know the real-time conditions of curb facilities [18], [19], and integrating human dynamics using user-generated data to track curb use patterns [20], [21]. With these improvements, we believe our low-cost and fast inventory framework will play an important role in bridging the real city and its digital copy, as well as mitigating the uneven technological developments across different cities.

ACKNOWLEDGMENT

The authors acknowledge the NSF-funded Smart City Digital Twin Convergence Conference for the opportunity to present the work. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] G. Marsden, I. Docherty, and R. Dowling, "Parking futures: Curbside management in the era of 'new mobility' services in British and Australian cities," *Land Use Policy*, vol. 91, Feb. 2020, Art. no. 104012, doi: 10.1016/j.landusepol.2019.05.031.

[2] M. F. Mitman, S. Davis, I. B. Armet, and E. Knopf, "Curbside management practitioners guide," Inst. Transport. Eng., Washington, DC, USA, 2018.

[3] J. Woetzel et al., *Smart Cities: Digital Solutions for a More Livable Future*. Bengaluru, India: McKinsey & Company, 2018, p. 152.

[4] B. Barth, "Curb control." 2019. Accessed: Jul. 17, 2022. [Online]. Available: https://www.planning.org/planning/2019/jun/curbcontrol/

[5] P. Butrina, S. Le Vine, A. Henao, J. Sperling, and S. E. Young, "Municipal adaptation to changing curbside demands: Exploratory findings from semi-structured interviews with ten U.S. cities," *Transp. Policy*, vol. 92, pp. 1–7, Jun. 2020, doi: 10.1016/j.tranpol.2020.03.005.

[6] N. Mohammadi and J. E. Taylor, "Thinking fast and slow in disaster decision-making with smart city digital twins," *Nature Comput. Sci.*, vol. 1, no. 12, pp. 771–773, 2021, doi: 10.1038/s43588–021-00174–0.

[7] J. E. Taylor, G. Bennett, and N. Mohammadi, "Engineering smarter cities with smart city digital twins," *J. Manag. Eng.*, vol. 37, no. 6, Aug. 2021, Art. no. 2021001, doi: 10.1061/(ASCE)ME.1943–5479.0000974.

[8] Y. Wang, H. Hao, and C. Wang, "Preparing urban curbside for increasing mobility-on-demand using data-driven agent-based simulation: Case study of City of Gainesville, Florida," *J. Manag. Eng.*, vol. 38, no. 3, 2022, Art. no. 05022004, doi: 10.1061/(asce)me.1943–5479.0001021.

[9] C. Diehl, A. Ranjbari, and A. Goodchild, "Curbspace management challenges and opportunities from public and private sector perspectives," *Transport. Res. Rec.*, vol. 2675, no. 11, pp. 1413–1427, 2021. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/03611981211027156

[10] A. Campbell, A. Both, and Q. Sun, "Detecting and mapping traffic signs from Google street view images using deep learning and GIS," *Comput. Environ. Urban Syst.*, vol. 77, Sep. 2019, Art. no. 101350, doi: 10.1016/j.compenvurbsys.2019.101350.

[11] S. Lumnitz, T. Devisscher, J. R. Mayaud, V. Radic, N. C. Coops, and V. C. Griess, "Mapping trees along urban street networks with deep learning and street-level imagery," *ISPRS J. Photogram. Remote Sens.*, vol. 175, pp. 144–157, May 2021, doi: 10.1016/j.isprsjprs.2021.01.016.

[12] B. Kang, S. Lee, and S. Zou, "Developing sidewalk inventory data using street view images," *Sensors*, vol. 21, no. 9, p. 3300, 2021, doi: 10.3390/s21093300.

[13] Y. Zhou et al., "Street-view imagery guided street furniture inventory from mobile laser scanning point clouds," *ISPRS J. Photogram. Remote Sens.*, vol. 189, pp. 63–77, Jul. 2022, doi: 10.1016/J.ISPRSJPRS.2022.04.023.

[14] Y. Wu et al., "Detectron2." 2019. [Online]. Available: https://github.com/facebookresearch/detectron2

[15] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[16] M. Cordts et al., "The cityscapes dataset for semantic Urban scene understanding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 3213–3223, doi: 10.1109/CVPR.2016.350.

[17] S. Tuohy, D. O'Cualain, E. Jones, and M. Glavin, "Distance determination for an automobile environment using inverse perspective mapping in OpenCV," in *Proc. ISSC*, 2010, pp. 100–105, doi: 10.1049/cp.2010.0495.

[18] F. Yao and Y. Wang, "Towards resilient and smart cities: A real-time urban analytical and geo-visual system for social media streaming data," *Sustain. Cities Soc.*, vol. 63, Dec. 2020, Art. no. 102448 doi: 10.1016/j.scs.2020.102448.

[19] H. Hao and Y. Wang, "Assessing disaster impact in real time: Data-driven system integrating humans, hazards, and the built environment," *J. Comput. Civil Eng.*, vol. 35, no. 5, 2021, Art. no. 04021010, doi: 10.1061/(ASCE)CP.1943–5487.0000970.

[20] Y. Wang, S. Gao, N. Li, and S. Yu. "Crowdsourcing the perceived urban built environment via social media: The case of underutilized land," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101371, doi: 10.1016/j.aei.2021.101371.

[21] R. Wang, N. Li, and Y. Wang, "Does the returners and explorers dichotomy in urban human mobility depend on the observation duration? An empirical study in Guangzhou, China," *Sustain. Cities Soc.*, vol. 69, Jun. 2021, Art. no. 102862, doi: 10.1016/j.scs.2021.102862.