Self-calibration of Traffic Surveillance Camera Systems for Traffic Density Estimation on Urban Roads

Zijian Hu[1], Wei Ma[1]*, William H.K Lam[1], S.C. Wong[2], Andy H.F.Chow[3]
[1]Civil and Environmental Engineering, The Hong Kong Polytechnic University
[2]Department of Civil Engineering, The University of Hong Kong
[3]Systems Engineering and Engineering Management, City University of Hong Kong

November 30, 2020

Keywords— Automatic camera calibration; Vehicle detection and classification; Traffic density estimation; Hong Kong

1 Introduction

Traffic density estimation plays a key role in the Intelligent Transportation System (ITS), which is associated with with various smart mobility applications such as intersection signal control and traffic state prediction. In many cities, the traffic surveillance cameras have been installed to cover most arterial roads and expressways. The collected video or image data can be regarded as a potential data source for traffic state estimation since it provides real-time information and wide coverage of the transportation networks. Take Hong Kong as an example, there are in 368 surveillance cameras installed in the current traffic monitor system, while most of the data collected by the cameras are used for visual inspection and no automatic tools are developed to extract the real-time traffic information. In addition, another challenge is that the spatio-temporal frequency of the surveillance camera data is usually low. For the Hong Kong's camera system[1], the image resolution is 320 × 240 pixels and the updating frequency is two minutes (Wu and Lam 2010). Therefore, how to make full use of the data collected from traffic surveillance camera systems present a pressing challenges for the public agencies and the smart city development.

This project develops a framework for estimating traffic density with uncalibrated surveillance cameras under low frame rates and resolutions benefiting most regions in Hong Kong. Giving an interesting region, the density can be approximately considered as the quotient of the total length of vehicles in a lane unit and the length of each lane. For the former case, the number and type of vehicles are extracted to obtain the total length of vehicles in the lane. For the latter case, a camera calibration method is involved to get the real length of lane from the image.

Results show that the error of camera calibration is 7% and the accuracy of detection and classification are both around 84%. Based on the proposed framework, traffic density estimated from the low-resolution surveillance camera can replicate the estimation from high-resolution videos (regarded as ground truth) in the same region. Incorporating the historical data with our framework, the traffic capacity can be found from the rush hour, which provides fruitful information for the downstream traffic analysis.

*Corresponding author, +1(412)-708-3668, wei.w.ma@polyu.edu.hk

[1] https://www.td.gov.hk/en/transport_in_hong_kong/its/its_achievements/closed_circuit_television_images_on_the_internet/index.html

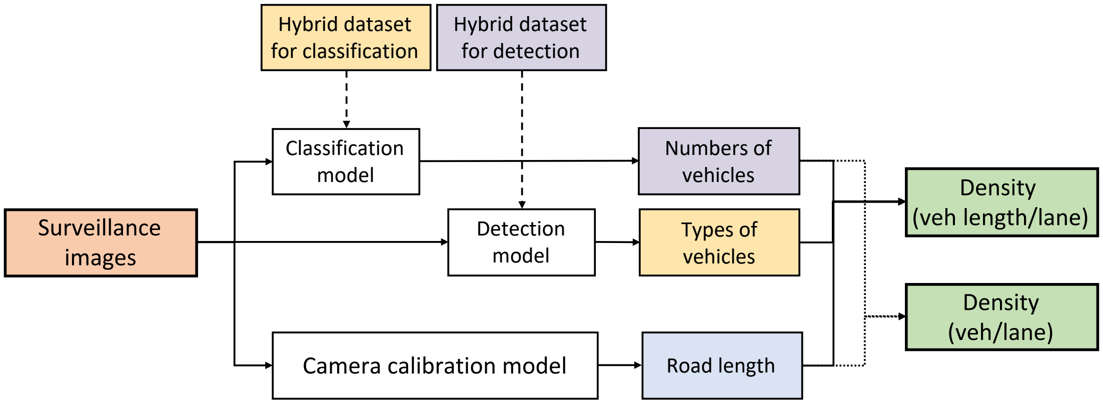


Figure 1: Framework of method for traffic density estimation in Hong Kong

2 Methodology

The framework of our project is shown in Fig.1. Overall, there are two major components in the framework: 1) camera calibration; 2) vehicle detection and classification. For the camera calibration, a two-stage semi-automatic method with multi-vehicles and multi-models is developed to estimate the real length of the interested lane from images. Multiple groups of candidate parameters are calculated in the first stage. In the second stage, parameters are filtered and optimized again considering the correlation between vehicles in the real space. Road length can be calculated after acquiring the optimal parameter for the camera. For vehicle detection and classification, deep-learning-based detection and classification frameworks (He et al. 2016; Redmon et al. 2016) can be utilized to extract the number and type of vehicles from images. Considering no labeled dataset for Hong Kong surveillance cameras has been published and two balances with illumination variance (day and night) and types of vehicles are difficult to mediate simultaneously, multi-source datasets are fused and re-sampled to enhance the diversity of training-set for detection and classification tasks separately.

2.1 Camera Calibration

Based on previous studies (Bartl and Herout 2019; Bhardwaj et al. 2018), we treat the camera calibration problem as a PnP (Perspective n Points) problem. The problem can be formulated in Equation 1.

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x&0&c_x\\0&f_y&c_y\\0&0&1\end{bmatrix}\begin{bmatrix}r_{11}&r_{12}&r_{13}&t_1\\r_{21}&r_{22}&r_{23}&t_2\\r_{31}&r_{32}&r_{33}&t_3\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{1}$$

where $[u, v]^T \in \mathbb{R}^2$ and $[X, Y, Z]^T \in \mathbb{R}^3$ are the pairwise key points in image plane and real space. The matrice on the right are intrinsic and extrinsic values that we need to be optimized. A two-stage camera calibration method is developed in this project. Due to the poor quality of images, we manually labeled key points of vehicles in images and measured the corresponding location in real space. In the first stage, the focal length is fixed as a known value, and the rotation and translation matrices are optimized through the EPnP algorithm (Lepetit et al. 2009). Involving multi-images and multi-models of vehicles, we optimized each combination and obtained multiple solutions as candidates. In the second stage, the global information of angles and distances of vehicles in real space are considered, and a hybrid loss function that incorporated both reprojection errors in images and real spaces is designed to optimize the intrinsic and extrinsic matrice. Since it is a non-convex optimization problem, the Covariance Matrix Adaptation (CMA) algorithm (Hansen and Ostermeier 1996) is leveraged to minimize the reprojection loss.

2.2 Vehicle detection and classification

Considering no labeled datasets have been published for Hong Kong surveillance cameras and evaluation data from high-resolution videos are also not annotated, it is laborious to label both datasets in the project. Multi-source datasets that incorporated re-identification, autonomous driving, and vehicle detection datasets are used in our project to derive general vehicle detection and classification models (Deng et al. 2009; Dong et al. 2015; Everingham et al.; Lin et al. 2014; Lou et al. 2019; Luo et al. 2018; Lyu et al. 2017; 2018; Wen et al. 2020; Yu et al. 2018; Zhang et al. 2017). For detection task, the ratio of images in a day and at night is the priority to balance, Resampling the above hybrid datasets, 52.3K images are selected for training and 40% images are shot at night, You Only Look Once (YOLO) v5 is selected for vehicle detection from the surveillance cameras (Redmon et al. 2016). For classification models, different types of vehicles are equilibrated for better performance consideration. The dataset consists of images with the bus, sedan, truck and van, and the numbers for each type are 17,039, 28,080, 23,217 and 11,224 respectively. Residual neural network (ResNet) is used to classify all types of vehicles (He et al. 2016).

3 Results

3.1 Evaluation for camera calibration

In this section, we conduct a case study on the camera on the Chatham Road South, underneath the footbridge of the Hong Kong Polytechnic University. Two-minute surveillance images are collected from June 22nd to September 25th, 2020. To evaluate the result of camera calibration. The road marks can be utilized since the real size is known. 14 road marks with 3 lanes are annotated and the distance for adjacent road marks are 6 meters . The average length for each interval is 5.72 meters and the standard deviation is 0.41 meters.

3.2 Traffic density analysis

For both detection and classification, the accuracy is around 84 percent on the hybrid training set. These two models can be used to extract the numbers and types from surveillance cameras. For evaluation, the vehicle counts estimated from high-resolution video data are compared. Fig.2 shows the traffic for 22 hours on September 24th, 2020 at Chatham Road South, Hong Kong. Fig.2A shows the detection and classification results after training 2 models. Boxing the same region from images (Fig.2B) and videos (Fig.2C), the variation of vehicles' number are close except for night (shown in Fig.2D), which demonstrates that surveillance cameras can be a potential device for observing traffic density. An interesting effect of traffic oscillation can be detected from both data from 11 AM to 12 AM where congestion appears and disappears alternatively in a short time. The detection accuracy for surveillance images at night is not very decent, and a potential reason may the patterns of vehicles at night in surveillance images are quite poor, which can be improved by annotating vehicles in the dataset in the future.

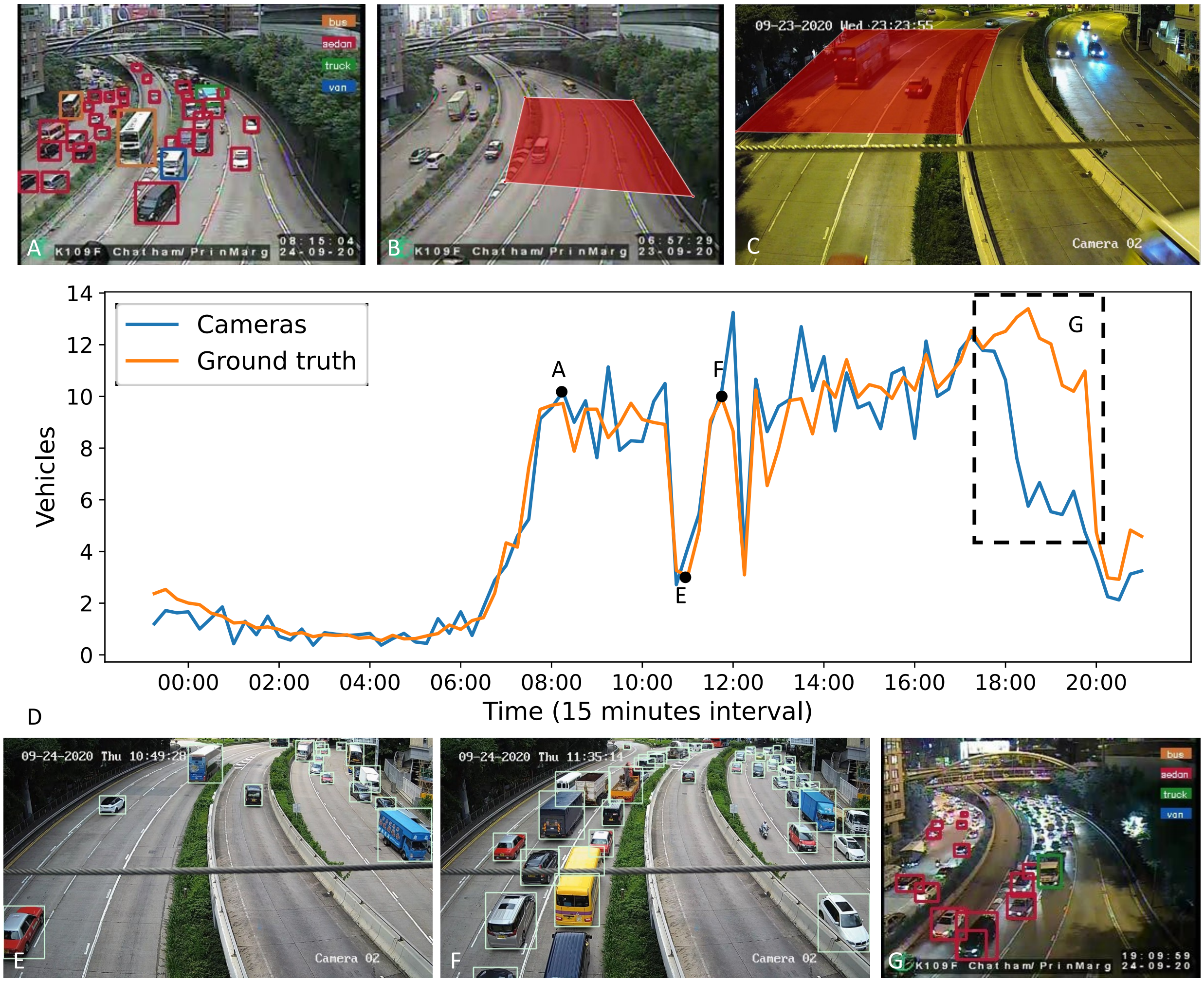
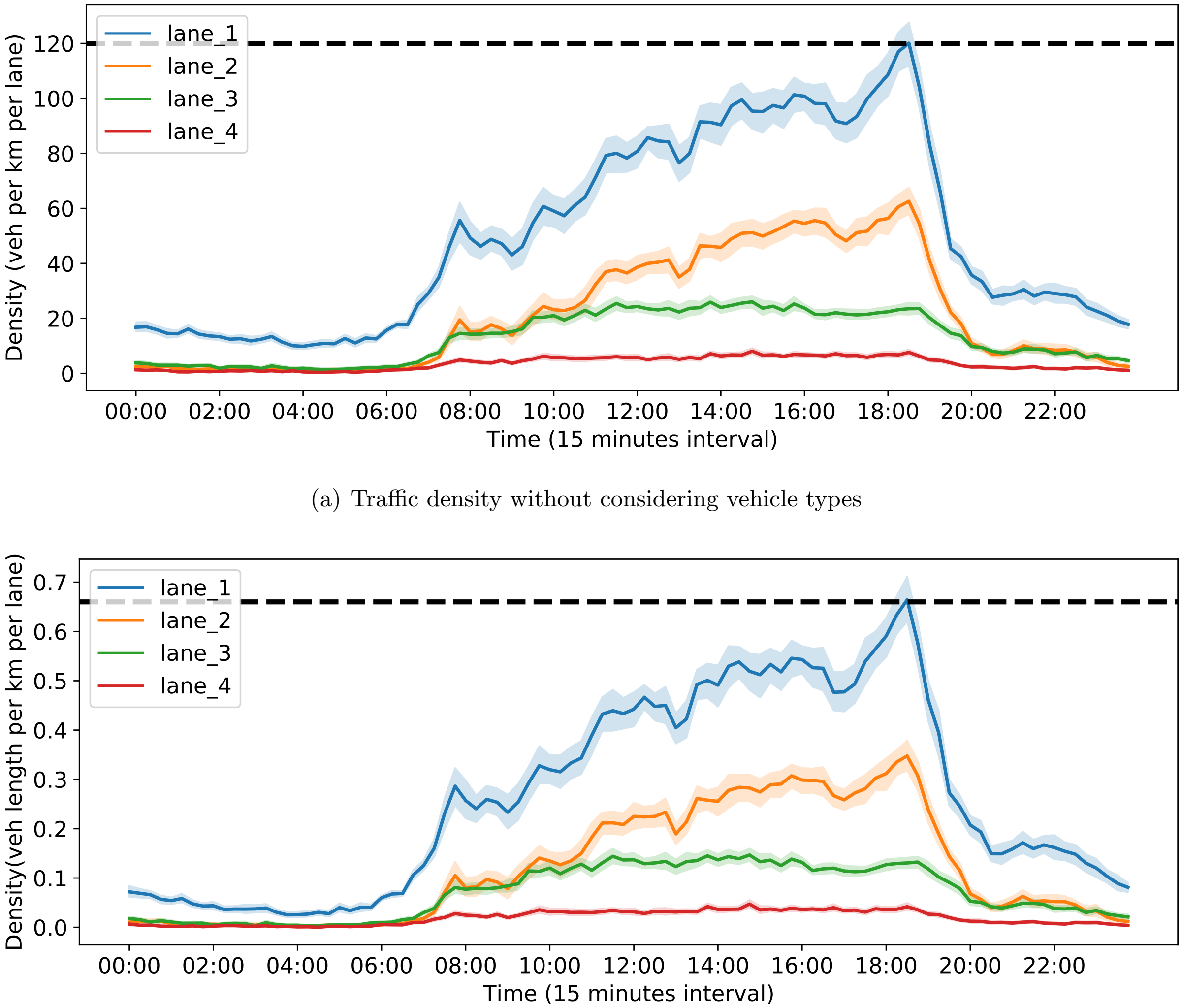


Figure 2: Vehicle counts at the footbridge of PolyU



(a) Traffic density without considering vehicle types

(b) Traffic density considering vehicle types

Figure 3: Traffic density estimation from historical data (June 22nd to Sep 22nd, 2020)

The traffic capacity of a certain road can be inferred from historical data. The data that we used is archived from June 22nd to September 22nd. Assuming that the traffic is always in congestion during the afternoon peak hours, The traffic capacity of the road is 120 vehicles per kilometers per lane (shown in Fig.3(a)). If we take the vehicle types into account and assign an average length for each type of vehicle, the traffic density is around 0.66 kilometers per kilometer per lane, which provides fruitful information for the local transportation department to regulate traffic conditions.

4 Conclusions

In this paper, we develop a traffic density estimation framework from low-resolution and low-frame-rate surveillance cameras. In the first step, a two-stage camera-calibration method is developed to calibrate the internal and external parameters from images. In the second step, deep learning-based detection and classification models are used to identify and classify the vehicles on the road. The results show that the estimation is consistent with the result from high-resolution videos and this framework could be used for large-scale and real-time traffic density estimation in a local traffic network.

Acknowledgments

The work described in this study was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU R5029-18) and a grant funded by the Hong Kong Polytechnic University (Project No. P0033933). The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein.

References

V. Bartl and A. Herout. Optinopt: Dual optimization for automatic camera calibration by multi-target observations. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, Sep. 2019. doi: 10.1109/AVSS.2019.8909905.

Romil Bhardwaj, Gopi Krishna Tummala, Ganesan Ramalingam, Ramachandran Ramjee, and Prasun Sinha. Autocalib: Automatic traffic camera calibration at scale. *ACM Trans. Sen. Netw.*, 14(3–4), November 2018. ISSN 1550-4859. doi: 10.1145/3199667. URL https://doi.org/10.1145/3199667.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Zhen Dong, Yuwei Wu, Mingtao Pei, and Yunde Jia. Vehicle type classification using a semisupervised convolutional neural network. *IEEE transactions on intelligent transportation systems*, 16 (4):2247–2256, 2015.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation*, pages 312–317, May 1996. doi: 10.1109/ICEC.1996.542381.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Epnp: An accurate o(n) solution to the pnp problem. *Int. J. Comput. Vision*, 81(2):155–166, February 2009. ISSN 0920-5691. doi: 10.1007/s11263-008-0152-6. URL https://doi.org/10.1007/s11263-008-0152-6.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3235–3243, 2019.

Zhiming Luo, Frederic Branchaud-Charron, Carl Lemaire, Janusz Konrad, Shaozi Li, Akshaya Mishra, Andrew Achkar, Justin Eichel, and Pierre-Marc Jodoin. Mio-tcd: A new benchmark dataset for vehicle classification and localization. *IEEE Transactions on Image Processing*, 27 (10):5129–5141, 2018.

Siwei Lyu, Ming-Ching Chang, Dawei Du, Longyin Wen, Honggang Qi, Yuezun Li, Yi Wei, Lipeng Ke, Tao Hu, Marco Del Coco, et al. Ua-detrac 2017: Report of avss2017 & iwt4s challenge on advanced traffic monitoring. In *Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on*, pages 1–7. IEEE, 2017.

Siwei Lyu, Ming-Ching Chang, Dawei Du, Wenbo Li, Yi Wei, Marco Del Coco, Pierluigi Carcagnì, Arne Schumann, Bharti Munjal, Doo-Hyun Choi, et al. Ua-detrac 2018: Report of avss2018 & iwt4s challenge on advanced traffic monitoring. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2018.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

Longyin Wen, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Computer Vision and Image Understanding*, 2020.

Zhongyi Wu and W.H.K. Lam. Using online CCTV image sequences for real-time traffic estimation. In *Proceeding of the 89th Annual Meeting of the Transportation Research Board (TRB)*, number 10-3846, Jan 2010.

Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2(5):6, 2018.

Shanghang Zhang, Guanhang Wu, Joao P Costeira, and Jose MF Moura. Understanding traffic density from large-scale web camera data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5898–5907, 2017.