Gathering Trajectories and Traffic Metrics from Street Level Video Recordings using current Deep Learning Algorithms: A Traffic Engineer's View Focused on Cycling

Wilkes, Gabriel[1]; Buck, H. Sebastian; Vortisch, Peter; Zeidler, Verena

Karlsruhe Institute of Technology, Institute for Transport Studies

Keywords: video data, traffic data collection, trajectories, microscopic flow simulation, bicycle traffic

Problem statement

Since bicycle traffic has been of rising relevance in traffic planning and engineering recently, there is increased need for microscopic traffic flow models that include bicycle traffic. However, existing microsimulation models that include cycling are of limited scope and often not sufficiently validated [1]. This might also be because little data sets are available to modelers. Collecting and using data for cycling is often more complicated than working with data about motorized vehicles, because lateral and longitudinal behavior of cyclists is more intertwined. Therefore, trajectories – vehicle position data continuously recorded over time – are specifically valuable to calibrate and validate bicycle traffic models.

The automated analysis of video recordings intuitively seems to be a suitable method to obtain trajectories. In contrast to methods e.g. based on GPS tracking, this allows for detection of all cyclists during the observation period. Traditional computer vision methods such as implemented in the software Traffic Intelligence [2] are used in bicycle modeling research to retrieve trajectories [3, 4]. However, using those methods requires extensive calibration and they are restricted to specific situations. Modern deep learning-based object detection algorithms have the ability to "understand" complex structures of an object category and therefore have the potential to deliver more robust results.

In the field of traffic engineering, first tests using such algorithms have been performed recently. For example, Jensen et al. [5] used deep learning techniques on thermal images to perform a traffic safety study. However, further details of the limitations and possibilities remain open questions. This includes the assessment of deriving traffic metrics, like traffic volume and individual and aggregated speeds, based on such trajectories.

Research objectives

The goal of our work is to test the actual applicability of image-based object detection for gathering trajectories and other traffic metrics of bicycle traffic from a traffic engineer's perspective. Questions in this context are: Is the accuracy satisfying using rather simple approaches? How time-consuming are all steps from recording videos towards the application of the algorithms? What is the cost-value-ratio compared to traditional methods? Which further development and research do we need? Could a traffic engineer apply those methods?

Methodological approach

We recorded videos of various urban street traffic situations using ordinary cameras in easily accessible mounting locations. Recordings of four situations were used to test the object detection of bicycles, and two of these to determine trajectories and further analysis. The four situations we used are two mixed traffic situations where cyclists are predominant, and two situations of cycling lanes or mixed cycling and pedestrian lanes at intersections.

[1] Corresponding author, gabriel.wilkes@kit.edu, +49 721 608-42253

We used the algorithms YOLO version 3 [6] in three different configurations and Mask R-CNN [7], which are among the best publicly available algorithms in current benchmarks and literature [8]. To keep the process simple, we did not perform manual annotation and training. Instead, we used publicly available models that are trained on the COCO dataset [9], which is based on everyday photos and includes a “bicycle” category. We developed a straightforward algorithm that groups the detections based on their time-location-similarity to build trajectories in image-based coordinates. The object detection was executed using a high performance GPU in a cloud computing service (Tesla P100).

For evaluating the results, at first we assessed the detection of bicycles in single video frames for each of the four situations. Secondly, we tested the trajectory building considering consecutive video frames. Third, we examined the accuracy in determining traffic volume by comparing the results to manual counts. At last, we assessed the determination of other important traffic metrics, including speed, and trajectories in world coordinates.

To understand the limitations and necessary computational effort, we tracked the computation time and tested different image resolutions and parameters for the trajectory building algorithm.

(Expected) results

The quality of the detection of bicycles in single video frames depends highly on the resolution, lighting conditions, distance and recording angles. For three out of the four algorithms, the rate of detected bicycles is around 90 to 100 %, if the objects have a distance of between 17 and 30 m to the camera. For larger distances, if the viewing angle gets sharper or backlight occurs, the detection rate lowers to around 40 to 70 % with the best algorithms.

The object detection randomly fails in some frames. Therefore, the trajectory building algorithm was built to be robust against missing detections. Allowing one or two frames of missing detections, the trajectory building works well; in most cases the trajectories can be created from the moment when a bicycle enters the image until it exits – as long as the object is constantly visible.

Consequently, the accuracy in calculating the traffic volume is high in these circumstances; in four consecutive samples of 15 minutes the difference to manual counted traffic volumes was 7 % at the most. However, the trajectory building is challenging in situations where occlusions occur frequently, e.g. due to mixed traffic situations with other passing vehicles or due to a high bicycle traffic density.

The available position information based on the object detection are the image coordinates of the “bounding box” or the outer “mask” of the object, depending on the algorithm. The calculation of real-world coordinates requires considerations about the geometry and angle between camera and objects in the specific situation. Depending on the specific requirements, this may produce sufficiently precise results. However, occlusions are especially problematic, and further development concerning the trajectory building is needed to handle these situations.

Several traffic metrics such as speed and waiting times can be approximated without a conversion into world coordinates through cross sections in the image coordinate system. Tests concerning these metrics produced mostly plausible and consistent results. In order to obtain a high quality in these metrics it is crucial to use well-constructed trajectories.

The computation time for the object detection depends primarily on the used algorithm and varies in the performed tests between 13 and 350 milliseconds per frame. A satisfying trade-off between performance and result quality is obtained with YOLO SPP, which required 93 milliseconds per frame in Full HD resolution. Based on the street level video recordings, a framerate of about 8 frames per second (fps) is sufficient to generate suitable results (e.g. to only determine traffic volumes), but 25 fps and more are required to determine speed and acceleration. The cost for evaluating one hour of observation at a framerate of 8 fps is at least 1.00 €, as GPU-including cloud computing instances cost around 1.47 € (1.60 USD) per hour. On higher frame rates and with other algorithms costs increase to several ten Euros per hour of observation.

In conclusion, under certain circumstances the hereby developed approach is a suitable method for gathering trajectories and traffic metrics of bicycle traffic with rather low efforts. At recording angles as used in this study, the automatic retrieval of trajectories works well in bicycle-only situations, as long as the traffic density is not too high. Therefore, at first a use in bicycle lanes and similar situations will be preferred. By using more sophisticated algorithms for the trajectory building and evaluation, the presented techniques might work well also in more complex traffic situations. Concerning the task of gathering traffic metrics, determining traffic volumes is especially easily achievable. For other traffic metrics, a precise view on the localization is needed. Concerning the determination of trajectories in world coordinates, in specific situations the positions can be approximated.

Literaturverzeichnis

[1] TWADDLE, Heather ; SCHENDZIELORZ, Tobias ; FAKLER, Oliver: *Bicycles in Urban Areas*. In: *Transportation Research Record: Journal of the Transportation Research Board* 2434 (2014), Nr. 1, S. 140–146

[2] JACKSON, Stewart ; MIRANDA-MORENO, Luis F. ; ST-AUBIN, Paul ; SAUNIER, Nicolas: *Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioral Analysis*. In: *Transportation Research Record: Journal of the Transportation Research Board* 2365 (2013), Nr. 1, S. 90–98

[3] TWADDLE, Heather ; GRIGOROPOULOS, Georgios: *Modeling the Speed, Acceleration, and Deceleration of Bicyclists for Microscopic Traffic Simulation*. In: *Transportation Research Record: Journal of the Transportation Research Board* 2587 (2016), S. 8–16

[4] AMINI, Sasan ; TWADDLE, Heather ; LEONHARDT, Axel: *Modelling of the tactical path selection of bicyclists at signalized intersections*. Presented at TRB Annual Meeting 2016. 2016

[5] JENSEN, Morten B. ; AHRNBOM, Martin ; KRUITHOF, Maarten ; ÅSTRÖM, Kalle ; NILSSON, Mikael ; ARDÖ, Håkan ; LAURESHYN, Aliaksei ; JOHNSSON, Carl ; MOESLUND, Thomas B.: *A Framework for Automated Analysis od Surrogate Measures of Safety from Video Using Deep Learning Techniques*. Presented at TRB 2019 Annual Meeting. 2019

[6] REDMON, Joseph ; FARHADI, Ali: *YOLOv3: An Incremental Improvement*. 09.04.2018

[7] WALEED ABDULLA: *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*. Github, 2017

[8] LIU, Li ; OUYANG, Wanli ; WANG, Xiaogang ; FIEGUTH, Paul ; CHEN, Jie ; LIU, Xinwang ; PIETIKÄINEN, Matti: *Deep Learning for Generic Object Detection: A Survey*. 06.09.2018

[9] LIN, Tsung-Yi ; MAIRE, Michael ; BELONGIE, Serge ; BOURDEV, Lubomir ; GIRSHICK, Ross ; HAYS, James ; PERONA, Pietro ; RAMANAN, Deva ; ZITNICK, C. Lawrence ; DOLLÁR, Piotr: *Microsoft COCO: Common Objects in Context*. 01.05.2014

[10] BUCH, N. ; VELASTIN, S. A. ; ORWELL, J.: *A Review of Computer Vision Techniques for the Analysis of Urban Traffic*. In: *IEEE Transactions on Intelligent Transportation Systems* 12 (2011), Nr. 3, S. 920–939