TACS: A Calibrated Highway Surveillance Dataset for Traffic Analysis

Audhav Durai[†], Siddhartha Rangavajjula[†], Sachin Satishkumar, Rusheel Nadipally
Thomas Jefferson High School for Science and Technology
Alexandria, United States
{2023adurai, 2023srangava, 2023ssatishk, 2023rnadipal}@tjhsst.edu

2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM) | 978-1-6654-5963-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIGMM55396.2022.00021

***Abstract*—Access to live streams of traffic surveillance cameras is becoming more common across the world. Despite this, many current traffic analysis algorithms do not utilize such cameras; rather, they rely on other sources of data. This can be attributed to the arduous process of extracting the necessary traffic information for such algorithms. The TACS (Traffic Analysis on Calibrated Streams) dataset aims to remove this barrier by annotating significant traffic information on realistic traffic surveillance video. We collect live video from the Virginia Department of Transportation and perform accurate camera calibration. After calibrating the camera, we obtain a top-down view of the video stream and extract the scene scale. From this, we calculate the real world position and speed of each vehicle. On top of such information, we calculate vehicle lane positions and two types of road congestion metrics. We collect between 10,000 and 15,000 frames from each of its 10 traffic video sources. Thus, the TACS dataset contains a wide variety of data and provides researchers with rich annotations for traffic-related algorithms.**

***Index Terms*—traffic analysis, traffic dataset, computer vision**

I. Introduction

With the abundance of live traffic streams, constant road surveillance using computational methods have great potential. Computational traffic monitoring have many applications, including speed estimation, reckless driving detection, and accident prediction. For these applications, proper traffic datasets are necessary for both development and evaluation of algorithms. Despite this, current traffic datasets tend to focus on only vehicle detection and tracking, leaving out other important traffic information. Furthermore, these data sources rarely collect traffic video that is representative of a typical CCTV source: often recording video streams at a unreasonably high quality or from non-surveillance cameras (e.g., drones, dashboard cameras). These dataset limitations serve as major barriers for research in this field. Thus, in this paper, we propose the Traffic Analysis on Calibrated Streams dataset (TACS) for traffic monitoring algorithms. The TACS dataset contains realistic surveillance video on roads, specifically straight highways, and annotations for significant traffic related information. The TACS dataset contains the following novel additions:

- Camera calibration on video streams and creation of a birds-eye-view video stream
- Calculation of the real world scene scale and vehicle speeds that is robust to video processing and bandwith issues
- Annotation of lane position and road congestion

[†]Both authors contributed equally to this work

II. Materials & Methods

To provide original, realistic traffic surveillance video, we collect live streams from the Virginia Department of Transportation 511 CCTV database [2]. A YOLOv4 model combined with a DeepSORT re-identification algorithm, which is accurate in traffic surveillance video [1] performs vehicle detection and tracking. Using the process shown in [1], we perform an accurate, two-vanishing-point-based camera calibration algorithm and extract a birds-eye-view perspective using the process shown in Figure 1. The frames of this birds-eye-view perspective are saved as a video to be used for further processing. We also save the transformation matrix to convert vehicle positions in the original video stream to positions in the top-down view.

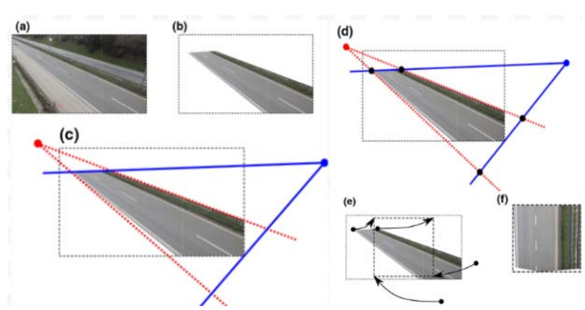

Fig. 1: Construction of the perspective transformation using the first and second vanishing points [4]

We obtain the scene scale by comparing the vehicle speeds in the top down view video to speed limits on the recorded roads. In order to prevent video processing issues and dropped frames from skewing the

978-1-6654-5963-1/22/$31.00 ©2022 IEEE
DOI 10.1109/BigMM55396.2022.00021

scene scale, we disregard vehicles that are detected in less than 50 frames in the calculation. The extraction of the birds-eye-view transformation matrix and scene scale is extremely useful for traffic analysis because it makes it extremely easy to obtain real world vehicle position. This is because the birds-eye-view perspective transformation converts the image pixel coordinates to real world coordinates on the road plane.

This birds-eye-view perspective also is used to calculate lane positions. The horizontal positions of vehicles are passed into a mean shift algorithm to extract the number of clusters in the data, which represents the number of lane lines in each video. Then, the horizontal positions and the number of clusters are passed into a K-Means clustering algorithm to identify the lane positions of each vehicle. Lastly, we record two different types of road congestion information. For each vehicle, we record the distance to the closest vehicle in the same lane, adjusted using the scene scale. Additionally, we save the number of vehicles in each frame to get the total congestion of a road at a specific point in time. For traffic planning, these congestion metrics are vital. Additionally, lane change frequency, vehicle speed, and car congestion information can be used for reckless driving and accident prediction [3].

III. Results & Discussion

We collect video streams from 10 different highways and record between 10,000 and 15,000 frames for each source. The video streams are recorded at 15 frames per second and a resolution of 320 x 240 pixels [2]. Such a quality is similar to that of most CCTV cameras. We limit the amount of manual interference in the collection of the data to provide a realistic dataset; only the accuracy of the camera calibration algorithm was manually verified by checking the correctness of the birds-eye-view transformations. Figure 2 shows frames from both the original and birds-eye-view video streams in one of the recorded roads.

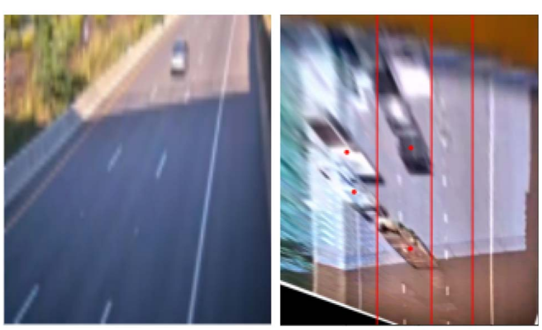

Fig. 2: Frames from the original video (left) and the transformed, birds-eye-view video stream with lane lines (right)

It can be seen that the red lines, which are the detected lane lines, are slightly to the left of the actual lane lines. This is caused by a slight error in the birds-eye-view transformation, which is almost unavoidable when recordings are taken at any orientation. However, a similar horizontal error is also present in the calculation of the positions of the vehicles in the birds-eye-view video. Thus, we observe that the lane positions of the vehicles are still calculated correctly, which is sufficient for the purpose of this dataset and traffic monitoring algorithms.

IV. Conclusion

This study presents the TACS (Traffic Analysis on Calibrated Streams) for the development traffic monitoring algorithms. First, we use a YOLOv4-DeepSORT model to detect and track vehicles in the original video stream. For each stream, we perform camera calibration [1] and create a new stream from a birds-eye-view perspective. This new perspective and the scene scale are used to extract real word vehicle positions and speeds. The positions of the vehicles are used to identify the lane lines and classify lane positions of vehicles. Lastly, we annotate road congestion in two types of ways: total road congestion, the number of vehicles on the road in each frame, and vehicle-specific congestion, the distance from each vehicle to the closest vehicle in the same lane.

As the field of computational traffic monitoring expands, algorithms must be developed and tested on realistic data sources. The TACS dataset provides such data by collecting streams from cameras that have a video quality representative of the majority of CCTV cameras and using mostly automatic data extraction methodologies. Additionally, the TACS dataset annotates a significant amount of traffic information so that researchers can focus solely on the application of their traffic analysis algorithms, rather than forcing them to do arduous preprocessing. Ultimately, this dataset further facilitates traffic research by removing major barriers to the development of traffic analysis programs.

References

[1] Durai, A., Vempati, S., Bindschaedler, L. (2022). Accurate Automatic Camera Calibration on Low-Quality CCTV Traffic Video Streams. In 2022 IEEE International Conference on Computers and Automation (CompAuto). IEEE.

[2] Virginia Department of Transportation (2012). VDOT's New 511 Traffic Information Program. https://www.virginiadot.org/news/resources/ 511_video/VDOT_-_Media_Day_Info_Session.pdf

[3] Heo, T., Nam, W., Paek, J., & Ko, J. (2020, December). Autonomous reckless driving detection using deep learning on embedded GPUs. In 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS) (pp. 464-472). IEEE.

[4] Kocur, V., & Ftáčnik, M. (2020). Detection of 3D bounding boxes of vehicles using perspective transformation for accurate speed measurement. Machine Vision and Applications, 31(7), 1-15.