A dataset for audio-video based vehicle speed estimation

1st Slobodan Djukanović
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
slobdj@ucg.ac.me

2nd Nikola Bulatović
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
nbulatovic@ucg.ac.me

3rd Ivana Čavor
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
ivana.ca@ucg.ac.me

***Abstract*—Accurate speed estimation of road vehicles is important for several reasons. One is speed limit enforcement, which represents a crucial tool in decreasing traffic accidents and fatalities. Compared with other research areas and domains, the number of available datasets for vehicle speed estimation is still very limited. We present a dataset of on-road audio-video recordings of single vehicles passing by a camera at known speeds maintained stable by the on-board cruise control. The dataset contains thirteen vehicles, selected to be as diverse as possible in terms of manufacturer, production year, engine type, power and transmission, resulting in a total of 400 annotated audio-video recordings. The dataset is fully available and intended as a public benchmark to facilitate research in audio-video vehicle speed estimation. In addition to the dataset, we propose a cross-validation strategy which can be used in a machine learning model for vehicle speed estimation. Two approaches to training-validation split of the dataset are proposed.**

***Index Terms*—Audio, dataset, traffic monitoring, vehicle speed estimation, video**

I. Introduction

The demands of improving traffic safety, mobility and efficiency, reducing air pollution and mitigating the impact of traffic problems (e.g. the impact of congestion on the economy) have led to the introduction of Intelligent Transport System (ITS). Therefore, an ITS applies a combination of leading-edge information and communication technologies to carry out sensing, analysis, control, and communication tasks.

One of the key ITS features is the ability of accurate speed estimation of the road vehicles, which is important for various reasons. First, since speeding increases both the risk of traffic accidents and the severity of consequences [1], speed limit enforcement is considered a crucial tool in decreasing traffic accidents and fatalities. For example, in the vicinity of speed cameras, the number of speeding vehicles and crashes is reduced up to 35% and 25%, respectively. In addition, the rate of accident reduction is directly proportional to the intensity or level of enforcement [2]. Consequently, the number of speed cameras installed worldwide has been constantly growing. Second important reason is that the knowledge of the traffic speed (average speed of all vehicles in multiple road segments) can be used in adaptive traffic signal control, real-time traffic aware navigation, detection of traffic jams and accidents.

Speed estimation of the road vehicles encompasses multiple tasks, including synchronized data recording, representation, detection and tracking, as well as distance and speed estimation. The data themselves are collected by sensors of different nature, such as radar, laser or cameras. Depending on the application, the accuracy and robustness of speed estimation can be at different levels. Regarding speed enforcement, the estimation accuracy has to be very high, since speed offenders can be fined, lose their drivers' license or even get imprisoned as a consequence. Therefore, sensors commonly used for speed measurement in sensitive applications are high-precision range sensors, such as radar (which uses the Doppler effect), LiDAR (based on the time of flight), or intrusive sensors embedded in pairs under the road surface (piezoelectric sensors and induction loops). These sensors provide very accurate speed estimations, although high price and the costs of installation and maintenance prevent their widespread use.

The focus of this paper are datasets for vehicle speed estimation. We consider video and audio data, i.e. data obtained from video cameras. The available audio-video datasets are described in Section II. Section III describes a dataset of 400 audio-video recordings collected for the purpose of vehicle speed estimation. Section IV concludes the paper.

II. Available datasets

A. Video datasets

In traffic monitoring, vision offers several important advantages with respect to other modalities. Cameras provide rich vehicle-based information such as visual features of vehicles, their geometry and path. A single camera can be utilized to detect and classify vehicles in multiple lanes. In addition, installation and maintenance of cameras in roadways is significantly less expensive and less disruptive than in intrusive systems. Even though vehicle speed estimation is a popular research topic, the number of available datasets is still very limited compared to other research areas.

Recently, the most widely used dataset for traffic and vehicle speed detection is the AI City Challenge [3]. It is part of an annual challenge originally proposed by NVIDIA in 2017, with the aim of pushing the boundaries of research and development in intelligent video analysis for various use cases in smart cities. Each year the number of videos and samples provided varies. Specifically, the challenge in 2018 focused on problems such as estimating traffic flow characteristics,

including traffic speed [4]. For the challenge purpose, 142 videos of different resolutions and length, with no ground truth of speed, were provided.

The BrnoCompSpeed dataset [5] contains 21 full-HD (1920×1080 pixels) videos, each with length of approximately 1 hour. Vehicles in the videos (20865 in total) were annotated using LiDAR and verified with several reference GPS tracks.

Another useful dataset has been collected by the Federal University of Technology of Paraná [6]. It contains 20 full-HD videos of total time of 291 minutes. The videos, containing 8849 vehicles in total, were divided into five sets depending on the weather and recording conditions. The ground truth speeds were obtained using a high precision speed meter based on an inductive loop detector. The bulk of the recorded speeds in the dataset is within range $40 - 60$ km/h.

A thorough analysis of the vision-based vehicle speed estimation topic is given in [7].

B. Audio datasets

Vision-based vehicle speed estimation suffers from reduced performance due to partial occlusion, shadows and illumination variation. What is more, video processing can be computationally expensive and time consuming. In terms of traffic monitoring, acoustic sensors offer many advantages over cameras. The list includes lower price, lower energy consumption, significantly reduced storage space, lower installation and maintenance costs [8].

Similarly to vision-based vehicle speed estimation, research progress in acoustic vehicle speed estimation is limited by lack of rich datasets, preferably annotated ones. Datasets used in experiments in the studies [9]–[15] are usually very small. For example, Cevher et al. use ten audio recordings, corresponding to nine different vehicles [9]. In [10] and [11], seven recordings were used, corresponding to three cars, a bus and a motorbike. Two different cars, four different speeds per car and two recordings per speed, were used in [12]. Research [13] used the sound of a car driven on a parking lot, with no car manufacturer, speed or the number of laps specified. Only one recording, with the length of 240 seconds, containing 22 and 2 motorbikes, was used in [14]. Göksu [15] used the sound of one vehicle, with speeds ranging from 30 km/h to 80 km/h.

III. VS13 DATASET

This paper presents a dataset of on-road audio-video recordings of vehicles passing by the camera at constant speeds. Each recording in the dataset contains a single drive of a single vehicle. Thirteen different vehicles were used, with a total of 400 annotated audio-video recordings. The dataset, referred to as VS13, is fully available, intended as a public benchmark to facilitate research on audio-video vehicle speed estimation.

VS13 has been compiled following these requirements [16]:

1) each recording contains a single drive of a single vehicle,
2) recordings are made in an urban environment,
3) recordings are real field ones,
4) vehicles are as diverse as possible in terms of manufacturer, production year, engine type (petrol or diesel), power and transmission (manual or automatic),
5) vehicles are equipped with cruise control (speed control), so that speed can be maintained stable while the vehicle passes by the camera.

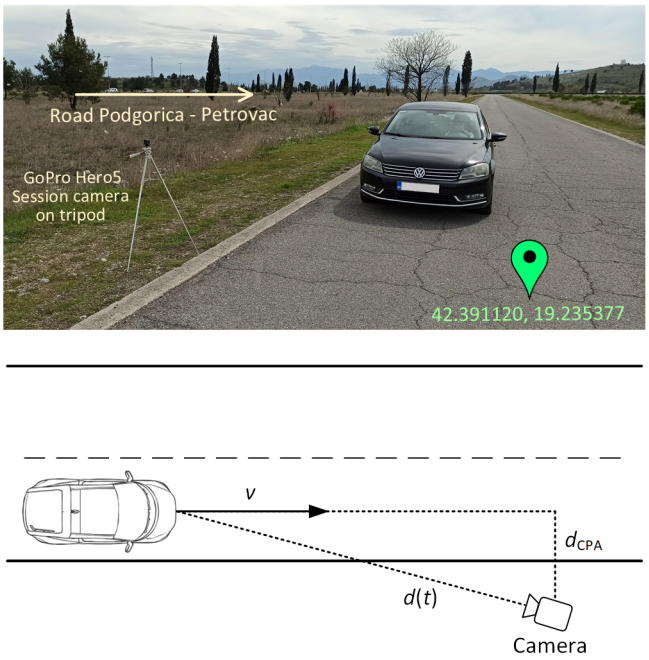

Fig. 1. *Top:* Screenshot of the recording setup. GoPro Hero5 Session camera mounted on a tripod, distance of ≈ 0.5 m from the road and at a height of ≈ 1.2 m. *Bottom:* Vehicle moving at a constant speed v. $d(t)$ is the distance between a vehicle and the camera at time instant t, whereas d_{CPA} is the distance at the closest point of approach (CPA).

The first requirement implies that we address estimating the speed of individual vehicles rather than measuring the average speed of all vehicles along a road segment, i.e. the traffic speed. Requirements 2) and 3) are important for the context of acoustic vehicle speed estimation. More precisely, the two imply that the pass-by sound of vehicles (prominent sound source) can be corrupted by the sound of other nearby vehicles and natural sounds (e.g. wind, bird chirps, crickets).

The dataset (recordings with annotations) is available for download at http://slobodan.ucg.ac.me/science/vs13/. Audio files have been extracted from the recordings and provided separately for download, to facilitate research in acoustic vehicle speed estimation.

A. Dataset collection

The dataset was recorded on a local road, 622 m long, located 90 m away from the main road connecting the cities of Podgorica and Petrovac in Montenegro (see arrow in Fig. 1 (top)). The reasons for selecting this road are as follows [16]:

- it is long enough so that stable speeds can be achieved prior to the pass-by instant,
- it is isolated enough to allow measurements without too many disturbances,
- it is close to other roads so that requirements 2) and 3) are met.

TABLE I
VS13 VEHICLES AND SPEEDS

Vehicle (Short name)	Engine type	Power (kW)	Transmission	Prod. year	Record. sessions	Speeds (km/h)
Citroen C4 Picasso 1.6 HDI (CitroenC4Picasso)	Diesel	88	Manual	2015	1	35, 38, 41, 44, 48, 51, 54, 57, 59, 63, 65, 68, 72, 74, 78, 80, 83, 85, 87, 92, 94, 96, 101
Kia Sportage 1.6 GDI (KiaSportage)	Petrol	97	Manual	2021	1	31, 33, 35, 38, 41, 44, 46, 48, 51, 53, 55, 58, 61, 63, 65, 68, 69, 72, 74, 77, 78, 80, 83, 85, 86, 89, 91, 93, 96, 98, 100, 103, 105
Mazda 3 Skyactive (Mazda3)	Petrol	74	Manual	2015	1	30, 33, 35, 38, 40, 43, 45, 47, 50, 52, 55, 57, 60, 62, 64, 67, 70, 72, 75, 79, 81, 84, 86, 88, 90, 92, 94, 96, 99, 101, 103, 105
Mercedes AMG 550 (MercedesAMG550)	Petrol	350	Automatic	2006	3	30, 33, 35, 38, 40, 42, 45, 47, 50, 52, 55, 58, 60, 62, 65, 67, 70, 73, 75, 78, 80, 82, 85, 87, 90, 93, 95, 98, 100, 105
Mercedes GLA 200D (MercedesGLA)	Diesel	100	Automatic	2017	1	30, 33, 36, 39, 41, 42, 45, 47, 48, 49, 52, 54, 55, 59, 61, 63, 65, 68, 70, 72, 75, 78, 81, 83, 85, 88, 90, 92, 93, 96, 100, 101, 103, 104
Nissan Qashqai 1.5 DCI (NissanQashqai)	Diesel	81	Manual	2018	1	35, 38, 40, 42, 45, 48, 50, 53, 55, 58, 60, 61, 64, 65, 68, 70, 73, 75, 78, 80, 82, 85, 88, 90, 93, 94, 96, 98, 102
Opel Insignia 2.0 CDTI (OpelInsignia)	Diesel	96	Automatic	2010	1	31, 35, 38, 41, 44, 47, 50, 53, 55, 58, 61, 64, 66, 68, 70, 72, 73, 76, 78, 80, 83, 86, 89, 91, 94, 97, 100
Peugeot 208 1.4 HDI (Peugeot208)	Diesel	50	Automatic	2014	1	30, 32, 34, 37, 40, 43, 45, 47, 50, 51, 54, 57, 60, 62, 64, 67, 68, 71, 73, 76, 77, 79, 82, 84, 87, 90, 92, 95, 96
Peugeot 3008 1.6 HDI (Peugeot3008)	Diesel	84	Automatic	2013	2	40, 43, 45, 47, 50, 52, 54, 55, 56, 58, 60, 61, 63, 65, 67, 68, 70, 72, 74, 75, 78, 80, 83, 85, 87, 89, 90, 92, 95, 97, 100
Peugeot 307 2.0 HDI (Peugeot307)	Diesel	100	Manual	2007	1	30, 33, 35, 38, 40, 43, 45, 47, 48, 50, 53, 56, 59, 60, 63, 66, 69, 72, 73, 76, 79, 82, 85, 88, 91, 94, 97, 101, 103
Renault Captur 1.5 DCI (RenaultCaptur)	Diesel	66	Automatic	2015	1	30, 33, 36, 38, 40, 41, 44, 46, 47, 48, 50, 52, 56, 58, 60, 63, 66, 68, 70, 72, 76, 78, 80, 83, 86, 88, 90, 92, 94, 97, 98, 100, 102
Renault Scenic 1.9 DCI (RenaultScenic)	Diesel	96	Manual	2010	2	30, 35, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 57, 60, 62, 64, 66, 68, 70, 71, 72, 74, 75, 77, 80, 82, 84, 86, 87, 90, 91, 94, 95, 98, 101
VW Passat B7 1.6 TDI (VWPassat)	Diesel	77	Manual	2011	2	30, 35, 39, 40, 42, 45, 47, 49, 50, 52, 54, 55, 57, 60, 61, 64, 65, 67, 70, 71, 72, 73, 75, 78, 80, 81, 82, 85, 88, 90, 91, 94, 96, 98, 100

For dataset recording, we used a GoPro Hero5 Session camera. It was installed by the road, mounted on a tripod, approximately 0.5 m from the road and at a height of approximately 1.2 m. Screenshot of the recording setup is presented in Fig. 1 (top). The camera position varied with respect to the road, that is, it was installed on both sides of the road and at different angles[1]. The recording sessions (one session per day) took place from December 2019 to February 2022. Thirteen vehicles were used, as listed in Table I. The number of recording sessions per vehicle is given in the sixth column of Table I.

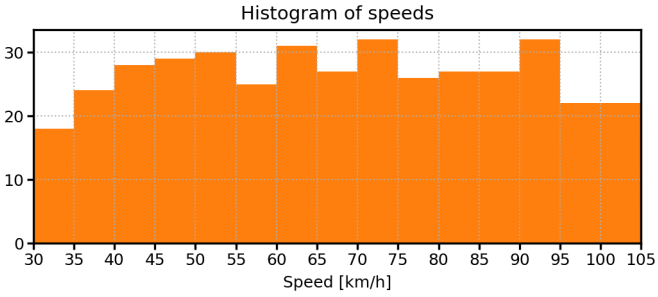

Fig. 2. Histogram of VS13 speeds (15 equal-width bins).

B. Dataset speeds

Speeds in the dataset range from 30 to 105 km/h, with the exact values given in the last column in Table I. For lower speeds, under 30 km/h, the cruise control cannot operate with the selected vehicles (for Peugeot 3008, even below 40 km/h). For higher speeds, above 105 km/h, we couldn't carry out stable and secure measurements in the selected road. Speed step varies between 1 to 3 km/h and all speeds from 30 to 105 km/h are included in VS13. Histogram of speeds is given in Fig. 2. The reported speeds are stable at least 3 seconds before and after the pass by. Outside that 6-second interval, minor speed variations are possible.

C. Dataset preprocessing: Video and audio

The original recordings were cut into 10-second video files (MP4 format, full HD, 30 fps) so that the pass-by instants of vehicles are around the middle of the file. For that purpose, we used the Format Factory application.

VS13 contains 13 folders, with 400 video files in total. Each folder contains 10-second video files and text annotations that correspond to one vehicle.

For the purpose of acoustic vehicle speed estimation, we extracted audio files (44100 Hz sampling rate, WAV format, 32-bit float PCM) from the corresponding video files using the Audacity application. The extracted audio files and the corresponding annotations are also available for download.

D. Dataset annotations

Video and audio files in VS13 are accompanied by annotation text files which contain two pieces of data each: the vehicle's speed and the pass-by-camera instant (with two-decimal precision). Relative time from the beginning of the file is given, measured in seconds. The pass-by-camera instant

[1]Sample video files of each vehicle can be seen at http://slobodan.ucg.ac.me/science/vs13/.

corresponds to the timestamp of a video frame that contains the vehicle starting to exit the camera view. That instant approximately corresponds to the closest point of approach (CPA), as depicted in Fig. 1 (bottom).

E. Naming convention

While naming the VS13 files, the following convention has been respected:

`shortVehicleName_vehicleSpeed`

where `shortVehicleName` and `vehicleSpeed` represent the short vehicle name (see the first column in Table I) and vehicle speed, respectively. For example, `MercedesGLA_68.mp4`, `MercedesGLA_68.wav` and `MercedesGLA_68.txt` represent the names of video, audio and annotation files, respectively, of Mercedes GLA 200D driven at 68 km/h.

F. Cross-validation strategy

In addition to the VS13 dataset, we propose a cross-validation (CV) strategy which can be used in a machine learning model for vehicle speed estimation. Since there are 13 vehicles in the dataset, a natural solution would be to use 13-fold CV. One (out of 13) CV round implies that one fold (vehicle) is used for testing, whereas the remaining twelve folds are used for training and validating the model. For training-validation split, we propose two approaches. In the first approach, for each vehicle, the files are divided into the training and validation ones according to the 80%–20% rule. To that end, we carry out the following procedure:

1) sort the speeds into ascending order,
2) divide the sorted speeds into batches of five speeds,
3) randomly select one speed in each batch to be used for validation, the other ones for training.

This approach ensures that low-, medium- and high-speed files are used in both training and validation. Each VS13 folder contains a file `Train_valid_split.txt` with labels *train* or *valid* associated with each file. Acoustic vehicle speed estimation methods [16]–[18] use this strategy.

The second approach is to split the folds, not speeds within the folds, into the training and validation ones. Nine folds can be used for training and the remaining three for validation.

In order to evaluate the model as precisely as possible, the proposed two approaches can be implemented as iterated k-fold CV with shuffling, which consists of applying k-fold CV multiple times, shuffling the data every time before splitting it k ways. The final score is the average of the scores obtained at each run of k-fold CV. Depending on the model architecture, this approach can require very high computational power.

IV. Conclusion

In this paper, we presented a dataset of on-road audio-video recordings of vehicles passing by camera at known constant speeds. The dataset, referred to as VS13, contains 400 annotated audio-video recordings of vehicles selected to be as diverse as possible in terms of manufacturer, production year, engine type, power and transmission. The dataset is publicly available, freely accessible, and intended as public benchmark to facilitate research in vehicle speed estimation. In addition to the dataset, we proposed a cross-validation strategy which can be used in a machine learning model for vehicle speed estimation, as well as two approaches for training-validation dataset split.

References

[1] W. H. Organization *et al.*, "Speed management: a road safety manual for decision-makers and practitioners," 2008.

[2] R. Elvik, "Developing an accident modification function for speed enforcement," *Safety Science*, vol. 49, no. 6, pp. 920–925, 2011.

[3] M. Naphade, S. Wang, D. C. Anastasiu, Z. Tang, M.-C. Chang, X. Yang, Y. Yao, L. Zheng, P. Chakraborty, C. E. Lopez, A. Sharma, Q. Feng, V. Ablavsky, and S. Sclaroff, "The 5th ai city challenge," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021.

[4] M. Naphade, M.-C. Chang, A. Sharma, D. C. Anastasiu, V. Jagarlamudi, P. Chakraborty, T. Huang, S. Wang, M.-Y. Liu, R. Chellappa *et al.*, "The 2018 NVIDIA AI city challenge," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018) workshops*, 2018, pp. 53–60.

[5] J. Sochor, R. Juránek, J. Špaňhel, L. Maršík, A. Širokỳ, A. Herout, and P. Zemčík, "Comprehensive data set for automatic single camera visual speed measurement," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1633–1643, 2018.

[6] D. C. Luvizon, B. T. Nassu, and R. Minetto, "A video-based system for vehicle speed measurement in urban roadways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1393–1404, 2016.

[7] D. Fernández Llorca, A. Hernández Martínez, and I. García Daza, "Vision-based vehicle speed estimation: A survey," *IET Intelligent Transport Systems*, vol. 15, no. 8, pp. 987–1005, 2021.

[8] M. Won, "Intelligent traffic monitoring systems for vehicle classification: A survey," *IEEE Access*, vol. 8, pp. 73 340–73 358, 2020.

[9] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 30–47, 2008.

[10] S. Barnwal, R. Barnwal, R. Hegde, R. Singh, and B. Raj, "Doppler based speed estimation of vehicles using passive sensor," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2013, pp. 1–4.

[11] R. López-Valcarce, C. Mosquera, and F. Pérez-González, "Estimation of road vehicle speed using two omnidirectional microphones: A maximum likelihood approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 8, pp. 1–19, 2004.

[12] J. Giraldo-Guzmán, A. G. Marrugo, and S. H. Contreras-Ortiz, "Vehicle speed estimation using audio features and neural networks," in *2016 IEEE ANDESCON*. IEEE, 2016, pp. 1–4.

[13] H. V. Koops and F. Franchetti, "An ensemble technique for estimating vehicle speed and gear position from acoustic data," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, pp. 422–426.

[14] P. Marmaroli, J.-M. Odobez, X. Falourd, and H. Lissek, "Pass-by noise acoustic sensing for estimating speed and wheelbase length of two-axle vehicles." in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 040030.

[15] H. Göksu, "Vehicle speed measurement by on-board acoustic signal processing," *Measurement and Control*, vol. 51, no. 5-6, pp. 138–149, 2018.

[16] S. Djukanović, J. Matas, and T. Virtanen, "Acoustic vehicle speed estimation from single sensor measurements," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23 317–23 324, 2021.

[17] N. Bulatović and S. Djukanović, "An approach to improving sound-based vehicle speed estimation," in *2022 Zooming innovation in consumer technologies conference (ZINC 2022)*. IEEE, 2022.

[18] ——, "Mel-spectrogram features for acoustic vehicle detection and speed estimation," in *2022 26th International Conference on Information Technology (IT)*. IEEE, 2022, pp. 1–4.