https://doi.org/10.2352/EI.2025.37.15.AVM-106

An Annotated Database for Pedestrian Temporal Action Recognition

Itsaso Rodríguez-Moreno[1], Brian Deegan[2], Dara Molloy[3], José María Martínez-Otzeta[1], Martin Glavin[2], Edward Jones[2], Basilio Sierra[1]

[1]Department of Computer Science and Artificial Intelligence, University of the Basque Country, 20018 Donostia-San Sebastián, Spain
[2]School of Engineering, University of Galway, University Road, Galway, Ireland
[3]Valeo, Tuam, Co. Galway H54 Y276, Ireland

Abstract

In this paper, we present a database consisting of the annotations of videos showing a number of people performing several actions in a parking lot. The chosen actions represent situations in which the pedestrian could be distracted and not fully aware of her surroundings. Those are "looking behind", "on a call", and "texting", with another one labeled as "no_action" when none of the previous actions is performed by the person. In addition to actions, also the speed of the person is labeled. There are three possible values for such speed: "standing", "walking" and "running". Bounding boxes of people present in each frame are also provided, along with a unique identifier for each person. The main goal is to provide the research community with examples of actions that can be of interest for surveillance or safe autonomous driving. The addition of the speed of the person when performing the action can also be of interest, as it can be treated as a more dangerous behavior "running" than "waking", when "on a call" or "looking behind", for example, providing the researchers with richer information.

Introduction

The field of autonomous driving has made remarkable strides in recent years, propelled by the development of sophisticated datasets that capture diverse and realistic driving scenarios. These datasets form the backbone of training and evaluating machine learning models, enabling robust perception, prediction, and control mechanisms. However, despite significant progress, understanding and predicting pedestrian behavior remains a critical challenge due to its inherent unpredictability and the varying contexts in which interactions occur. Addressing this gap is vital, as pedestrians often engage in actions that compromise their awareness of the environment, such as texting or speaking on the phone, especially in dynamic settings like parking lots.

Parking lots present unique challenges to autonomous systems, characterized by low-speed maneuvers and frequent close-proximity interactions between vehicles and pedestrians. Unlike traditional roadway environments, the unstructured nature of parking lots necessitates a higher level of interpretability and adaptability in perception models. Pedestrians navigating these environments may exhibit distracted behaviors, such as ”looking behind” or ”on a call,” increasing the complexity of ensuring safety. Additionally, the interplay of pedestrian speeds, such as ”standing,” ”walking,” or ”running,” further complicates the task of risk assessment and prediction.

While many existing datasets focus on structured driving scenarios, such as highway navigation and intersection management, they often lack the granularity needed for pedestrian action analysis in confined and irregular spaces. Notable efforts, including KITTI[3] and nuScenes[2], have paved the way for standardized benchmarking of autonomous driving algorithms but have left a gap in pedestrian-centric studies. This paper addresses that gap by presenting an annotated database explicitly designed to capture pedestrian behavior in parking lots. By focusing on distracted behaviors and their associated speeds, this dataset adds a crucial dimension to understanding human actions in contexts where awareness and reaction times are critical for safety.

The presented database shares similarities with existing efforts in action recognition with applications extending far beyond autonomous driving, encompassing areas such as surveillance, human-computer interaction, and public safety. Unlike many datasets tailored for specific environments or purposes, this database offers a unique focus on pedestrian actions in unstructured parking lot scenarios, bridging a critical gap in the research landscape. Action recognition, as a broader domain, has seen advancements in datasets such as AVA-Kinetics[7] or HMDB[6]. However, these efforts lack the contextual specificity required to address safety-critical environments like those encountered in autonomous driving applications.

A key strength of this dataset lies in its practicality. Recorded using an off-the-shelf camera, it ensures accessibility and cost-efficiency, enabling researchers and developers to replicate the data collection process without specialized equipment. This design choice significantly lowers the barrier for adoption, facilitating the seamless integration of findings from models trained on this dataset into real-world applications. The straightforward setup also aligns with industry requirements for scalable solutions, ensuring that the methods developed can transition from controlled experiments to operational systems with minimal modifications.

Moreover, the database's versatility allows it to contribute to cross-disciplinary applications. For instance, the detailed action annotations and speed categorizations can inform surveillance systems, enhancing their ability to detect risky pedestrian behaviors. In human-computer interaction, such nuanced data can improve gesture recognition and situational awareness in collaborative environments. This adaptability ensures that the insights gained extend well beyond the autonomous driving domain, enriching research and applications in various contexts.

By prioritizing simplicity in hardware requirements while maintaining robust annotation standards, this dataset stands out as a resource that combines accessibility with high utility. Its relevance to both autonomous systems and broader fields underscores its potential to accelerate progress across multiple domains, driving innovation in action recognition and its practical deployment in real-world settings.

This paper is structured as follows: first, we present the related work, contextualizing the contribution within the broader research landscape. Next, we outline the methodology for dataset collection and annotation, followed by a description of its structure and content. We then present preliminary analyses and discuss potential applications. Finally, we address limitations and future directions.

Related Work

The study of autonomous driving heavily relies on datasets that offer comprehensive and varied scenarios for model training and validation. Notable contributions include KITTI [3], which has served as a benchmark for 3D object detection, tracking, and visual odometry, and nuScenes [2], which extended the scope to include multimodal data and annotations across diverse urban environments. While these datasets have significantly advanced perception and localization tasks, they primarily focus on structured road environments, with limited attention to pedestrian behaviors in unstructured contexts like parking lots.

Efforts to address pedestrian-related challenges have led to datasets such as PIE (Pedestrian Intention Estimation)[8], which annotates pedestrian behaviors and crossing intentions. The JAAD (Joint Attention in Autonomous Driving) dataset[4] focuses on pedestrian actions and environmental conditions, offering insights into contextual interactions in urban environments. However, these datasets often lack detailed annotations for specific distracted behaviors or the speeds at which these actions occur, which are critical for understanding pedestrian dynamics in high-interaction zones.

HighD[5] and inD[1] datasets, which provide bird's-eye views of vehicle and pedestrian interactions at highways and intersections, have contributed to understanding traffic dynamics. Similarly, INTERACTION[9] captures complex driver-pedestrian interactions but focuses more on vehicle behaviors at intersections rather than the nuanced actions of pedestrians. These efforts underscore the importance of studying pedestrian actions but highlight the need for datasets tailored to specific scenarios, such as parking lots, where pedestrian distractions and varying speeds play a crucial role.

This paper builds upon these efforts by presenting a specialized dataset that captures distracted pedestrian behaviors and their corresponding speeds in parking lot settings. By incorporating lessons from prior work and addressing specific gaps, this dataset aims to complement existing resources, advancing research in both pedestrian behavior modeling and autonomous systems development.

Methodology

The videos were recorded in a single parking lot scenario, ensuring a consistent environment while capturing diverse interactions between pedestrians and the surroundings. The choice of a fixed camera location ensures a full view of the scene, reducing occlusions and enabling comprehensive annotation. To facilitate machine learning applications, the dataset provides annotations in widely used formats, compatible with popular training and evaluation pipelines.

Database Description

The presented database is designed to capture and annotate pedestrian actions in unstructured environments, particularly parking lots. It focuses on actions indicative of distraction and their corresponding speeds, providing a unique dataset for advancing autonomous driving and surveillance applications. The database consists of annotated videos recorded with a single off-the-shelf camera, ensuring accessibility and ease of reproduction.

Each video frame is manually annotated with bounding boxes for all visible pedestrians, accompanied by unique identifiers for each individual. These identifiers allow tracking across frames, enabling temporal action analysis. The dataset includes four distinct action categories: "looking behind", "on a call", "texting", and a baseline category, "no_action", for moments when none of the specified actions are being performed. Additionally, the dataset incorporates speed annotations with three distinct categories: "standing", "walking", and "running". This dual labeling approach allows for a nuanced analysis of pedestrian behaviors, considering both the nature of the action and its dynamic context. Some examples of the dataset and its annotations can be seen in Figure 1, where the person ID and the action performed are written in red while the color of the bounding box indicates the speed (gray for "standing", blue for "walking" and green for "running").

The videos were recorded in a single parking lot scenario, ensuring a consistent environment while capturing diverse interactions between pedestrians and the surroundings. The choice of a fixed camera location ensures a full view of the scene, reducing occlusions and enabling comprehensive annotation. To facilitate machine learning applications, the dataset provides annotations in widely used formats, compatible with popular training and evaluation pipelines.

To enhance usability, the dataset also includes descriptive statistics such as:

- The total number of frames in the dataset.
- The number of frames in which each individual appears.
- Frequency distributions for each action and speed combination.
- Histograms of bounding box sizes, offering insights into pedestrian proximity and scaling considerations.

Table 1 provides a detailed breakdown of the dataset, summarizing the number of frames per individual, action, and speed category. Each row corresponds to a unique pedestrian identified by an ID, while columns provide:

- **Total Frames:** The total number of frames in which the pedestrian appears.
- **Actions:** The distribution of frames across different actions, including *no action*, *looking behind*, *on a call*, and *texting*.
- **Speeds:** The breakdown of frames by movement speed: *standing*, *walking*, and *running*.

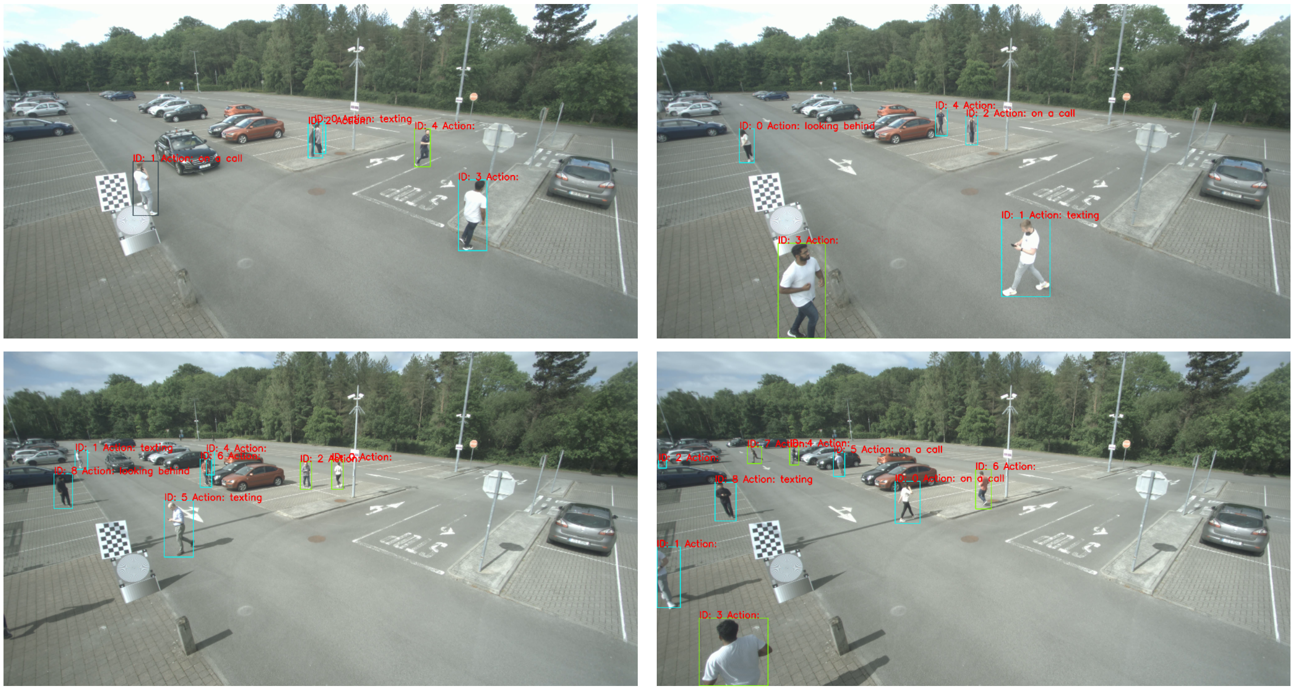

Figure 1. *Some examples of the annotations from the recorded videos, where bounding boxes, IDs, actions and speeds are shown. Speed is represented by the color of the bounding box: gray for "standing", blue for "walking" and green for "running".*

This breakdown helps researchers analyze how often each action occurs and how pedestrian speeds vary with different activities. The table also facilitates model benchmarking by providing a reference for class distributions within the dataset.

Figure 2 presents two histograms: the left histogram depicts the distribution of sequence lengths for each action category in the dataset, while the right histogram shows the distribution of bounding box sizes for different action categories. These insights help to understand the frequency and duration of various pedestrian activities and how pedestrian appearance varies with different actions.

Similarly, Figure 3 displays the distribution of sequence lengths for each pedestrian speed category on the left and the bounding box size distribution across different speed categories on the right. This provides insights into how pedestrian movement affects bounding box variations.

In Figure 4, we analyze the combined effect of actions and speeds on pedestrian annotations. The left histogram illustrates the distribution of sequence lengths for different action-speed pairs, while the right histogram presents the corresponding bounding box size distributions. This figure offers a comprehensive view of variations in pedestrian appearance and behavior dynamics.

Conclusions and Future Work

The annotated database presented in this study provides a novel resource for advancing pedestrian action recognition, particularly in unstructured environments such as parking lots. By focusing on actions indicative of distraction—such as ”looking behind,” ”on a call,” and ”texting”—and incorporating speed annotations, this dataset contributes to safer and more adaptive autonomous systems. The bounding box annotations and unique person identifiers further enhance its utility for tasks like pedestrian tracking and behavior modeling. The results obtained from this database have significant implications for autonomous driving and surveillance applications, where understanding pedestrian behavior is critical to ensuring safety and efficiency.

Despite its contributions, the database has several limitations. The dataset was recorded in a controlled environment with a single off-the-shelf camera, which, while practical, limits its scope in terms of diversity and environmental variability. Scenarios involving extreme weather, varied lighting conditions, or densely crowded areas are not adequately represented. Furthermore, the dataset focuses on a specific set of actions and speeds, which may not encompass all potential pedestrian behaviors of interest in real-world settings. These limitations underline the need for more comprehensive datasets to address these gaps.

Future work will focus on addressing these limitations by expanding the dataset to include a broader range of scenarios, including diverse environmental conditions, complex pedestrian interactions, and crowded spaces. Integrating data from multiple cameras or other sensor modalities, such as LiDAR, could provide richer spatial and temporal information, enabling more robust action recognition. Moreover, future efforts could explore automated annotation methods to scale dataset creation while maintaining high annotation quality. These enhancements aim to establish a more comprehensive and versatile resource for the research community.

By addressing these challenges, the dataset can serve as a foundation for developing advanced machine learning models capable of understanding nuanced pedestrian behaviors. This progress will support safer and more reliable autonomous driving

Table 1. *Descriptive information of the dataset, including the total number of frames and the number of frames for each individual, action and speed.*

Total number of frames: 21580								
ID	Total	Action				Speed		
		No action (default)	Looking behind	On a call	Texting	Standing	Walking	Running
0	20502	12215	981	3338	3968	583	16347	3572
1	19233	10617	543	3812	4261	470	15170	3593
2	20864	13779	671	4291	2123	461	17470	2933
3	19474	13459	269	3086	2660	1049	13983	4442
4	16909	12683	339	1576	2311	446	11133	5330
5	10620	7028	240	1521	1831	831	7462	2327
6	10321	7368	219	1219	1515	966	7346	2009
7	7630	4735	-	1413	1482	700	5244	1686
8	9799	7078	139	1291	1291	800	6716	2283

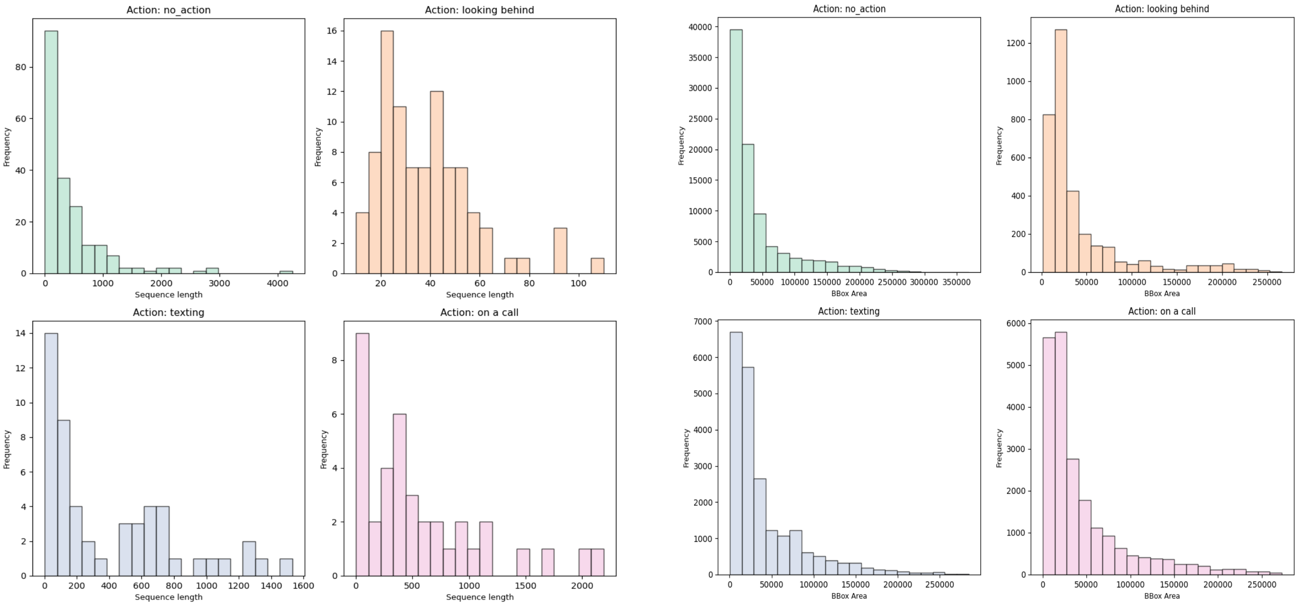

Figure 2. *Histograms of sequence lengths and bounding box sizes for different action categories.*

systems and extend the dataset's applicability to other domains, such as urban planning, surveillance, and human-computer interaction.

Acknowledgments

This work has been partially funded by the Basque Government, Spain, under Grant number IT1427-22; the Spanish Ministry of Science (MCIU), the State Research Agency (AEI), the European Regional Development Fund (FEDER), under Grant number PID2021-122402OB-C21 (MCIU/AEI/FEDER, UE). It was also supported, in part, by Science Foundation Ireland grants 13/RC/2094 P2 and 18/SP/5942, and co-funded under the European Regional Development Fund through the Southern and Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software (www.lero.ie), and by Valeo Vision Systems.

References

[1] Julian Bock, Robert Krajewski, Tobias Moers, Steffen Runde, Lennart Vater, and Lutz Eckstein. The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1929–1934. IEEE, 2020.

[2] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A multi-

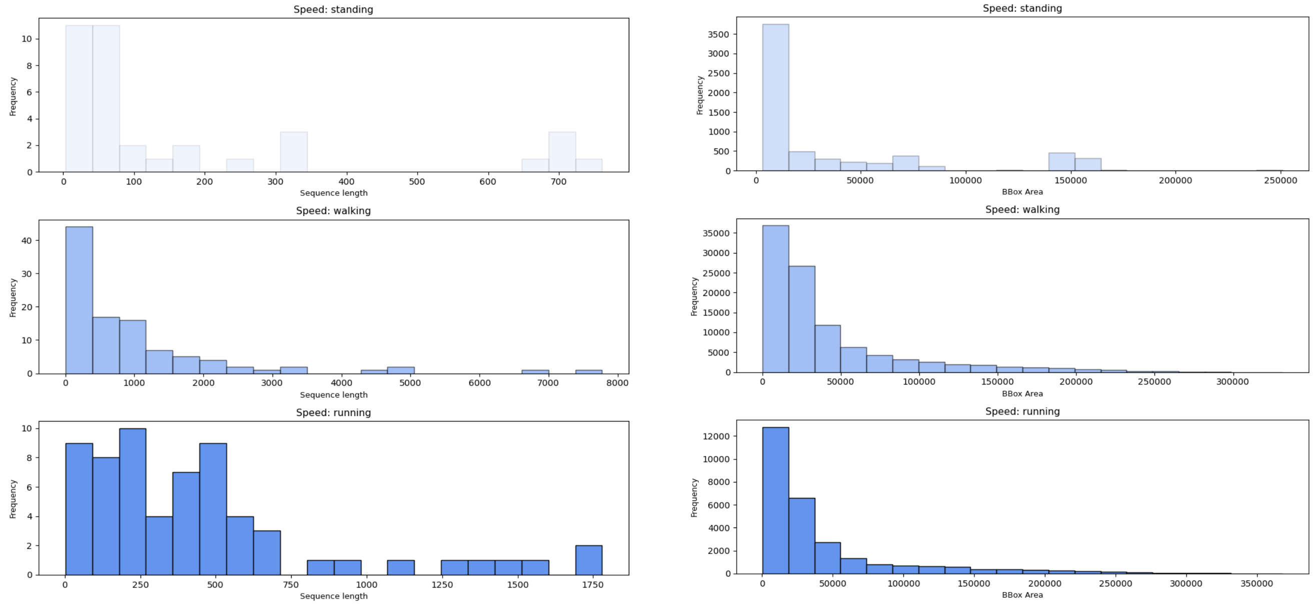

Figure 3. *Histograms of sequence lengths and bounding box sizes for different speed categories.*

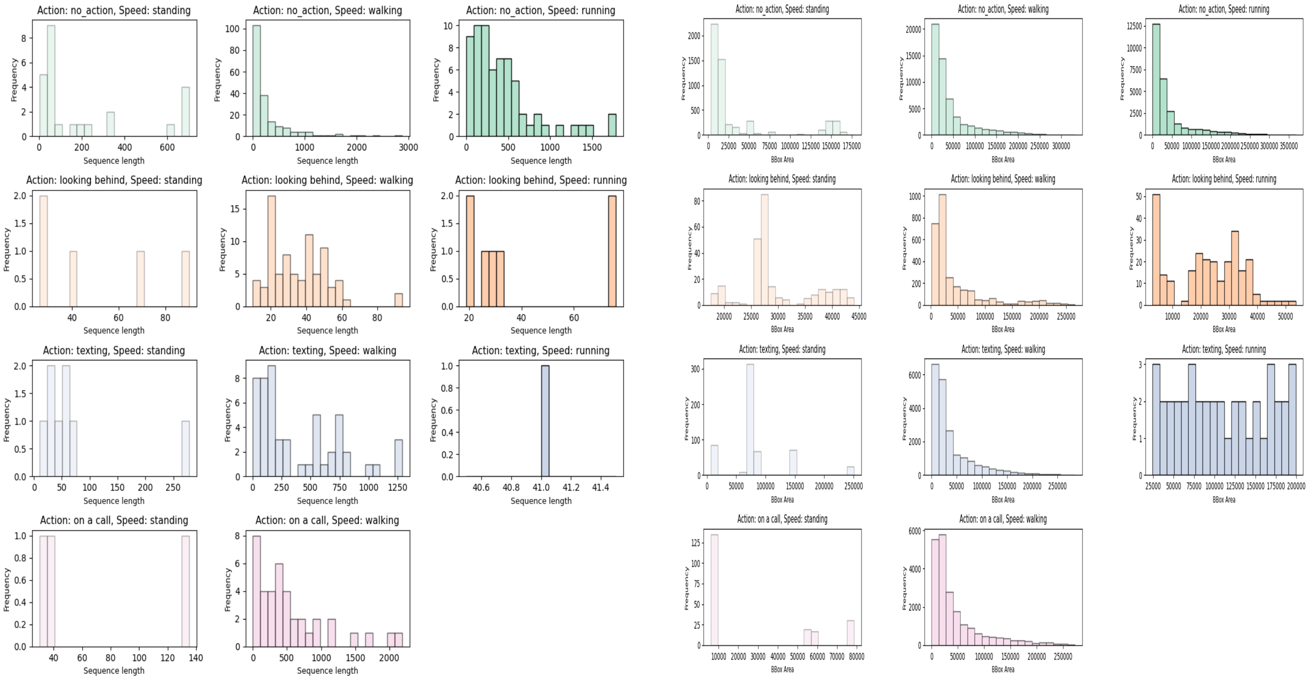

Figure 4. *Histograms of sequence lengths and bounding box sizes for different action-speed pairs*

modal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[3] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[4] Iuliia Kotseruba, Amir Rasouli, and John K Tsotsos. Joint attention in autonomous driving (JAAD). *arXiv preprint arXiv:1609.04741*, 2016.

[5] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In *2018 21st international conference on intelligent transportation systems (ITSC)*, pages 2118–2125. IEEE, 2018.

[6] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. HMDB: a large video database for human motion recognition. In *2011 International conference on computer vision*, pages 2556–2563.

IEEE, 2011.

[7] Ang Li, Meghana Thotakuri, David A Ross, João Carreira, Alexander Vostrikov, and Andrew Zisserman. The AVA-Kinetics localized human actions video dataset. *arXiv preprint arXiv:2005.00214*, 2020.

[8] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K Tsotsos. PIE: A Large-Scale Dataset and Models for Pedestrian Intention Estimation and Trajectory Prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6262–6271, 2019.

[9] Wei Zhan, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilian Naumann, Julius Kummerle, Hendrik Konigshof, Christoph Stiller, Arnaud de La Fortelle, et al. INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. *arXiv preprint arXiv:1910.03088*, 2019.

Author Biography

Itsaso Rodríguez-Moreno *received the B.Sc., M.Sc. and Ph.D. degrees in computer science from the University of the Basque Country in 2018, 2019 and 2023, respectively. She is currently an Assistant Professor at the University of the Basque Country and a member of the Robotics and Autonomous Systems Group. Her research interests include machine learning, computer vision and robotics.*

Brian Deegan *received his B.Sc. in Computer Engineering (2004) and M.Sc. in Biomedical Engineering (2005) from the University of Limerick, and his Ph.D. from the University of Galway (2011). He worked at Valeo Vision Systems (2011–2022) on image quality research. Since 2022, he has been a Lecturer and Researcher at the University of Galway, focusing on imaging, LED flicker, and machine vision.*

Dara Molloy *received the B.E. (Hons.) degree from the University of Galway in 2018. He is currently pursuing a Ph.D. degree at the University of Galway. Dara is currently working as a member of the Connaught Automotive Research (CAR) group under the supervision of Prof. Martin Glavin and Prof. Edward Jones. His research interests include computer vision and sensor availability within an autonomous vehicle context.*

José María Martínez-Otzeta *received the B.Sc. and Ph.D. degrees in computer science from the University of the Basque Country in 1993 and 2008, respectively. He is currently a Postdoctoral Researcher with the Department of Computer Sciences and Artificial Intelligence, University of the Basque Country. He is also a member of the Robotics and Autonomous Systems Group. His research interests include machine learning, computer vision, and robotics.*

Martin Glavin *received his B.E. and Ph.D. degrees in electronic engineering from the University of Galway in 1997 and 2004, respectively. He is a Joint Director of the Connaught Automotive Research (CAR) Group and a Funded Investigator at Lero. His research interests include signal processing and embedded systems for automotive and agricultural applications.*

Edward Jones *received his B.E. and Ph.D. in electronic engineering from the University of Galway. He is a Professor at the University of Galway, with experience in academia and industry. His research focuses on DSP algorithms, embedded systems, and applications in autonomous vehicles, biomedical engineering, and speech processing.*

Basilio Sierra *received the B.Sc. and Ph.D. degrees in computer science from the University of the Basque Country in 1990 and 2000, respectively. He is currently a Full Professor at the University of the Basque Country and co-director of the Robotics and Autonomous Systems Group at the same university. He has authored over 70 journal articles in the fields of machine learning, data analysis, computer vision and robotics.*

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!

• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org