Inferring and Operating Pedestrian Behaviour Models on Autonomous Vehicles

Fanta Camara
University of Leeds
Institute for Transport Studies

Submitted in accordance with the requirements for

Doctor of Philosophy

June, 2022

Intellectual Property & Publications

The candidate confirms that the work submitted is her own, except where work has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 2 of this thesis has appeared in publication as follows:
Fanta Camara, *Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Charles Fox. Pedestrian Models for Autonomous Driving Part I: Low-Level Models, From Sensing to Tracking, in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 10, pp. 6131-6151, Oct. 2021, doi:* https://doi.org/10.1109/TITS.2020.3006768.
I developed the main idea for this work under the guidance of Charles Fox and Nicola Bellotto. I performed the narrative review work and wrote the manuscript. Charles Fox, Nicola Bellotto, Serhan Cosar and Matthias Althoff provided recommendations on the narrative review work, the structure of the manuscript and comments on the quality of the citations. The manuscript was improved by comments from all the co-authors.

The work in Chapter 3 of this thesis has appeared in publication as follows:
Fanta Camara, *Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Gustav Markkula, Anna Schieben, Fabio Tango, Natasha Merat, Charles Fox. Pedestrian Models for Autonomous Driving Part II: High-Level Models of Human Behavior, in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 9, pp. 5453-5472, Sept. 2021, doi:* https://doi.org/10.1109/TITS.2020.3006767.
I developed the main idea for this work under the guidance of Charles Fox and Matthias Althoff. I performed the narrative review work and wrote the manuscript. Charles Fox and Matthias Althoff provided recommendations on the narrative review work, the structure of the manuscript and comments on the quality of the citations. The manuscript was improved by comments from all the

co-authors.

The work in Chapter 4 of this thesis has appeared in publication as follows:
***Fanta Camara**, Patrick Dickinson, Charles Fox. Evaluating pedestrian interaction preferences with a game theoretic autonomous vehicle in virtual reality, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 78, 2021, Pages 410-423, ISSN 1369-8478,* https://doi.org/10.1016/j.trf.2021.02.017.
I developed the main idea for this work under the guidance of Charles Fox and Patrick Dickinson. I designed the study, collected and analysed the data. I interpreted the results and wrote the manuscript. Patrick Dickinson and Charles Fox provided recommendations on the study design, modelling work and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 5 of this thesis has appeared in publication as follows:
***Fanta Camara**, Charles Fox. Space Invaders: Pedestrian Proxemic Utility Functions and Trust Zones for Autonomous Vehicle Interactions. International Journal of Social Robotics 13, 1929–1949 (2021).* https://doi.org/10.1007/s12369-020-00717-x.
I developed the main idea for this work under the guidance of Charles Fox. I performed the modelling work, interpreted the results and wrote the manuscript. Charles Fox provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 6 of this thesis is accepted for publication as follows:
***Fanta Camara**, Charles Fox. Extending Quantitative Proxemics and Trust to HRI. 31st IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), 2022. (Accepted)*
I developed the main idea for this work under the guidance of Charles Fox. I performed the modelling work, interpreted the results and wrote the manuscript. Charles Fox provided recommendations on the modelling and comments on the results. The manuscript was improved by comments from all the co-authors.

The work in Chapter 7 of this thesis is a manuscript under review:

Fanta Camara*, Chris Waltham, Grey Churchill, Charles Fox. OpenPodcar: an Open Source Vehicle for Self-Driving Car Research.* https://arxiv.org/abs/2205.04454v1 *(Under review)*

I developed the main idea for this work under the guidance of Charles Fox. I performed the physical podcar automation work including the mechanical design, printed circuit board (PCB), ROS localisation and mapping, path planning, people detector and tracker integration, Gazebo 3D simulation worlds of Lincoln and Leeds university campuses, testing, and wrote the documentation for the physical podcar and the manuscript. Chris Waltham designed the initial electronics circuit and safety systems, wrote the Arduino code for the speed control and participated in the initial remotely-controlled driving test. Grey Churchill developed the podcar simulator with path planning in ROS/Gazebo and wrote the documentation for the simulator. Charles Fox supervised the work, wrote the manuscript, and wrote some ROS code. The manuscript was improved by comments from all the co-authors.

Acknowledgements

"No duty is more urgent than that of returning thanks." James Allen (1864–1912)

"Let us be grateful to the people who make us happy; they are the charming gardeners who make our souls blossom." Marcel Proust (1871–1922)

I would like to express my sincere and deep gratitude to my PhD supervisor, Charles Fox. I am thankful for his unfailing support, his great mentoring and for giving me so much freedom in doing my research. I also can't thank him enough for his remarkable patience, kindness and for his true friendship. I could not have wished for a better and nicer supervisor. Thanks Charles, you are awesome !

I am very grateful to Samantha Jamson for also advising me during the final stages of my thesis. This work would not have been possible without the many collaborations I've had, including Nicola Bellotto, Serhan Cosar, Patrick Dickinson, Matthias Althoff, Jingyuan Wu, Johannes Ruenz and many others. Thank you all for your invaluable time, insightful advice and interesting discussions.

I am very thankful to my examiners, David Hogg and Kris Kitani, for their invaluable time and very useful feedback.

I would like to thank Zia Wadud and Dan Johnson for their constant support during my studies in ITS. I also thank the University of Leeds for my PhD scholarship and Natasha Merat for the EU interACT project funding. A big thank you to the University of Lincoln and L-CAS members for welcoming me so kindly into the group and for their help in many occasions. I also thank the Robotics at Leeds Exec and PGR Network committee members for their kindness and support.

My list of "charming gardeners" is really too long for mentioning all the individual names here, but I am sincerely thankful to all my supervisors, teachers, collaborators, colleagues, mentors, friends and family. Thank you so much!

Wherever I have lived, I have been fortunate to meet some really fantastic friends to whom I am very grateful for their kindness, camaraderie and friendship. My sincere and deep gratitude also goes to Rémi, Mr Gosset and many other people who always supported me and believed in me.

Finally, I am deeply grateful and forever indebted to my family for everything. None of this would be possible without you, I dedicate this thesis to you.

Abstract

Interacting with pedestrians remains challenging for autonomous vehicles (AVs). In most current AVs, for safety and legal reasons, pedestrians are considered as obstacles, such that the AVs always stop for them. But their highly safe nature may lead pedestrians to take advantage over them and slow their progress.

When a pedestrian wishes to cross the road in front of the vehicle at an unmarked crossing, the pedestrian and AV must compete for the space, which may be considered as a game-theoretic interaction in which one agent must yield to the other. Game theoretic approaches have been used for decades to model the interactions between rational decision-makers, but have run in parallel streams with psychology research on human proxemics and trust. Results from game theory and psychology studies have yet to be operationalised for autonomous vehicles, this thesis thus aims to bridge the gap between these separate fields.

We first contribute with a comprehensive review of the literature in which we propose a new unifying taxonomy of pedestrian models required for autonomous driving, linking the low-level and high-level models of behaviour for the first time. We find that the low-level models are mature enough to be deployed in the real world but the high-level models such as game theory approaches still require more research and development. We therefore proceed with the evaluation of pedestrian interaction preferences with a game theoretic AV in a virtual reality experiment. Knowledge of such preferences could then be used by future AVs to predict and control for pedestrian behaviour. However, game theory approaches require the use of credible threats such as crash probabilities in order to make AVs progress on the roads, but another possible and more friendly solution that is explored in this work is Hall's theory of proxemics. Hence we propose a novel Bayesian method to infer pedestrian proxemic utility functions and the concept of physical trust requirement (PTR) for game theoretic AV interactions. We show how this PTR model can accurately generate Hall's empirical zone sizes, and then extend it to more general human-human and human-robot interactions. Finally, operating and deploying pedestrian behaviour models require the use of a physical AV platform, we hence introduce OpenPodcar, a low-cost, open source hardware and software platform developed for real-world AV research experiments with pedestrians. Thus, the present thesis forms a step towards the first operational game theoretic autonomous vehicles with pedestrian proxemic and trust behaviour.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Why did the chicken cross the road?"

1 Background

Autonomous vehicles (AVs), also called self-driving cars, intelligent/automated vehicles or autonomous driving systems (ADS), are appearing on the roads [35], thanks to huge improvements on Simultaneous Localisation And Mapping (SLAM) algorithms [5, 33] together with new, cheap sensors and computation technologies [14] and also thanks to huge public and private investments. For instance, in 2015, the UK government alone invested £100 million in research and development (R&D) for the deployment of Connected Autonomous Vehicles (CAV) technologies [22] and the global market is estimated to be worth £907 billion in 2035 [7]. Self-driving cars can navigate safely on roads, promising a future society with a better mobility system with fewer accidents and traffic in cities [21], less energy consumption [15, 36] and air pollution [3].

The Society of Automation Engineers (SAE) defined five levels of automation for these autonomous vehicles [31], as shown in Fig. 1.1. Many automotive companies currently claim having developed level 3 automated vehicles, the race is now towards the full automation [29]. But after decades of development and despite the global enthusiasm around AVs and the big

Figure 1.1: SAE levels of driving automation (source: sae.org)

investments, some major challenges still remain [16]. In fact, before the fully autonomous vehicle revolution happens, AVs will have to share space with and will be challenged by human drivers and pedestrians, who are much harder to model and act upon than passive environments. Navigating around humans in dynamic environments requires the understanding of human social behaviour and remains an active research question [32].

Unlike static environments, pedestrians are complex interactive agents having their own goals, utilities, and decision making systems, and interactions with them must take these into account in order to predict their actions and plan accordingly. This is critical in environments where traffic rules do not clearly define priority, such as at unmarked intersections, where AVs and pedestrians have to negotiate over who will pass the other [27]. The ability of human drivers to read the intention of other road users and predict their future behaviour and then interact with them is currently missing on autonomous vehicles [28, 29]. Brooks identifies the need for these higher levels of interaction as 'the big problem with self-driving cars' [4].

Currently AVs are designed to be as safe as possible, and must always yield to other road users in competition with them. But recent real-world autonomous vehicle studies have shown that pedestrians may then take advantage of their predictable behaviour [8, 19, 21], pushing in front of them for priority eventually in *every* negotiation, so that the vehicles will hence make zero progress in busy areas, this has been known as the 'freezing robot problem' [34]. Similar difficulties arise with Autonomous Ground Vehicles (UGV) on high street pedestrianized areas. AVs thus need better prediction and decision-making models.

To address this problem in this thesis, we argue that AVs, like human-driven vehicles, may thus sometimes need to maintain a credible threat of colliding with or otherwise causing some smaller negative utilities to pedestrians in order to sometimes obtain priority over them and make progress. Autonomous vehicles hence need to actively infer and predict pedestrian utilities and decisions in order to control interactions with them and navigation around them. This requires an understanding of human negotiation strategies and social behaviour, which have long been studied in game theory and psychology but mainly in parallel streams. Results from these fields have not yet been translated into robotic control systems, and leave many questions still unanswered. In this thesis, we thus attempt to bridge the gap between these separate fields and we propose methods and solutions to bring them to an operational level for fully automated vehicles' interactions with pedestrians.

2 Research Gaps and Objectives

In this thesis, our aim is to create bridges between game theory and psychology approaches in order to enable and improve real-time pedestrian behaviour inference and operation for autonomous vehicles. For this purpose, we split our work into small steps, and in particular, we identify several key research gaps and a set of objectives to address them.

The first research gap is that the above SAE levels of automation define the ultimate behaviour that autonomous vehicles should have, but they do not provide a clear guidance about the types of pedestrian behaviour models required to reach a specific level of automation. Previous literature reviews have focused on specific topics from a specific field of research, such as detection [9] and tracking [18] from machine vision community, pedestrian crossing behaviour [20, 24] from transport studies or human nonverbal communication [30] from robotics. This thus leads to the first objective, which is to write a comprehensive review of the literature that identifies the existing pedestrian models from these different fields and propose a suitable framework that can connect and classify them into the relevant SAE levels. For instance, a model defined at level 1 should be able to pass its outputs to the model used at level 2, whose output should then pass to the model at level 3, and so on. Defining this framework is not only needed to connect game theory and psychology models together but is also necessary to connect them with the rest of the AV software stack. Chapter 2 will focus on the low-level models of pedestrian behaviour and Chapter 3 will address the high-level models of the stack.

The second gap is that there is a lack of a clear tractable decision-making model for real-time AV control. Interaction is recursive and complex: the AV's own actions will affect the pedestrian's actions and vice versa over time. This is most critical in environments where traffic rules do not clearly define priority for any of the participants, such as at unmarked intersections, where AVs and pedestrians have to negotiate with one another for priority. Conflict rates at unsignalized intersections are much higher than in other types of intersections [17] because the priority is not defined, and each agent acts based on their own interpretation. As a first step, the objective is to identify a suitable decision-making model which would address the 'Freezing Robot Problem' and allow the AV to make some progress during its interactions. Also, when drivers speed up and slow, these are not just ways to control their progress but also send information about their risk preferences to pedestrians engaged in negotiations for priority. Hence, the AV decision-making model should take pedestrian risk-taking preferences into account. The chosen AV decision-making model will be introduced in

Chapter 3 and detailed further in Chapter 4.

The third research gap that we identify is the lack of suitable methods and data to infer pedestrian interaction preferences with vehicles. To address this, the objective is to run empirical studies with human participants and use the data to fit parameters to the AV decision-making model. This could be done using lab experiments data or in a virtual reality environment for more realism. These studies should validate the AV interaction control model and provide specific parameter values that measure pedestrian behaviour preferences. We should also explore ways to unveil pedestrian preferences from different AV driving styles in different environments in order to find out interaction preferences for the AV behaviour itself. Chapter 4 will detail the methods and experiments used in this work.

Another gap in this field is that social norms dictate pedestrian behaviour [28, 32], hence the AV decision-making model needs to integrate psychology results but these are often provided in the form of qualitative or discrete data, which cannot be easily operationalised onto autonomous vehicle control software. The objective is then to perform a review of human nonverbal cues, which would allow us to assess the different forms of reading and giving nonverbal cues between agents, whilst keeping the interactions safe. In particular, we should identify pedestrian nonverbal cues and explore methods to integrate them into the AV decision-making model. Also, understanding human psychology and cognitive precesses is a topic of great interest in general human-human and human-robot interactions, we should therefore explore ways to generalize our models to dual agent interactions. Chapter 3 will review psychology-based interaction cues, their quantitative development and possible integration to the AV decision-making will be introduced in Chapter 5.

Further, we identify another research gap, and not the least, that is the clear lack of an experimental platform on which to operate pedestrian behaviour models for AV research. To make autonomous driving a reality, it is not sufficient to test algorithms in simulation or virtual reality, but testing them on a physical and realistic platform is of utmost importance. However, autonomous vehicles are composed of many different sensors and components, it is therefore very expensive to own or develop such a vehicle in every research lab. It is rare to test one's own algorithms on a real AV platform, because most of the autonomous vehicle platforms for research belong to big research labs or automotive companies, and these are often used to collect datasets that are then shared with the rest of the community, such as *nuscenes* [6] from Motional part of Hyundai Motor Group or ApolloScape [37] from Apollo part of Baidu. Moreover, most open source AV research platforms are RC-scaled car projects such as F1Tenth [1], AutoRally [11] and MIT Racecar [2], which are not large enough for realistic pedestrian interactions. Bigger open source hardware platforms exist such as PixBot [26] and Tabby EVO [23], but these are too complex and require expert skills and potentially dangerous processes for replication. The medium-sized AV research vehicles, SMART [25] and iCab (Intelligent Campus Automobile) [12], that could have facilitated the evaluation of pedestrian models are not open source. Therefore the objective in this work is to propose a low-cost and open source solution that is large and safe enough for realistic pedestrian interactions, and which would make AV testing more accessible to the wider community. Such an AV platform would be useful to implement and validate the work produced in this thesis. Chapter 7 will present our solution to that problem.

3 Thesis Outline and Contributions

This section presents the outline and contributions of this thesis.

Chapter 2 presents a paper entitled "Pedestrian Models for Autonomous Driving Part I: Low-Level Models, From Sensing to Tracking". This paper focuses on reviewing the existing pedestrian models required for the SAE low levels 0 to 2. We propose a novel unifying taxonomy of pedestrian models based on probability theory to provide possibilities for quantitative computational interfaces for the different levels. The models for levels 0 to 2 are found to mainly come from the machine vision and robotics fields, and we show that pedestrian modelling requirements increase as the SAE levels increase. We map pedestrian sensing methods to SAE level 0, detection methods to SAE level 1 and recognition and tracking mehtods to SAE level 2. We find that pedestrian models at these low levels are mature, hence we find and suggest several standard software implementations for these models, and we discuss the current state and future directions of research in these areas. This Chapter lays the foundations for the models required in Chapter 3.

Chapter 3 presents a paper entitled 'Pedestrian models for autonomous driving Part II: high-level models of human behavior'. This paper builds upon the unifying taxonomy started in Chapter 2 but focuses here on the high-level models of pedestrian required for SAE levels 3 to 5. The models reviewed here are found to come from various fields of research such as machine vision, robotics, data science, transport studies, psychology and game theory. We map pedestrian prediction methods to SAE level 3, and we show that pedestrian interaction, game theory and signalling methods are all mapped to both SAE levels 4 and 5. This is because in practice the AV functioning is exactly the same at both levels, since no human intervention is required, except in very rare cases. In particular, we identify a mathematical model for the game of chicken, a discrete sequential game theory model, called the Sequential Chicken Game [10], for negotiation between autonomous vehicles and pedestrians at an unsignalized intersection. This model shows that not only the first agent to yield is more likely to lose the game but also if the AV only uses its position to signal its intent, there must exist a small probability for a collision to occur. This collision probability can be used as a threat for the pedestrian, preventing them from stepping intentionally in front of the AV, thus avoiding the freezing robot problem. We also identify a form of nonverbal cue called proxemics or interpersonal distancing. It is based on Hall's theory [13] that people experience a psychological sense of comfort or discomfort in interactions, depending on specific *discrete* spatial zones around them. We then discuss the current state and future directions for pedestrian models required at these high levels and we find that these are still premature and not yet operational for fully autonomous vehicles. This motives the rest of the work presented in this thesis, especially the development and testing of the Sequential Chicken game theory model detailed in Chapter 4.

Chapter 4 presents a paper entitled 'Evaluating Pedestrian Interaction Preferences with a Game Theoretic Autonomous Vehicle in Virtual Reality'. This paper focuses on the implementation of a game theoretic model on a autonomous vehicle and the inference of its parameters from human experiments. We implement the mathematical model for Sequential Chicken Game identified in Chapter 3 on an AV in a virtual reality (VR) environment. Simulation results of the Sequential Chicken model previously showed that the AV must be programmed with a small crash probability, i.e. a very high negative utility, that is used as a credible threat in order to make the AV progress on the roads, hence the virtual AV was designed not to stop in its interactions. To test this approach, we performed road-crossing experiments with human participants interacting with this virtual AV and evaluated their crossing behaviour. We then show how game-theoretic predictive parameters can be fit into pedestrians' natural and continuous motion during these road-crossings, and how predictions can be made about their interactions with AV controllers in similar real-world settings. This providing a more realistic, empirical understanding of the human factors intelligence required by future autonomous vehicles. We make use of dynamic programming to compute the optimal game theoretic solution form, then find pedestrian behavioural parameters via a Gaussian Process regression analysis. We also evaluate pedestrian preferences for the AV driving style using a gradient descent approach in two virtual environments using two vehicles sizes. We show that pedestrian utility preferences are more realistic using the VR setup compared to previous empirical lab experiments and that pedestrian behaviour is similar in different environments and with different vehicle sizes. We finally discuss the limitations and future work. The legal and ethical implications of this work suggest that an alternative must be found for the engineered crash probability in order to operate this game theoretic approach in the real world. This motivates the development of the novel proxemic model introduced in Chapter 5.

Chapter 5 presents a paper entitled 'Space Invaders: Pedestrian Proxemic Utility Functions and Trust Zones for Autonomous Vehicle Interactions'. This paper focuses on the mathematisation of pedestrian proxemics and trust, and link them quantitatively in the limited case of pedestrian-autonomous vehicle interactions. The game theory model used in Chapter 4 suggested that the large negative utility of a collision could be replaced with more frequent but smaller negative utilities inflicted on pedestrians and progress still be made by AVs without having to hit any pedestrians. We here explore the use of an alternative and physically harmless solution for this credible threat required by the Sequential Chicken model, this is provided by the theory of proxemics reviewed in Chapter 3. Proxemics offers this nice alternative, because invading someone's personal space without actually touching them creates a psychological discomfort equivalent to a small negative utility, which is currently not quantified. We hence propose a novel Bayesian approach to infer pedestrian proxemic utility in the form of *continuous* functions in order to be operational with the chicken model. We then develop a new concept and mathematisation of 'physical trust requirement' (PTR) for pedestrian–AV interactions, which can numerically generate and explain Hall's empirical proxemic zones within 4% error. We then evaluate this method on two public datasets and find empirical results of pedestrian proxemic utility functions and trust zones. We finally discuss the limitations of this work, its legal implications and potential applicability to other human-robot interaction scenarios.

Chapter 6 presents a paper entitled 'Extending Quantitative Proxemics and Trust to HRI'. This paper addresses some of the limitations mentioned in Chapter 5 and generalizes the PTR model to include more general human-human and human-robot interactions (HRI).

We present new results comparing the extended model's predictions to empirical data and find that this new model can explain Hall's empirical proxemic zones within 1% error. These links could enable research to be shared and operational between models of proxemics, trust, and robotic interactions for the first time. We finally discuss the limitations of this work and the different applications that could benefit from more quantitative models of proxemics and trust.

Chapter 7 presents a paper entitled 'OpenPodcar: an Open Source Vehicle For Self-Driving Car Research'. This paper introduces OpenPodcar, a light-weight, low-cost, easy-to-build and robust open source hardware and software autonomous vehicle platform. OpenPodcar was developed with the aim to enable the rapid development and testing of AV algorithms including the interaction and decision-making models introduced above. We explain the build steps and tests carried out to convert a mobility scooter into an autonomous (self-driving) podcar. We open source the hardware and software designs of this platform in order to give the opportunity to other researchers to reproduce this work, have their own autonomous vehicle platform, test their algorithms and run experiments in conditions similar to the real world. The platform is large enough to transport one person or delivery containers at speeds up to 15km/h. It is small and safe enough to be parked in a standard research lab and be used for realistic and real-world human-vehicle interaction studies. The OpenPodcar has several layers of safety features and its total build cost (∼USD7,000 in 2022) is also low enough to be reproduced by a large number of researchers. It thus provides a good balance between real world utility, safety, cost and research convenience. We show how this platform could be used for different use cases such as general self-driving car or HRI research. The platform includes a simulator for initial testing, and has a lower-level and a higher-level ROS stack such as GMapping and move_base with Timed-Elastic Band planner for SLAM and path planning, respectively. Further models could be plugged into this higher-level stack such as the work presented in Chapters 4, 5 and 6 but with a real physical vehicle and in real-world conditions this time. We finally discuss plans for possible extensions and future work using this platform.

4 Publications

A significant portion of research undertaken during this work is excluded in this thesis in order to reduce the length of this document. The papers are listed below.

Included Research:

These following papers form the chapters of this thesis.

- ***Fanta Camara****, Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Charles Fox. Pedestrian Models for Autonomous Driving Part I: Low-Level Models, From Sensing to Tracking, in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 10, pp. 6131-6151, Oct. 2021, doi:* https://doi.org/10.1109/TITS.2020.3006768*.*

- ***Fanta Camara****, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Gustav Markkula, Anna Schieben, Fabio Tango, Natasha Merat, Charles Fox. Pedestrian Models for Autonomous Driving Part II: High-Level Models of Human Behavior, in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 9, pp. 5453-5472, Sept. 2021, doi:* https://doi.org/10.1109/TITS.2020.3006767*.*

- ***Fanta Camara****, Patrick Dickinson, Charles Fox. Evaluating pedestrian interaction preferences with a game theoretic autonomous vehicle in virtual reality, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 78, 2021, Pages 410-423, ISSN 1369-8478,* https://doi.org/10.1016/j.trf.2021.02.017*.*

- ***Fanta Camara****, Charles Fox. Space Invaders: Pedestrian Proxemic Utility Functions and Trust Zones for Autonomous Vehicle Interactions. International Journal of Social Robotics 13, 1929–1949 (2021).* https://doi.org/10.1007/s12369-020-00717-x*.*

- ***Fanta Camara****, Charles Fox. Extending Quantitative Proxemics and Trust to HRI. 31st IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), 2022. (Accepted)*

- ***Fanta Camara****, Chris Waltham, Grey Churchill, Charles Fox. OpenPodcar: An Open Source Vehicle For Self-Driving Car Research.* https://arxiv.org/abs/2205.04454v1 *(Under review)*

Excluded Research:

The following papers have been excluded to reduce the length of this document.

- *Fox, Charles;* ***Camara, Fanta****; Markkula, G.; Romano, R.; Madigan, R. & Merat, N.* When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. *Proceedings of VEHITS 2018: 4th International Conference on Vehicle Technology and Intelligent Transport Systems, 2018. (Vol. 1, pp. 431-439) SciTePress*

- ***Camara, Fanta****; Romano, R.; Markkula, G.; Madigan, R.; Merat, N. & Fox, C.* Empirical game theory of pedestrian interaction for autonomous vehicles. *Proceedings of Measuring Behavior 2018, Measuring Behavior 2018: 11th International Conference on Methods and Techniques in Behavioral Research, Manchester Metropolitan University, 2018.*

- ***Camara, Fanta****; Giles, O.; Madigan, R.; Rothmüller, M.; Rasmussen, P. H.; Vendelbo-Larsen, S. A.; Markkula, G.; Lee, Y. M.; Garach, L.; Merat, N. & Fox, C.*Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. *Proceedings of the 15th International Conference on Intelligent Autonomous Systems (IAS-15) Workshops, 2018.*

- ***Camara, Fanta****; Cosar, S.; Bellotto, N.; Merat, N. & Fox, C.* Towards pedestrian-AV interaction: method for elucidating pedestrian preferences. *Proceedings of IEEE/RSJ Intelligent Robots and Systems (IROS) Workshop Towards Intelligent Social Robots: From Naive Robots to Robots Sapiens, 2018.*

- ***Camara, Fanta****; Giles, O.; Madigan, R.; Rothmüller, M.; Rasmussen, P. H.; Vendelbo-Larsen, S. A.; Mark'kula, G.; Lee, Y. M.; Garach, L.; Merat, N. & Fox, C.*Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. *In Proceedings of 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 2098-2103.* https://doi.org/10.1109/ITSC.2018.8569282.

- ***Camara, Fanta****; Dickinson, P.; Merat, N. & Fox, Charles* Examining pedestrian-autonomous vehicles interactions in Virtual Reality *Transport Research Arena (TRA) 2020. (conference canceled)*

- ***Camara, Fanta****; Cosar, S.; Bellotto, N.; Merat, N. & Fox, C.* Continuous game theory pedestrian modelling method for autonomous vehicles

In C. Olaverri-Monreal, F. García-Fernández, & R. J. F. Rossetti (Eds.), Human Factors in Intelligent Vehicles. River Publishers, 2020.

- ***Camara, Fanta***; *Merat, N. & Fox, Charles*
  A heuristic model for pedestrian intention estimation. *In Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC), 2019, pp. 3708-3713.* https://doi.org/10.1109/ITSC.2019.8917195.

- ***Camara, Fanta*** *& Fox, Charles*
  Game theory for self-driving cars. *Proceedings of UK-RAS Conference, 2020.*

References

[1] F1tenth. https://f1tenth.org/.

[2] MIT RaceCar. https://mit-racecar.github.io/.

[3] J. Anderson, N. Kalra, K. Stanley, P. Sorensen, C. Samaras, and O. Oluwatola. *Autonomous Vehicle Technology: A Guide for Policymakers.* 2014.

[4] Rodney Brooks. The big problem with self-driving cars is people and we'll go out of our way to make the problem worse. https://spectrum.ieee.org/transportation/self-driving/the-big-problem-with-selfdriving-cars-is-people, 2017. IEEE SPECTRUM.

[5] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J.J. Leonard. Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. *IEEE Transactions on Robotics*, (6), 2016.

[6] 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 multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[7] Transport Systems Catapult. Market forecast for connected and autonomous vehicles. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/642813/15780_TSC_Market_Forecast_for_CAV_Report_FINAL.pdf, 2017.

[8] Shuchisnigdha Deb, Lesley Strawderman, Daniel W Carruth, Janice DuBien, Brian Smith, and Teena M Garrison. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. *Transportation research part C: emerging technologies*, 84:178–195, 2017.

[9] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, April 2012.

[10] Charles W. Fox, Fanta Camara, Gustav Markkula, Richard Romano, Ruth Madigan, and Natasha Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *Proceedings of VEHITS 2018: 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, January 2018.

[11] Brian Goldfain, Paul Drews, Changxi You, Matthew Barulic, Orlin Velev, Panagiotis Tsiotras, and James M Rehg. Autorally: An open platform for aggressive autonomous driving. *IEEE Control Systems Magazine*, 39(1):26–55, 2019.

[12] D Gomez, P Marin-Plaza, Ahmed Hussein, A Escalera, and J Armingol. Ros-based architecture for autonomous intelligent campus automobile (icab). *UNED Plasencia Revista de Investigacion Universitaria*, 12:257–272, 2016.

[13] Edward Twitchell Hall. *The hidden dimension*, volume 609. Garden City, NY: Doubleday, 1966.

[14] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada. An open approach to autonomous vehicles. *IEEE Micro*, 35(6):60–68, Nov 2015.

[15] Tschangho John Kim. Automated autonomous vehicles: Prospects and impacts on society. *Journal of Transportation Technologies*, 2018.

[16] Keith Kirkpatrick. Still waiting for self-driving cars. *Commun. ACM*, 65(4):12–14, mar 2022.

[17] Miaomiao Liu, Yongsheng Chen, Guangquan Lu, and Yunpeng Wang. Modeling crossing behavior of drivers at unsignalized intersections with consideration of risk perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 45:14 – 26, 2017.

[18] Wenhan Luo, Xiaowei Zhao, and Tae-Kyun Kim. Multiple object tracking: A review. *CoRR*, abs/1409.7618, 2014.

[19] Ruth Madigan, Tyron Louw, Marc Dziennus, Tatiana Graindorge, Erik Ortega, Matthieu Graindorge, and Natasha Merat. Acceptance of automated road transport systems (arts): An adaptation of the utaut model. *Transportation Research Procedia*, 14:2217 – 2226, 2016. Transport Research Arena TRA2016.

[20] S. Mamidipalli, V. P. Sisiopiku, B. Schroeder, and L. Elefteriadou. A review of analysis techniques and data collection methods for modeling pedestrian crossing behaviors. *Journal of Multidisciplinary Engineering Science and Technology*, Vol. 2, Issue 2, February 2015.

[21] Adam Millard-Ball. Pedestrians, autonomous vehicles, and cities. *Journal of Planning Education and Research*, 38(1):6–12, 2018.

[22] House of Lords. Connected and autonomous vehicles: The future? `https://publications.parliament.uk/pa/ld201617/ldselect/ldsctech/115/115.pdf`, 2017.

[23] Open Motors. Tabby EVO. `https://www.openmotors.co/evplatform/`.

[24] Eleonora Papadimitriou, George Yannis, and John Golias. A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3):242 – 255, 2009.

[25] Scott Pendleton, Tawit Uthaicharoenpong, Zhuang Jie Chong, Guo Ming James Fu, Baoxing Qin, Wei Liu, Xiaotong Shen, Zhiyong Weng, Cody Kamin, Mark Adam Ang, et al. Autonomous golf cars for public trial of mobility-on-demand service. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1164–1171. IEEE, 2015.

[26] PixMoving. Pixbot. `https://gitlab.com/pixmoving/pixbot`.

[27] Evangelia Portouli, Dimitris Nathanael, and Nicolas Marmaras. Drivers' communicative interactions: on-road observations and modelling for integration in future automation systems. *Ergonomics*, 57(12):1795–1805, 2014. PMID: 25204887.

[28] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles*, 3(1):61–70, March 2018.

[29] Amir Rasouli and John K. Tsotsos. Joint attention in driver-pedestrian interaction: from theory to practice. *CoRR*, abs/1802.02522, 2018.

[30] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

[31] SAE International. SAE J3016 levels of driving automation. https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic, 2019.

[32] Andrea Thomaz, Guy Hoffman, and Maya Cakmak. Computational human-robot interaction. *Found. Trends Robot*, 4(2–3):105–223, dec 2016.

[33] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. Intelligent robotics and autonomous agents. MIT Press, 2005.

[34] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803, Oct 2010.

[35] Michael Wade. Silicon valley is winning the race to build the first driverless cars. https://theconversation.com/silicon-valley-is-winning-the-race-to-build-the-first-driverless-cars-91949, February 2018.

[36] Z Wadud, D MacKenzie, and P Leiby. Help or hindrance? the travel, energy and carbon impact of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, 86:1–18, April 2016. (c) 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[37] Peng Wang, Xinyu Huang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

Chapter 2

Pedestrian Models for Autonomous Driving Part I: Low-Level Models, From Sensing to Tracking

Abstract

Autonomous vehicles (AVs) must share space with pedestrians, both in carriageway cases such as cars at pedestrian crossings and off-carriageway cases such as delivery vehicles navigating through crowds on pedestrianized high-streets. Unlike static obstacles, pedestrians are active agents with complex, interactive motions. Planning AV actions in the presence of pedestrians thus requires modelling of their probable future behaviour as well as detecting and tracking them. This narrative review article is Part I of a pair, together surveying the current technology stack involved in this process, organising recent research into a hierarchical taxonomy ranging from low-level image detection to high-level psychology models, from the perspective of an AV designer. This self-contained Part I covers the lower levels of this stack, from sensing, through detection and recognition, up to tracking of pedestrians. Technologies at these levels are found to be mature and available as foundations for use in high-level systems, such as behaviour modelling, prediction and interaction control.

1 Introduction

Many organisations are vigorously developing autonomous vehicles (AVs). The technology for vehicles moving in static environments – localising, mapping, planning, and controlling – is well developed [219] and is now available as open-source software [116]. However, in real-world driving environments, human drivers regularly make decisions involving social decision-making that are harder to automate. Autonomous vehicles need additional social intelligence to operate in these complex social environments.

Interacting with pedestrians is a particular type of social intelligence. Autonomous vehicles will need to utilize many different models of pedestrians, each addressing different aspects of perception and intelligence from low-level machine vision detection to high-level psychological and social reasoning. Each of these models can be based on empirical science

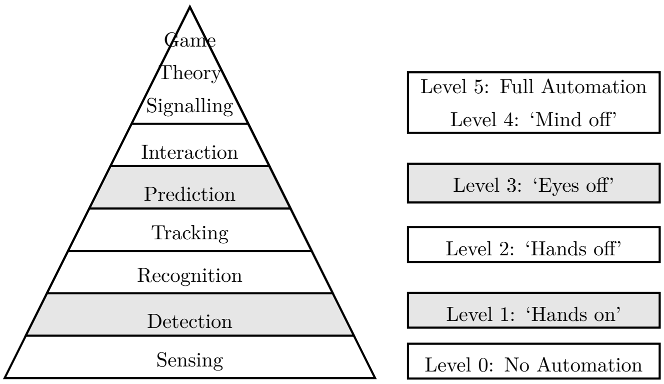


Figure 2.1: Main structure of the review.

results or obtained via machine learning. So far, the required models have typically been developed by different research communities, so their integration is currently premature.

At the lower levels of the technology stack, pedestrian modelling requires perceptual methods to detect pedestrians, track their positions and velocities over time, and predict their movements to avoid colliding with them. These methods mostly originate from computer vision and robotics.

At the higher-levels, as researched by psychologists and taught in advanced driver training programmes, drivers may infer the personality of other humans, predict their likely behaviours, and interact with them to communicate mutual intentions. At the higher levels, researchers infer psychological information from perceptual information, for example recognizing pedestrian body language, gestures, and demographics information, to better predict their likely goals and behaviours. Despite the importance of bridging the research between the higher and lower levels, their connection is still thin, both conceptually and in terms of implementations.

A promising method to bridge the higher and the lower levels is probability theory, providing possibilities for quantitative computational interfaces: for example, a pedestrian detector can pass a detection probability to a gesture recognizer, which computes probabilities of particular gestures based on this information, which in turn can be passed to a psychological or game-theoretic behaviour predictor, before the information is finally used to probabilistically compute optimal steering and speed values. Such a unified probabilistic stack requires models at all levels to realise quantitative, probabilistic inferences and predictions. Besides surveying the required building blocks, we also examine the maturity of each required level.

Many papers have been published presenting pedestrian models at various levels, but no

Table 2.1: Proposed mapping from SAE levels to pedestrian model requirements.

SAE LEVEL	DESCRIPTION	MODEL REQUIREMENTS	SECTION
0	No Automation. Automated system issues warnings and may momentarily intervene, but has no sustained vehicle control.	Sensing	Sec. 2
1	Hands on. The driver and the automated system share control of the vehicle. For example, adaptive cruise control (ACC), where the driver controls steering and the automated system controls speed. The driver must be ready to resume full control when needed.	+Detection	Sec. 3
2	Hands off. The automated system takes full control of the vehicle (steering and speed). The driver must monitor and be prepared to intervene immediately. Occasional contact between hand and wheel is often mandatory to confirm that the driver is ready to intervene.	+Recognition +Tracking	Sec. 4 Sec. 5
3	Eyes off. Driver can safely turn attention away from the driving tasks, e.g. use a phone or watch a movie. Vehicle will handle situations that call for an immediate response, like emergency braking. The driver must still be prepared to intervene within some limited time.	+Unobstructed Walking Models, Known Goals +Behaviour Prediction, Known Goals +Behaviour Prediction, Unknown Goals	Part II Sec. 2.1 Part II Sec. 2.2 Part II Sec. 2.3
4	Mind off. No driver attention is required for safety, except in limited spatial areas or special circumstances.Outside of these areas or circumstances, the vehicle must be able to safely abort or transfer control to the human.	+Event/Activity Models +Effects of Class on Trajectory +Pedestrian Interaction Models +Game Theory and Signalling Models	Part II Sec. 2.4 Part II Sec. 2.5 Part II Sec. 3 Part II Sec. 4
5	Full automation. No human intervention is required at all.	+Extreme Robustness and Reliability	

Note: '+X' means that 'X' is required in addition to the requirements of the previous level.

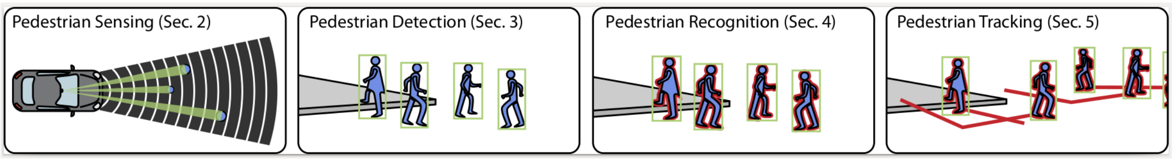


Figure 2.2: Structure of the paper.

unifying theory to connect them has yet been produced. The present study is Part I of a linked pair which together survey and unify the stack of required skills from engineered low-level aspects up to high-level aspects involving social decision-making. This Part I reviews the lower-level parts of the stack from sensing, through detection and recognition, to tracking, which together create the required inputs for higher-level AI systems to control interactions reviewed in Part II [29].

Together, these two reviews contribute steps towards such a theory by bringing together, and organising into a new taxonomy (presented via the structure of the papers), research from different fields, including machine vision, robotics, data science, psychology and game theory. We suggest how models from these fields could be linked together into a single technology stack by probability theory. We support this goal by summarizing methods for translating qualitative concepts into simple quantitative statistical models.

Fig. 3.1 provides an overview of the main structure of the review and links the structure to five levels of driving automation defined by the Society of Automotive Engineers (SAE), ranging from simple driver assistance tools to full self-driving [190]. In our taxonomy, we approximately map requirements for pedestrian modelling to each of these levels, with requirements increasing as levels increase. Table 3.1 gives an overview of SAE levels and requirements mapping.

To reach level 0, no automation is required, but some basic sensing is needed to inform the human driver. Very simple sensors can be used, such as the ultrasonic reverse parking sensors currently available commercially, together with very basic signal processing such as distance thresholds causing an audible signal. More complex concepts from our reviews *may* also be added to inform the driver of higher-level information, such as the identity of the particular pedestrian they are about to hit, but this is not *necessary* to *reach* level 0.

To reach level 1, the AV needs to provide driving assistance tools, such as lane keeping and adaptive cruise control (ACC). To do this, it needs to *detect* the road structure and the surrounding objects to help the driver. The AV needs to detect these objects in order to avoid them, but does not yet need to *recognise* them as specific individuals because this is not necessarily needed for obstacle avoidance.

To reach level 2, the AV and the driver must share the driving task, with the vehicle taking full control of the vehicle at certain times. To take full control, it is not sufficient to only detect objects, but it is also necessary to *recognize* and track them over time in order to make *short-term* predictions of their motion and safely avoid them, possibly often passing control to the human, when these simple predictions do not work.

To reach level 3, drivers can turn their attention away from the driving task, but must be prepared to take control occasionally within a certain time. This requires better prediction of pedestrian motion than level 2 in order to reduce take-over requests to humans. For example, adding concepts of likely routes and destinations to pedestrian models reduces the human take-over requests.

Finally, to reach levels 4 and 5, we believe that the AV must understand the driving task as good as a good human driver. Human drivers use complex psychology of pedestrian behaviour as well as their negotiating and signalling behaviours, so these must be replicated by the AV.

This Part I begins at the lowest levels of machine vision with sensing (SAE level 0) and detection (SAE level 1), and considers recognition and tracking (SAE level 2) based on them. This Part I is intended to lay the foundations for Part II [29], which then moves up the technology stack to consider SAE levels 3-5. Part II also reviews data sources and other experimental resources useful for building and testing models at all levels.

Pedestrians are here defined as humans moving on and near public highways including roads and pedestrianised areas, who walk using their own locomotive power. This excludes, for example, humans moving on cycles, wheelchairs and other mobility devices, skates and skateboards, or those transported by other humans. This review does not cover interactions of traffic participants without pedestrians: a survey on trajectory prediction of on-road vehicles is provided in [133] and a survey on vision-based trajectory learning is provided in [154].

The organization of the review serves as a new taxonomy from relatively well understood quantitative engineering methods at the lower levels, towards less clear qualitative psychological theories of behaviour and interaction. It summarizes some progress in translating these qualitative concepts into simple quantitative statistical models, and identifies a strong need for this process towards quantifying psychological, social, group and interactive models into algorithms for real-world AV control. Each section has an introduction and discussion, which should be readable by researchers from other, especially neighbouring, fields who would like to get an overview of the state of the art and consider how their own field could connect both conceptually and computationally to it. Statistics on included papers are shown in the supplementary material Sect. 1. The remainder of this Part I is organized as shown in Fig. 3.2.

2 Pedestrian Sensing

Any pedestrian modelling system must begin by collecting sensor data about pedestrians. Detection, tracking and higher-level models may all depend on what information is present at this low-level, so a brief review is provided here. More details on automotive sensors are available in [85]. We classify our review into passive and active sensors. Active sensors actively send pulses into the environment that are reflected and detected while passive sensors detect physical phenomena already present in the environment. A summary of common AV sensors with their range and accuracy is provided in Table 2.2.

2.1 Passive Sensors

Manual Detection and Labelling The most basic method of sensing pedestrians is to use human perception, which is often used in offline studies, such as for conducting on-street surveys or annotate recordings of such surveys made with other sensors [30, 28]. Humans still have advantages over automated systems since they can use their full intelligence to subjectively annotate otherwise difficult events, such as the meanings of body language, emotions, and gestures. In particular, manual detection of pedestrians is needed and used as ground truth data for machine learning algorithms as in [247] where human experts were asked to detect people as a baseline for a comparison against machine algorithms.

Video Cameras One of the most commonly used sensors is the video camera, because it is cheap and easy to install. For example, [75] proposed a survey and experiments on pedestrian detection using monocular cameras. In [253] the shadow of moving objects is removed from the foreground images in order to improve the accuracy of the detection. In [107], shadows are automatically removed from the images in HSV color space. On the contrary, Wang and Yagi [226] treated shadow as helpful information for their appearance-based pedestrian detector.

Stereo Pair Video Cameras Traditionally, 3D machine vision was a less-developed research field than 2D image processing [102]. It uses two (or more) images from cameras, placed some distance apart, to estimate the stereo disparity between them and, ultimately, the distance in 3D space. Disparity describes the difference in location of corresponding features seen by the left and right cameras [212, ch. 11]. Disparity estimation methods fall into two classes: pixel-based methods (similar to optical flow), which estimates disparity at each pixel based on colour similarity to its neighbours; feature-based methods, which find a smaller number of statistically *interesting* points in the image (such as corners) and compute only their disparities. In recent years, these algorithms have become standard and very fast hardware implementations have enabled both real-time use and integration into consumer-style camera products [112]. Hence, it is now possible to consider a stereo camera as a single device at the sensor level for detecting humans. For example, in [117], pedestrians are detected using dense (i.e. pixel-based) stereo camera images. Ess *et al.* [76], instead, implemented a stereo vision-based detection algorithm that extracts visual features and performs pedestrian detection from a mobile platform.

Passive Infrared Imaging Pedestrians' bodies radiate heat in the infrared (IR) spectrum, which may be easier to detect than the visible one. For example, Xu *et al.* [82] developed a pedestrian detection and tracking method using a night-vision camera. [209] proposed a pedestrian detection method using infrared images. Cielniak *et al.* [48] presented a technique that combines color and thermal vision sensors data to track multiple people. Unlike visible light, IR does not allow to distinguish a single body from a group of pedestrians, but this technology can be useful for detecting and identifying objects in foggy conditions [143].

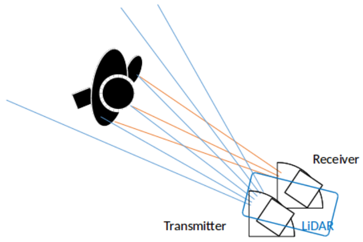
Passive Ultrasonic Sensors When a moving object enters and then leaves the detection area, the sound energy increases and then decreases: the role of a passive ultrasonic sensor is to measure the produced acoustic energy [72]. This technique is not very reliable, as it might not be able to detect single moving objects from groups, and it is also dependent on weather conditions.

Piezoelectric Sensors A review on tactile sensor detection of humans is provided in [218]. Piezoelectric sensors generate an electric impulse on touch contact, such as pedestrians stepping onto a sensed ground region, or making contact with an AV itself. This can become very expensive because it requires the installation of many piezoelectric sensors in the study area, for instance on the floor of the pedestrian infrastructure. It is useful as a last-resort sensor to detect actual collisions when other sensors have failed. In some limited (small but very high density) environments, it may be useful to monitor pedestrian movements around a sensor-filled floor, e.g. in a heavily pedestrianized area shared with last mile robots.

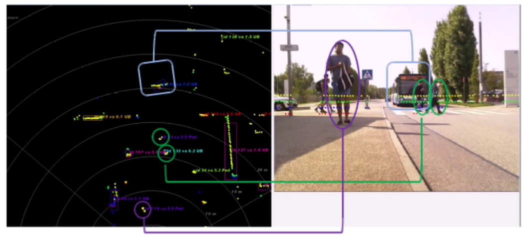
ID Sensors These devices are attached to or carried by pedestrians and they transmit unique identifying tags as well as simplifying localisation, and include infrared and RFID (Radio-Frequency IDentification) badges. Schulz *et al.* [198] developed a tracking system which combines ID sensor information with anonymous ones, such as lidar (see Sect. 2.2), in order to improve tracking accuracy. Versichele *et al.* [223] proposed to use Bluetooth for person tracking based on unique MAC (Media Access Control) addresses emitted continually by many personal devices already carried by pedestrians, such as mobile phones. In [94], camera images are fused with an omnidirectional RFID detection system using a particle filter in order to enable a mobile robot to track people in crowded environments.

2.2 Active Sensors

Lidar (Light Imaging Detection And Ranging) This sensor is mainly used for localisation and detection of traffic participants, such as pedestrians, cars, bicycles, etc. It makes use of laser beams and calculates the distance to obstacles (objects, walls, people) by measuring the time gap between sending and receiving impulses; some lidar have a 360 degrees detection range. It can be used to determine the direction, speed and trajectory of moving objects. For instance, Dewan *et al.* [67] presented a model-free detection and tracking of dynamic objects with 3D lidar data in complex environments. Objects are detected and segmented thanks to multiple motion cues, then their estimated motion model is used for tracking. Arras *et al.* [6] proposed a similar supervised classifier to detect people using a 2D lidar. In this case, AdaBoost (Adaptive Boosting), a binary boosting algorithm that combines a set of weak classifiers into a strong classifier, is used to detect features of the laser beams corresponding to peoples' legs in different environments. Gonzalez *et al.* [97] combined lidar and RGB camera data for pedestrian detection. Lidars can be used in any weather conditions, but they can be quite expensive, especially when a range of more than 30m is needed [15]. Fig. 2.3a shows the working principle of lidar and Fig. 2.3b shows the detection of road users using a lidar.



(a) The working principle of a lidar



(b) Detection of road users with a 2D lidar

Figure 2.3: The working principle of a lidar and its detection of road users.

Radar (Radio Detection And Ranging) This sensor was first used during World War II. Radars emit a radiation from their antenna, which receives back the radiation reflected by passing objects. There are two types of radar: one which transmits a continuous wave of constant frequency to determine the speed of moving objects based on the Doppler principle, where objects with no relative motion are not detected [122]. The second type, frequency modulated continuous wave (FMCW), transmits a continuous changing frequency, which can detect static and moving objects [45].

Active Infrared Sensors These sensors are composed of a transmitter that emits infrared light, a receiver that captures the reflected light, and a data collection unit that measures the time of flight of the emitted infrared light. Objects' speed can be detected by sending over two or more beams of infrared light. Their range varies from a few to tens of meters. The Kinect sensor [250], a popular RGBD (red, green, blue, depth) camera, is a particular example of an active infrared sensor. It uses a complex known pattern of thousands of rays and measures their movement in the reflected image to infer distance, similarly to a lidar. A review of computer vision techniques based on the Kinect sensor is proposed in [101].

Active Ultrasonic Sensors They emit sound waves and a detector senses the sound waves reflected by passing objects. This low-cost sensing method is immune to lighting conditions and does not require significant maintenance. However, it can be seriously affected by weather conditions and it is typically not accurate enough in certain areas [36].

Table 2.2: Range and Accuracy for common AV sensors.

SENSOR	RANGE	ACCURACY
STEREO CAMERAS	From 0.5m up to several tens of meters [19]	Disparity error of 1/10 pixel (correspond to about 1m distance error if the object is 100m far away) [169]
INFRARED	From a few cm to several meters [85, 108]	Temperature accuracy of +/-1°C, can measure temperatures up to 3,000°C [85]
ULTRASONIC	From 2cm to 500cm [36, 195]	About 0.3mm [36, 195]
RFID	Several meters [256, 84]	A few centimeters [256, 84]
LIDAR	Up to 300m [251, 193]	Up to 2cm[62, 193]
RADAR	• Short range: 40m, angle 130° [160, 103, 100] • Middle range: 70m to 100m, angle 90° [160, 103] • Long range automotive radar from less than 1m to up to 300m (opening angle up to +/-30°, a relative velocity range of up to +/-260km/h) [62, 160, 204]	• Short range: Less than 0.15m or 1% [160, 103, 100] • Middle range: Less than 0.3m or 1% [160, 103] • Long range: 0.1m e.g. Bosch LRR3 77 GHz, range 250m [62]

2.3 Discussion

Most autonomous vehicles today are using a mix of lidar, radar, and stereo vision. Visual RGB images are most commonly used as the base for detection, and feature-based localisation and mapping. Lidar or radar provide more reliable, but more expensive sensing capabilities for safety-critical aspects such as collision avoidance. While stereo cameras and radar are already used in commercially-available vehicles – for example in lane departure and adaptive cruise control systems, respectively – we expect that lidars will be used as well due to expected drops in prices. In recent years, lidar has been the main source of point cloud localisation and mapping in high-precision sensing for research work, but developments in millimeter radar and stereo cameras are making them increasingly competitive for this purpose. Manual annotation of image data remains necessary for recognition of difficult detailed features such as pedestrian eye contact and body language meanings, but for other tasks even including the creation of training sets for machine learning, is now replaced by automated methods, including semi-supervised approaches which allow quite small manual training sets to be bootstrapped with much larger unannotated data.

3 Pedestrian Detection Models

A previous review of pedestrian detection is presented in [71]. Here we summarize some of the key detection methods that are particularly relevant to AVs. Different techniques are used for detection, which can be classified into six main categories: visual appearance-based detection, motion-based detection, spatio-temporal feature detection, 3D feature detection models, deep learning methods and attention-windows detection. In computer vision, the detection problem can be viewed as a special case of image classification: given a candidate image window, the detection seeks to classify the latter as a pedestrian or non-pedestrian. The same concept applies to other types of sensors with their own detection windows. Fig 2.4 summarizes the sensing technologies and the pedestrian detection techniques described in this section.

3.1 Visual Appearance-Based Detection

Unlike motion-based methods, feature-based methods can operate with a single still image, as they look only for static patterns rather than changes over time.

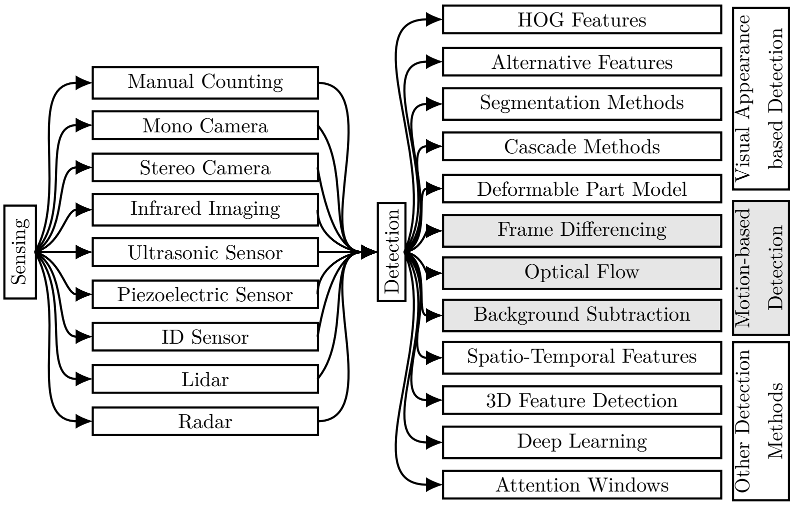
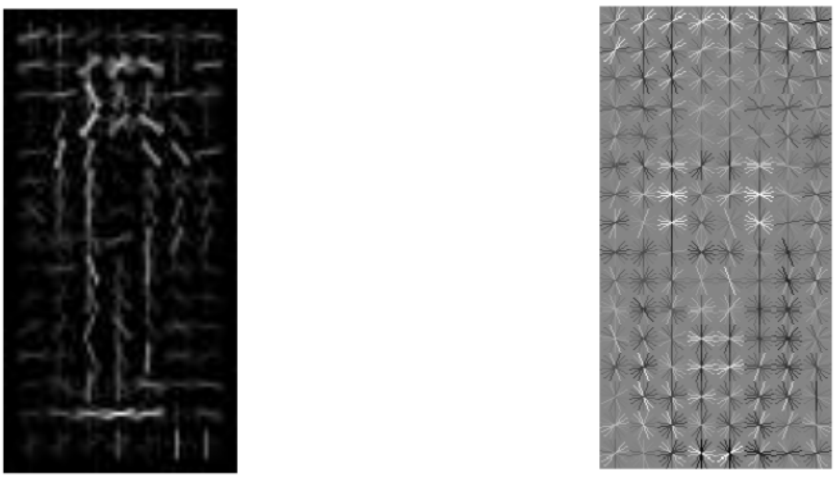


Figure 2.4: Pedestrian sensing and detection techniques.



(a) Pedestrian (b) Face

Figure 2.5: Examples of HOG features [60].

HOG-SVM One of the most commonly used pedestrian detectors is based on the combination of HOG (Histogram of Oriented Gradients) and SVM (Support Vector Machine). HOG [60] is a technique that was invented for the purpose of human detection. After training, a classifier can determine whether a proposed HOG corresponds to a pedestrian or not (Fig. 2.5). The OpenCV vision library [24] has a generic implementation of an object detector based on this method, which can be applied to pedestrian detection.

Alternative Features Sometimes used in place of HOG, alternative features including point descriptors, e.g. BRISK (Binary Robust Invariant Scalable Keypoints) and SIFT (Scale Invariant Feature Transform), are used to detect characteristic features of an image, such as corners or edges [192] [20]. Other forms of gradient features and edge detectors [33] are less sensitive to illumination compared to color descriptors. Texture features, such as Local Binary Patterns (LBP), assign a class to each local window. Groups of classes in nearby windows can then be classified as pedestrians or non-pedestrians. For example, [3] proposed a face recognition method based on the LBP feature descriptor. [163] used LBP with spatial pooling for a robust pedestrian detection.

Cascade-based Detection The detector proposed by [224] is composed of a sequence of classifiers, trained using Haar-like visual features, where each classifier can pass or not a sub-region to the following one. Zhu *et al.* [258] proposed a person detection method using a cascade (40 levels) of HOG-SVM detectors combined with Adaboost for feature selection. In [41], Chen *et al.* developed a person detection approach using a cascade classifier based on Adaboost with rectangle features and edge orientation histogram (EOH) features.

Segmentation Methods These include methods such as the Mean-Shift clustering [27], watershed, and grab-cut, which divide the image into regions typically having similar or smoothly changing colour and texture characteristics. These regions can then be tested directly for pedestrians presence through shape, texture and other statistics as in [188], where people were detected and segmented based on a probabilistic method that describes the shapes of their different postures.

Deformable Part Model Deformable Part Model (DPM) is a popular detection model. It has been originally proposed for the Pascal VOC challenge for object (including pedestrian) detection and recognition [77]. DPM splits an object into several parts arranged in a deformable configuration and can be used for pedestrian classification as in [79]. This method can deal with significant variations in shape and appearance. A fast implementation of DPM applicable for person detection is proposed in [233].

3.2 Motion-based Detection

Frame Differencing This method consists in computing the difference between the current frame and a reference one (usually the first frame). In [74], a person detector was developed using optical flow computed on regions selected by frame differencing on camera data recorded from a vehicle. Selected regions are then passed to a wavelet-based features classifier combined with template matching. Park *et al.* [165] proposed an approach that uses coarse-scale optical flow to stabilize camera frames with temporal difference features for pedestrian detection and human pose estimation, and tested on the Caltech pedestrian benchmark [70].

Optical Flow This technique assigns a direction and a velocity of motion to each pixel of two consecutive frames, as in [225]. Fernández-Caballero *et al.* [83] used optical flow and frame differencing for human detection on infrared camera images for a security mobile robot platform. Another use of optical flow for detection and tracking is proposed in [67] using 3D lidar data.

Background Subtraction This method builds a background model used as a reference model in order to detect moving objects. This modelling is based on the assumption that the background is static. It consists in extracting an estimate of the background from the rest of the image by using some methods such as mean filter, running Gaussian average, etc. Background modelling has two variants: the recursive algorithm, which updates each frame with the estimate of the background, and the non-recursive algorithm, which stores a buffer with the previous frames and the background estimated from them. In [201], Sheikh *et al.* developed a background subtraction model that can detect humans and objects in moving camera images. Their method builds background and foreground appearance models based on the background trajectory estimated by a RANSAC algorithm.

3.3 Other Detection Models

Spatio-Temporal Features These are commonly used in video codecs, such as Theora and H.264, because they are statistically efficient summary descriptors of natural video. As such, they are also candidates for informative classification features. Oneata *et al.* [162] used these features with a supervoxel method for human detection in videos.

3D Feature Detection These models rely on 3D sensors, such as depth cameras and 3D lidars. Depth information enables more robust detection algorithms. For example, the authors in [234] proposed an online learning method based on a 3D lidar cluster detector, a multi-target tracker, a human classifier and a sample generator. The cluster detection starts by removing the ground plane, then point clusters are extracted from the point clouds using the Euclidean distance in 3D space and finally a human-like volumetric model is fitted to the clusters for filtering. Yan *et al.* [235] took advantage of multiple (2D and 3D) sensor detectors to train an online semi-supervised human classifier for a mobile service robot. A depth-based person detector is presented in [151]. This detector applies template matching on depth images. To reduce the computational load, the detector first runs a ground plane estimation to determine a region of interest, which is the most suitable to detect the upper bodies of a standing or walking person. In [58], a mobile robot equipped with an RGB-D camera is used to detect people. Munaro and Menegatti [156] proposed a real-time detection and tracking system based on RGB-D camera data capable of detecting people within groups or standing near walls.

Attention Windows In their basic forms, the classifier-based detection methods above may assume that every possible location and size window of a 2D or 3D image will be tested for pedestrian detection. Such 'sliding windows' can be computationally slow, unless the tests are performed in parallel (e.g. on a GPU) or some form of attention model is used to restrict the search. It is common to use a simple, fast, and inaccurate detector set to have many false positives and few false negatives, to decide whether a window should be explored further or not [200]. In this case, a more advanced but slower method would be applied to test the most interesting windows. Prokhorov [173], for example, developed a road obstacle detector based on attention windows with potential application to pedestrian detection.

Neural networks ('deep learning') Neural networks [98] are hierarchical-in-the-parameters regression models which seek to minimise an error function E between N desired vector outputs $c^{(n)}$ for $n \in \{0, N-1\}$ and a function F of input vectors $x^{(n)}$ (including an element which is always 1) with parameters θ,

$$E = \sum_n \|c^{(n)} - F(x^{(n)};\theta)\|^2, \tag{2.1}$$

where F is comprised of layers of 'node' functions,

$$y_j = f(a_j), \quad a_j = \sum_i w_{ji} y_i, \tag{2.2}$$

and f is any nonlinear function, $w_{ji} \in \theta$ are weights from any node i in a lower layer to any node j in the layer above it, and y_i for the lowest layer are elements of the input vector $x_i^{(n)}$. The vector formed from y_l for all nodes l in the top layer is the value of F. E is then locally minimised by computing *backpropagation* terms Δ_i for each node,

$$\Delta_i = f'(a_i) \sum_j \Delta_j w_{ji}, \tag{2.3}$$

beginning by setting for the top layer nodes l,

$$\Delta_l = c_l^{(n)} - F(x^{(n)}; \theta)_l, \tag{2.4}$$

then updating the parameters w_{ji} along the direction,

$$-\frac{\delta E}{\delta w_{ji}} = -\Delta_j y_i. \tag{2.5}$$

Neural networks date from at least the 1970s [229], but have returned to popularity due to falls in prices of parallel hardware (specifically, graphics cards) which has enabled the use of 'deep' networks having more layers; together with the algorithmic improvements of sharing weights (convolutional neural networks, CNN), pooling [130] and dropout [125] which exploit statistical regularities found in most natural data.

The classifier-based detectors presented so far rely on a two-stage process of feature extraction followed by classification. Neural networks can be used in this way as classifiers given input vectors of features. But increased computing power now enables the raw image to be given directly as input to neural networks having more layers, which can learn their own feature sets in the lower layers, enabling features to be learned, rather than manually chosen, to optimise performance in specific tasks. For example, [5] proposed a real-time pedestrian detector using 'deep network cascades'.

Like other classifier-based detectors, neural networks themselves only learn a mapping from input to output vectors, so to apply them to *detection* of objects in images, some scheme like the attention windows of section 3.3 is needed to propose regions of interest. R-CNN [96] computes region proposals with any non-neural method such as 'selective search'. It computes features for each proposal region using a large CNN, then classifies these features sets using class-specific linear SVMs and also uses linear regression to refine the region from the features. Faster R-CNN [181] extends a CNN with layers for region proposals and layers for classification, using them to propose then classify regions. YOLO [178, 177, 179] similarly extends a CNN with layers for both region proposal and classification, but runs them at the same time with classification based on approximate rather than finally proposed regions. It is able to detect about twenty different classes such as people, cars, bicycles and trucks in real time video. Mask R-CNN [104] finds segmentations as well as rectangular regions, by extending Faster R-CNN with layers predicting masks for regions.

3.4 Discussion

Traditionally, a wide variety of image features have been developed by hand and matched with a wide variety of classifiers, to find good performance in pedestrian detection. Until recently, the HOG-SVM method was the best known [16]. Pedestrian detection, like most classification tasks, has however recently been revolutionized by price falls in parallel hardware such as GPUs, which have enabled classical neural network algorithms with small modifications ('deep learning') to outperform hand-crafted methods for the first time. It seems likely that neural network methods will completely replace all others. The same GPU hardware also enables pixel-wise algorithms, such as optical flow, to be massively accelerated. They might not be necessary though if neural networks alone achieve the required performance.

The implementation of a person detection method for an AV is one of the major practical challenges. OpenCV[1] library provides open-source implementation of many computer-vision algorithms (in C++ and Python), mainly aimed at real-time processing. It contains feature extraction methods such as HOG, SIFT, BRISK. It also includes a C++ implementation of DPM. In addition, LibSVM[2] is a popular implementation of SVM classification algorithm. The lidar-based leg detector in [6] is implemented as a Robot Operating System (ROS) module[3]. Again, the ROS implementation of the depth-based detector in [151] is available[4]. In addition, an offline version of the 3D lidar-based approach in [234] is implemented as a ROS module[5]. The authors of the RGBD-based detector in [156] provide the implementation of their algorithm[6]. Many DL-based approaches provide their code for reproducibility and comparison: YOLO[7], R-CNN[8], Faster R-CNN[9] and Mask R-CNN[10].

High performance of deep learning models comes at a price: they require larger training data (sometimes several millions of examples), longer training times (up to several days), and their computational cost is more important than for simpler detectors [248]. In some cases, DL methods cannot reach real-time performance [5] and are outperformed by simpler

[1] https://opencv.org/
[2] https://www.csie.ntu.edu.tw/~cjlin/libsvm/
[3] https://github.com/wg-perception/people
[4] https://github.com/strands-project/strands_perception_people/
[5] https://github.com/LCAS/FLOBOT
[6] http://pointclouds.org/documentation/tutorials/ground_based_rgbd_people_detection.php
[7] https://pjreddie.com/darknet/yolo/
[8] https://github.com/rbgirshick/rcnn
[9] https://github.com/rbgirshick/py-faster-rcnn
[10] https://github.com/facebookresearch/Detectron

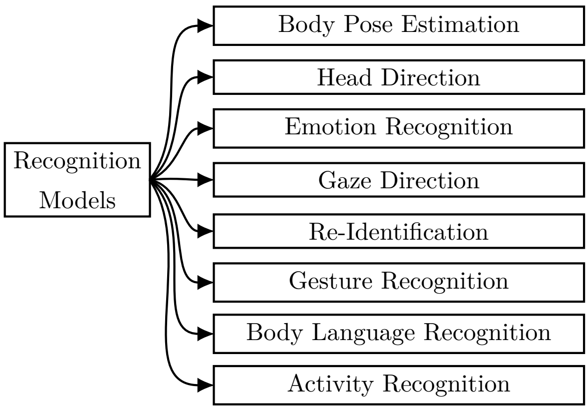


Figure 2.6: Pedestrian attributes for recognition models.

methods such as HOG [221].

4 Pedestrian Recognition Models

While detection refers to finding the presence or absence of pedestrians at locations and scales in images, *recognition* here refers to the recognition of attributes of pedestrians given by such detections. Recognition takes as input the localised window of visual or other sensor data forming the detection, and yields as output some information about the particular pedestrian detection. In some cases, this could include their actual identity – identity recognition – but our use of the term here also includes recognition of attributes such as their body pose and facial features. Recognition refers to these *tasks*, while *classification* here refers to processes that perform recognition specifically by mapping inputs into discrete rather than continuous output classes. Figure 2.6 presents a set of attributes used for pedestrian recognition and a summary of the recognition models and papers reviewed in this paper is given in the supplementary material Sect. 2.

4.1 Recognition of Body Pose

While full-body tracking is discussed below, some methods may attempt to classify from single images some basic information on pose, such as the head direction of the pedestrian into facing AV/not facing AV. Where the pedestrian body state is known – as resulting from skeleton and other body tracking – it may contain useful information about pedestrians' goals and intentions, which may be extracted by classifiers operating at a higher-level – on the tracked body configurations rather than on the raw images or other sensor data. Cao *et al.* [34, 35] presented OpenPose, a real-time multi-person pose estimation software that uses CNNs to detect people in 2D images and Part Affinity Fields (PAF) is used to associate body parts to the detected people. Shotton *et al.* [203] developed 3D human pose estimation based on body parts representation. Their method relies on depth features, randomized decision trees and forest algorithms for classification, and outputs a proposal position for each detected body part. The method was tested on motion capture and synthetic data.

Iqbal *et al.* [110] proposed a graphical model optimized by a integer linear programming (ILP) to estimate and track multiple people in videos; the used data is made available as a new dataset called PoseTrack. Tompson *et al.* [220] combined a deep CNN with a Markov Random Field to estimate human pose from monocular images. Fragkiadaki *et al.* [89] proposed a method using recurrent neural networks with an Encoder-Recurrent-Decoder (ERD) architecture to predict body joint displacements. ERD is an extension of LSTMs. Martinez *et al.* [144] proposed a method using RNN with Gated Recurrent Unit (GRU) architecture without requiring a spatial encoding layer and allows to train a single model on the whole human body. Tang *et al.* [214] proposed a model that extends the work in [89] and [144]. Their work is based on the observation of human skeleton sequences and uses deep neural networks (Modified High-way Unit (MHU)) to remove motionless joints, estimate next moves and perform human motion transfer. Gosh *et al.* [95] used a Dropout Autoencoder LSTM (DAE-LSTM) to extract structural and temporal dependencies from human skeleton data. Manual annotations are not needed because a tracker gives the actual direction of movement. Kohari *et al.* [123] used a CNN model to estimate human body orientation for a service robot.

4.2 Recognition of Head Direction

The primary use of extracting the head direction in pedestrian-AV interaction is epistemological: a pedestrian facing the AV – and/or establishing direct eye contact with it – is a good indicator that the pedestrian has seen the AV and knows it is there, and therefore will be planning their own behaviors on the assumption that they will have to interact with it. In contrast, a pedestrian who has not seen the AV, unless relying on auditory cues, may just step into the road with no idea that a potentially dangerous interaction is about to occur [230] [12]. Darrell *et al.* [61] developed a real-time human tracking and behaviour understanding system, called Pfinder. The system converts human head and hands into a statistical model of color and shape in order to deal with different viewpoints. Schulz and Stiefelhagen [197] estimated pedestrian head pose using multi-classifiers for different monocular grayscale images; depth information within the detection bounding box is also taken into account. Flohr *et al.* [86] proposed a model that can detect pedestrian body and head orientation from grayscale images based on a pictorial structure method.

4.3 Recognition of Gaze Direction and Eye Contact

Algorithms for gaze tracking and eye contact detection are not yet robust, and in laboratory eye tracking experiments require expensive precision equipment to be installed on the subjects' heads. Benfold and Reid [17] proposed a method which infers the gaze direction from a head pose detector based on HOG and colour features. The head pose is classified using randomised ferns, i.e., similar to decision trees, and the tracking is done frame-by-frame based on the head detector using multiple point features. Baltrusaitis *et al.* [9] developed the open-source OpenFace, running in real-time with a simple webcam. It is suitable for facial behavior analysis, in particular for facial landmark detections, head pose estimation, facial action recognition and eye-gaze estimation.

4.4 Emotion Recognition

Pedestrian emotion recognition might be useful to inform about their crossing intention. For example, an angry pedestrian might be more likely to behave more assertively in crossing the road in front of an AV. Cornejo *et al.* [56, 55] developed a facial expression recognition method that is robust to occlusions. The occluded facial expression is reconstructed with a robust principal component analysis (PCA) method, facial features are extracted using Gabor wavelets and geometric features in [56] and using CENTRIST features in [55], recognition is performed with KNN and SVM as classifiers. Cambria *et al.* [32] proposed a new categorization model for emotion recognition systems and [31] reviewed sentiment analysis methods. Poria *et al.* [171] developed a CNN model with a convolutional recurrent multiple kernel learning that can extract features from multimedia data such as audio, videos, and text. The method has been tested on Youtube videos and ICT-MMMO dataset. Den Uyl and Van Kuilenburg [65] developed the FaceReader, an online facial expression recognition system, which is robust to the head pose, orientation and lighting conditions.

4.5 Recognition of Pedestrian Identity for Re-Identification

Person re-identification (re-ID) is the problem of recovering the identify[1] of the same person with different clothing across different images, under different camera views, weather, lighting, and other environmental conditions. Ahmed *et al.* [2] developed a deep convolutional network that solves the re-identification problem by computing a similarity value between two image pairs. Their method has been tested on CUHK01, CUHK03 and VIPER datasets. Zheng *et al.* [252] proposed a person re-identification method based on the Bag-of-Words (BoW) model which extracts Color Names (CN) descriptor features from the input image, a Multiple Assignment (MA) is then used to find neighboring local features and finally TF-IDF finds the number of occurrences of visual words. Their method was tested on the Market1501 dataset. In [254], a CNN model with unlabeled images is used to re-identify people. Li *et al.* [134] proposed a filter pairing neural network (FPNN) model for person re-identification, capable of handling challenging conditions such as occlusions.

4.6 Gesture Recognition

Deliberate gestures are the most obvious form of communication from body pose. For example, a pedestrian may wave a vehicle on to show that they intend to give it priority in a crossing. A previous review on hand gesture recognition is provided in [150] and more recently Rautaray and Agrawal [176] presented a survey for interactions with a computer. Chen *et al.* [40] used a real-time tracker with hidden Markov models (HMM) to recognize hand gestures. Freeman and Roth [90] used orientation histograms for gesture recognition. Their real-time method can recognize about 10 different hand gestures. Ren *et al.* [182] developed a robust hand gesture recognition system for active infrared (Kinect) sensors. Their method is based on template matching for part-hand gesture recognition and a new distance metric called Finger-Earth Mover's Distance (FEMD) is used to measure the similarity between two hand shapes. Other gesture recognition methods based on HMMs are proposed in [23] [132].

4.7 Body Language Recognition

In addition to deliberate gestures, unconscious body language, including stance and gait (walking style), may also be a predictor of pedestrian assertiveness in interactions, and of other behaviours. As with gesture recognition, body language recognition relies on recognition

[1]Identity here is distinct from 'personal information' as defined by privacy laws such as the EU General Data Protection Regulation (GDPR).

of body pose, followed by classification of this pose. Quintero *et al.* [174] proposed a hidden Markov model for pedestrian intention recognition based on 3D positions and joint displacements along the pedestrian body. In [227], a human gait recognition method is proposed, combining background subtraction with PCA for dimensionality reduction. A supervised pattern classification is finally performed to recognize the gait.

4.8 Activity Recognition

Pedestrian activity recognition is of particular importance for autonomous vehicles. A lot of work is ongoing for service robots and AVs. A more complete review on human activity recognition methods is proposed in [68]. Chaaraoui *et al.* [37] used contour points of human silhouette to recognize human actions for real-time scenarios. Doll'ar *et al.* [69] used spatio-temporal features for both human and rodent behaviour recognition. Vail *et al.* [222] compared hidden Markov models to conditional random fields for human activity recognition. In [138], a coupled conditional random field is used with RGB and depth sequential information. Coppola *et al.* [54] developed one of the first RGBD-based social activity recognition methods for multiple people. Their method learns spatio-temporal features from skeleton data, which are fused using a probabilistic ensemble of classifiers called Dynamic Bayesian Mixture Model (DBMM).

4.9 Discussion

AVs need to recognize pedestrian attributes including pose and possibly identity to help them make more accurate predictions about pedestrians' likely future behaviours. Detection of pedestrians is now mature technology, but recognizing the attributes of these pedestrians within these detections, such as body pose, is a harder and still open research area. Eye direction and eye contact remain particularly difficult as it requires very precise estimation of the positions of small pupils and irises at a long distance. Humans have evolved to be particularly good at recognizing gaze direction for social purposes, but it is hard to replicate. Recognition of emotions may be useful to inform predictions of pedestrians' likely behaviours (e.g. an angry pedestrian may be more likely to push in front of us), and progress has been made in this area in non-real time systems, such as social networks' processing of photographs. But again, recognition from far distances and speeds travelled by AVs for real-time encounters remains challenging and open. It is likely, in the future, that neural network approaches will come to dominate this area as with detection.

Open-source implementations of pedestrian recognition models include Openpose[1] for pose estimation, OpenFace[2] for head pose and eye-gaze estimation and OpenTrack[3] for head tracking. To our knowledge, there is no generally accepted benchmark for pedestrian recognition models. Future research should thus explore the performance and computational efficiency of pedestrian recognition models in the context of autonomous driving.

Recognition of any attribute which enables recovery of a pedestrian's name or other formal identification will fall under data protection laws in most jurisdictions, such as the GDPR across the EU. While re-identification (re-ID) might be particularly useful, for example for use in delivery robots to confirm recipients' identities, the usage of this technology should be carefully assessed with respect to data privacy. The other recognition and tracking algorithms mentioned in this section extract features anonymously, i.e., extracted data does not allow the identification of individuals. Re-ID on the other hand can be used to record and store sensitive personal data, which yields the potential to be misused for public surveillance. For AVs, centralized re-ID might be useful to link individual traffic participants to their previously-observed behavior in traffic enhancing long-term path prediction, but at the cost of severe intrusion into the privacy of road users. This will raise a host of ethical and legal issues when such accuracy is reached by rapidly accelerating machine vision research, such as selling data of individual's locations and behaviours to insurance and advertising companies, or use by local authorities or law enforcement agencies [88].

5 Pedestrian Tracking Models

Pedestrian tracking is the process of updating the belief about a pedestrian's location from a temporal sequence of data. More specifically, tracking consists in determining the position and possibly orientation or velocity of a given object over time. A pedestrian track is a sequence of their locations over time. A pedestrian pose track is a sequence of a pedestrian's body pose states over time. When multiple pedestrians are present, tracking requires separating the pedestrians from each other and associating the identities of the pedestrians with tracks. This is a challenging problem for humans if their tracks overlap or disappear behind obstacles, and appears to require high-level social intelligence and knowledge to guess what most likely happened when tracks are temporarily hidden.

[1] https://github.com/CMU-Perceptual-Computing-Lab/openpose
[2] https://cmusatyalab.github.io/openface/
[3] https://github.com/opentrack/opentrack

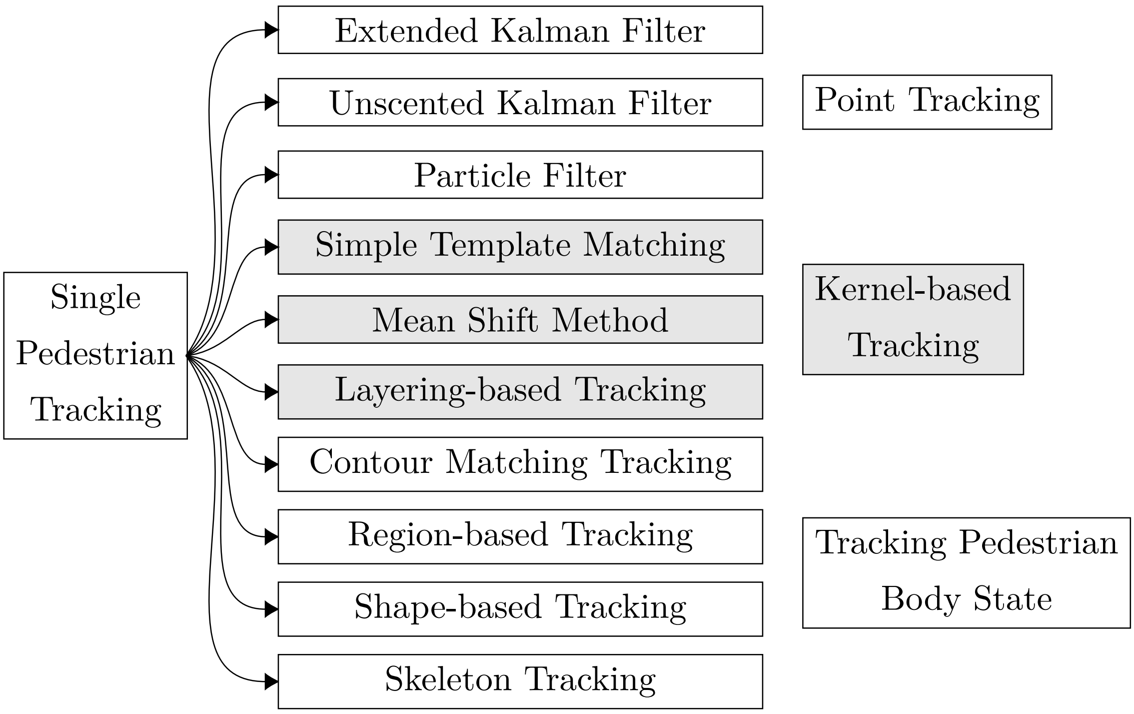


Figure 2.7: Single pedestrian tracking models.

Pedestrian tracking consists of two steps: (1) a prediction step to determine several likely next possible pedestrian states, (2) a correction step to check each of these predictions and select the best one. It often requires the estimation of non-linear, non-Gaussian problems due to the nature of human motion, pedestrian sizes, and posture changes [15]. Pedestrian tracking is a challenge for AVs because of the multiple uncertainties (e.g. occlusions) originating from complex environments. Many techniques have been employed for pedestrian tracking, see e.g. [239, 206]. Bar-Shalom *et al.* [11] presented state estimation algorithms and how they could be applied to tracking and navigation problems. Figure 2.7 summarizes single pedestrian tracking models.

Previous reviews on tracking methods for pedestrians can be found in [239, 152]. In this section, we first review two classes of methods for single pedestrian location tracking relevant to autonomous vehicle interactions (as previously classified by Yilmaz *et al.* [239]): point tracking and kernel-based tracking. We then review recent work in the more challenging tasks of body pose tracking and multiple pedestrian tracking. A summary of the tracking methods and papers reviewed in this Part I is provided in the supplementary material Sect. 3.

5.1 Single Pedestrian Point Tracking

Point tracking typically relies on probabilistic methods based on Bayes filtering [43, 191, 208]. Based on Bayes rule (2.6), the filter is composed of an initial state, a prediction step and a correction step. The initial state x_0 (2.7) presents the initial belief about the state x. The prediction step (2.8) consists in updating the belief using information about how the target typically moves around. Finally, the correction step (2.9) updates the state estimate with sensor measurements z, to give posterior beliefs $bel(x_t)$ about the state at each discrete time t, with a normalizer η, [186, 219].

$$p(x_t \mid z_t) = \frac{p(z_t \mid x_t)p(x_t)}{p(z_t)} \tag{2.6}$$

$$bel(x_0) = p(x_0) \tag{2.7}$$

$$\widehat{bel}(x_t) = \int p(x_t \mid x_{t-1}) \cdot bel(x_{t-1})dx_{t-1} \tag{2.8}$$

$$bel(x_t) = \eta \cdot p(z_t \mid x_t) \cdot \widehat{bel}(x_t) \tag{2.9}$$

The transition probability $p(x_t|x_{t-1})$ is of crucial interest as it provides the mathematical bridge from low to high-level pedestrian behavior models. In its lowest form – the standard Kalman filter – it may simply be a Gaussian with zero mean and variance set to model the scale of a (literal) random walk by the pedestrian. But we may have much more predictive information θ about the pedestrian behavior to form $p(x_t|x_{t-1}, \theta)$. Here θ could include mid-level information such as the pedestrian's pose, heading, and location on a map. For example, if the pedestrian is standing at the edge of the road, he/she is more likely to wait and cross. Information about the pedestrian's origin and destination could also help to predict the future trajectory. Further information about beliefs, intentions and desires of the pedestrian will also modify the trajectory probability. The transition probability thus provides the interface where all higher-level models, discussed later in Part II [29], will link to low-level pedestrian models. The following are some of the most popular variants of Bayesian Filtering used for pedestrian point tracking:

Kalman Filter (KF) A KF is a Bayes filter applied to linear systems with continuous states and Gaussian noise ϵ_t,

$$x_t = A_t x_{t-1} + B_t u_t + \epsilon_t, \tag{2.10}$$

where A_t is the system matrix and B_t is the control matrix.

The measurement probability also depends on a linear model C_t with Gaussian noise δ_t,

$$z_t = C_t x_t + \delta_t \tag{2.11}$$

where C_t is the measurement matrix.

The prediction step (control update step) increases the uncertainty in the robot's belief, while the measurement update step decreases it.

Extended Kalman Filter (EKF) An EKF is an extension of the Kalman Filter and approximates non-linear models via Taylor expansion. EKF is a tracking technique well performed in scenarios where there are few changes but it has a computational cost that could be not neglectable for large state and measurement vectors due to the linearization process, which can involve the calculation of big Jacobian matrices. One of the limitations of EKF is that the linearization decreases the accuracy of the system and therefore the pedestrian tracking performance [14]. For example, in [63], the authors try to solve this problem with a CNN detector combined to a Multi-Hypothesis Extended Kalman Filter (MHEKF) for vehicle tracking using low-resolution lidar data.

Unscented Kalman Filter (UKF) The UKF avoids the linearization problem by a second-order approximation, called the Unscented Transformation. It approximates a probability distribution with chosen weighted points called *sigma points* and estimates its mean and covariance. This leads to better performance in pedestrian tracking, as the Jacobian computation is not necessary anymore, with no or minimum increase of the computational cost [14].

Particle Filter This is a sample-based estimator widely used for pedestrian tracking, based on Monte Carlo methods [80, 145, 231]. Unlike EKF, which deals with Gaussian and linearized distributions, it performs state estimation of non-linear and non-Gaussian distributions. It represents the target distribution by a set of samples, called particles. An important step in particle filtering is the resampling, which consists in withdrawing 'weak' particles with low weights from the sample set, and increasing the number of 'strong' particles with high weights [219]. Particle Filtering demands high computation capabilities, when using many particles. A tutorial for implementing particle filters for detection and tracking purposes can be found in [7]. Moreover, Bellotto and Hu [14] evaluated different Bayesian filters, such as

EKF, UKF and Sequential Importance Resampling (SIR) particle filter, for people tracking and analysed the trade-off between performance and computational cost of each method.

5.2 Single Pedestrian Kernel-based Tracking

Simple Template Matching This is a brute force method. The goal is to compare a region of an image to a reference template image by minimizing the *sum-of-square-difference (SSD)*. For example, in [113], a template matching is proposed for real-time people tracking, which is robust to occlusions and variations of the illumination. In the approach proposed by Lipton *et al.* [137], moving objects are detected in camera images using frame differencing. By combining temporal differencing and template matching, the classified objects are then tracked in real-time on video. In [115], a feature selection method in image sequences is proposed to improve the performance of template matching tracking.

Mean Shift Method This is a visual tracking technique trying to match objects in successive frames, where each track is represented by a histogram. The histogram of the region of interest is compared to the histogram of the reference model. The technique iteratively clusters data points to the average of the neighbouring points using a kernel function, similar to k-means clustering [44]. In [52], the authors proposed a real-time object tracking using the mean-shift algorithm and the Bhattacharyya coefficient to localize the targets. This method is applied to non-rigid objects tracking observed from a moving camera. Collins [50] applied the mean-shift algorithm to 2D blob tracking and proposed a method to select the kernel scale for an efficient tracking of blobs. In particular, a difference of Gaussian (DOG) mean-shift kernel is chosen to efficiently track blobs through space.

Layering-based Tracking Layering consists in splitting an image into several layers by compensating the background motion to estimate the state of a moving object with a 2D parametric model [164]. Each layer is represented by its shape, motion, and appearance (based on intensity) [257]. For instance, in [215], the authors proposed a dynamic layering-based object tracker exploiting spatial and temporal information from its shape, motion and appearance. Their estimation is done using a Maximum-A-Posteriori (MAP) approach with the Expectation Maximisation (EM) algorithm. Layering-based trackers can handle multiple moving objects and occlusion. In [232], a layering-based method is combined with optical flow. A Bayesian framework is used to estimate the layers' appearance and a mixture model is used to segment the image into foreground/background regions. Other layering-based tracking methods applied to imaging sensors can be found in [73, 127].

5.3 Body Pose State Tracking

Tracking the whole state of a pedestrian's body – including skeleton pose, head direction, feet and walking directions – may provide useful information about the pedestrian's state and intention. These silhouette tracking methods are based on an accurate shape description of the pedestrian object. The general technique consists in finding the pedestrian region in each frame with an object model computed from the previous frames. The advantage is that it can cope with different types of shape, occlusion problems, etc.

Contour Matching Tracking Tracking is performed considering the contours of objects, which are dynamically updated in successive frames. Geiger *et al.* [93] proposed a contour tracking method that is based on Dynamic Programming (DP) to detect and track the contour of multiple shapes and provide the optimal solution to the problem. Techmer [217] developed a real-time approach to contour tracking relying on the distance transformation of contour images and tested it on real-world images. Baumberg and Hogg [13] proposed a method that combines dynamic filtering (Kalman filter) with an active shape model to track a walking pedestrian in real-time. However, this tracking technique is very sensitive to the initialization, so other solutions have been developed to overcome that issue [240].

Region-based Tracking This technique is based on the color distribution of objects. In [1], a tracking algorithm is proposed based on multiple fragments of object images, creating a histogram of the current frame that is compared to the histogram of the patches. Their method is able to handle occlusion and pose changes in an efficient manner. Other methods have employed depth, probabilistic occupancy maps and gait features to estimate a region's features, but in some cases (e.g. depth features) this requires the computation of multiple views of the same scene. Meyer and Bouthemy [146] developed a method to track objects over a sequence of images using a recursive algorithm based on image regions information, such as their position, shape and motion model.

Shape Matching Tracking Shape matching tries to match silhouettes found in two consecutive frames. Performed with Hough transform, it can handle occlusion problems. For instance, in [51], a silhouette-based model is used to identify people from their body shape and gait.

Skeleton Tracking for Body Language and Gesture Recognition Skeleton tracking, based on tracking human body parts, is a popular technique [92, 238, 196, 153]. Schwarz *et al.* [199] presented a full-body tracker using depth data from a Kinect sensor. 3D data is represented by a graph structure which can deal with variations in pose and illumination. A skeleton is then fitted to the 3D data by constrained inverse kinematics and geodesic distances between body parts. Sinthanayothin *et al.* [205] reviewed skeleton tracking methods using Kinect sensors. Make Human Community[1] is an open-source project building parametric models of humans based on realistic skeleton structures, mainly targeted at video games users, but also used as a generative machine vision and tracking model for 3D sensor data.

5.4 Multiple Pedestrian Tracking

Multiple pedestrian tracking (a form of MTT, Multi-Target Tracking) names the task of (rather than specific algorithm for) tracking the poses of several pedestrians at the same time. The pedestrians may be close, overlapping, or obstructing one another, and they may be indistinguishable from one another other than by their pose. This is required for AV interactions with multiple pedestrians, ranging from two well-separated pedestrians, to small groups of pedestrians (often crossing roads together) and to dense crowds. MTT creates a data association problem: how to know which pedestrian detection belongs to which track? A probabilistic MTT model would maintain beliefs at each time step about the state of every track and consider every possible association of detections to tracks; then, it would perform inference accordingly. However, the number of associations grows exponentially with the number of pedestrians, so this approach is unlikely to work in very crowded scenarios. Standard approximations then include making hard 'winner-take-all' assignments at each time step; maintaining search trees of recent possible assignments; and pruning association hypotheses. There are many possible variations on these approximations, all making use of basic individual-pedestrian trackers as components.

Leal-Taixé *et al.* [129] presented a benchmark for Multiple Object Tracking that was launched in 2014 and called *MOTchallenge*. This benchmark provides a framework for evaluating the performance of state-of-the-art MTT algorithms. About 50 methods have been

[1] http://www.makehumancommunity.org/

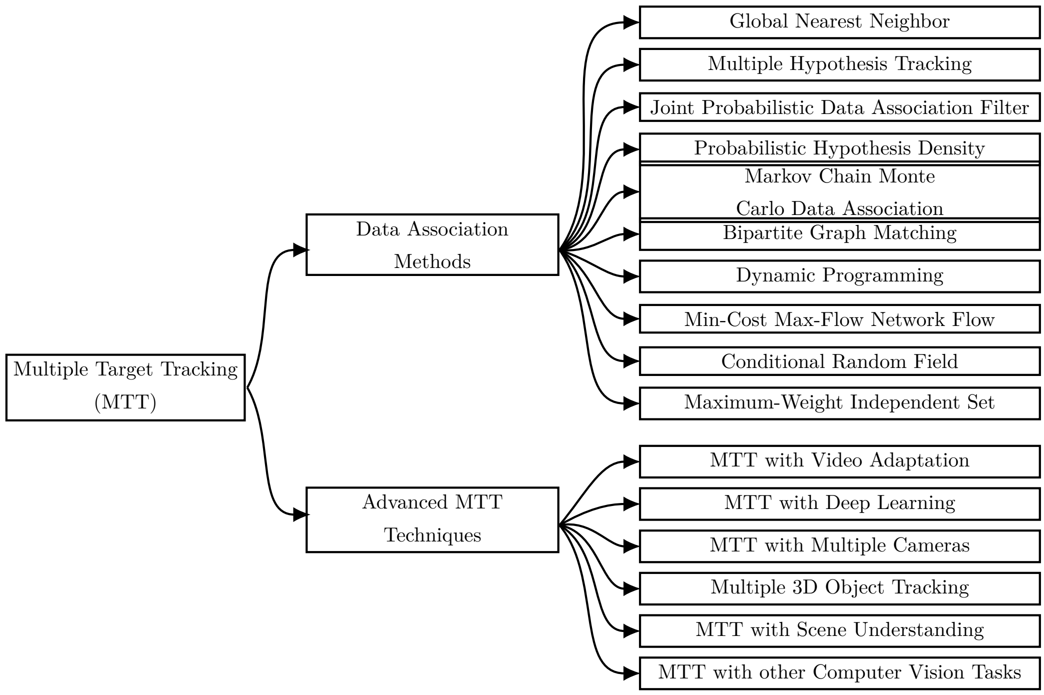


Figure 2.8: MTT data association and advanced techniques.

tested up to now on this benchmark. However, [129] does not describe these algorithms, while Fan *et al.* [78] only presents a survey on visual methods. A previous review on multiple object tracking was proposed in [142]. The remainder of this section will therefore extends their work for multiple person tracking and try to give an overview of the main methods, challenges and future directions of MTT techniques, which intelligent transportation systems heavily rely upon. Figure 2.8 summarizes the techniques described in this section.

5.4.1 Categories of MTT methods

The following paragraphs will develop the different categories of multi-target tracking methods that are defined according to their initialization method used, the processing method, or the tracking output.

Initialization Method The first category is characterised by the detection technique used before tracking. The most commonly used method is Detection-Based Tracking (DBT) where a program is trained in advance to detect the target object in the input data (e.g. images) [111]. This technique can deal with a variable number of target objects, but it cannot track unknown objects that were not part of the training. The other initialization method is Detection-Free Tracking (DFT), which requires manual initialization, i.e., an operator labels manually the target objects. In this case, the object detection is error-free but the tracking can usually only deal with a fixed number of target objects. Neiswanger *et al.* [157] proposed a method to track multiple people in video sequences without any pre-defined person detector. A Dirichlet process is used to find the clusters in the images and then a Sequential Monte Carlo (SMC) method with local Gibbs iterations and a Particle Markov Chain Monte Carlo (PMCMC) are used to infer the posterior of targets. Lin *et al.* [135] developed a detection-free multiple target tracking method which relies on video bundle representation and a spatio-temporal graphical model to infer the trajectories of people.

Processing Model This second category refers to the information processing mode: online or offline tracking. Online tracking [120] is a sequential tracking, which relies on up-to-date information. It is a causal method where only past and current observations are used. Offline tracking [118] instead uses information both from past and future observations, therefore it is not causal. In order to estimate the output, offline tracking needs to evaluate all the observations from all the frames, which requires a high computation cost. The manual assignment guarantees a tracking process free of false detections, but is not suitable for real-time applications. Both online and offline tracking methods are proposed in [242].

Tracking Output MTT methods can be grouped according to output. Output results are fixed for MTT methods relying on deterministic optimization, i.e., there is no randomness when these methods are run many different times, whereas for probabilistic optimization methods, output may vary for several trials cf. section 5.4.3.

5.4.2 Challenges of MTT Approaches

There are multiple challenges with the tracking of multiple objects. Here we summarise the most important ones.

Similarity Measurement The first problem is how to measure the similarity between objects in different frames. Different models have been proposed to deal with the *similarity measurement* between objects. The most commonly-used technique in visual tracking relies on the object's appearance, i.e., its visual features. There are local features, which can be obtained by the KLT algorithm or optical flow (if we treat each pixel as the finest local range) to get information about object motion patterns [202]. Region features are extracted from an image and represented by a bounding box. Three main types of region features exist: zero-order, first-order and up-to-second-order type. The zero-order type represents region features as color histogram or raw pixel templates. Although color is a common similarity measure, the problem is that it does not take into account the spatial layout of the object region. A first-order type uses gradient-based representations or level-set formulation to deal with region features [47]. Gradient-based representation is a robust technique because it describes well the shape of the object and it is less dependent to illumination conditions, but it cannot handle occlusion problems. An up-to-second-order type computes region covariance matrices to model the observed features [172]. This is a robust strategy but it requires a high computation capability.

Track Identification The second problem consists in recovering the identity of objects from the similarity measurement across frames. Different strategies *compute the similarity* between objects. A survey on similarity measures for probability density functions is provided in [249]. In case of a single cue, a distance measure is computed from two color histograms and then transformed into similarity using the exponential function or an affinity measure such as the Normalized Cross Correlation (NCC). When multiple cues are available, there are several strategies used to fuse the information [26]. *Boosting*, for example, consists in selecting the most representative features from a large set of proposed features using a machine learning algorithm such as AdaBoost [236]. *Concatenation* uses features from different cues and concatenates them for computation. *Summation* takes affinity values from different features and adds a weight to each value. *Product strategy* assumes independence between affinity values and computes their weighted product. *Cascading* uses diverse visual representations and tries to determine the finest model appearance [224]. To improve tracking prediction, exclusion models can be used to prevent physical collisions, assuming that two distinct pedestrians cannot be at the same place at the same time. Two types of constraints can be applied to the trajectory hypotheses: detection-level exclusion and trajectory-level exclusion [142]. Detection-level exclusion assumes that two detections in a frame cannot be assigned to the same target. Trajectory-level exclusion means that two trajectories cannot be too close to each other. In order to avoid that, a penalty is assigned to two hypotheses that are too close and which have different trajectories, to suppress one of them.

Occlusion The third problem is how to handle *occlusions* of tracking targets. Three major strategies are employed to face this challenge. *Part-to-whole* divides the object into several parts and then computes an affinity for each part. When an occlusion occurs, only the unoccluded parts are taken into account for estimation [210, 237]. In *hypothesize-and-test*, detection hypotheses are generated for two objects with different levels of occlusion, which are then tested for example using MAP or a multi-person detector [213]. The *buffer-and-recover* technique keeps the states of objects over several frames, before and during an occlusion. When it ends, the states of objects are recovered using the observations on the frame buffer [189].

5.4.3 Multi-Tracks and Data Association Methods

Probabilistic or deterministic optimization are the common methods to deal with multiple tracks and data association problems. Data association is about the uncertainty related to measurements, it aims at associating observed measurements with current known tracks or generate new tracks. Deterministic optimization methods are usually suitable for offline tracking, as they require observations from several or all the frames in advance [142], whereas probabilistic methods are commonly used for online or real-time tracking. Bar-Shalom and Li [10] presented several data association algorithms, such as Nearest Neighbors (NN), Multi-Hypothesis Tracking (MHT), Joint Probabilistic Data Association Filter (JPDAF), or Probability Hypothesis Density (PHD), and evaluated their performances.

Global Nearest Neighbour (GNN) GNN [22] is one of the simplest methods for data association. At every new time step, it 'hardly' assigns each current observation to a single best object without revising the past. In [124], GNN is described as a 5-step algorithm: (1) receive data for each scan; (2) each track is first defined as a cluster and if common observations are found for two tracks, they are merged into a 'super cluster'; (3) observations are assigned to each cluster using Munkres algorithm [126]; (4) tracks' states are updated using some estimation technique such as Kalman filter; (5) observations which are not associated to any existing tracks are used to create new tracks. The work in [8] developed a multiple person tracker where GNN is used for data association with a new distance function and a Kalman filter for state estimation. The proposed method is suitable for occlusion issues.

Multiple Hypothesis Tracking (MHT) This filter, originally proposed by Reid [180], is an iterative algorithm which can handle multiple tracking targets, with occlusions, and give optimal solutions. It makes predictions on each hypothesis for the succeeding frame. Each hypothesis represented a group of mutually separate tracks [219]. The aim of MHT is to overcome the wrong data association problem by representing the posterior belief with a mixture of Gaussians, where each Gaussian component is considered to be a track and relies on a unique data association decision. MHT is a more complex approach than GNN: it propagates assignment probabilities over time as a tree of the future observations in order to resolve past ambiguities. Luber *et al.* [141] proposed a model that uses social force model as a motion model for MHT. Motivations, principles and implementations of MHT are presented in [21]. MHT is generally considered to be too slow and memory-expensive for multi-target tracking methods as pruning and priming have to be applied in order to keep the size of the tree manageable [121]. Amditis *et al.* [4] proposed examples of MHT implementation for MTT using laser scanner data.

Joint Probabilistic Data Association Filter (JPDAF) This method has been proposed by [87]. It generates multiple tracks-to-measurement hypotheses and calculates the hypotheses probabilities. Then, it gives hard, unrevisable assignment of hypotheses that are merged to each track at each time step. This is more complex than GNN because the latter is greedy and just assigns each observation individually to its nearest object, while JPDAF allows some entanglement over space. In contrast, MHT filter allows some entanglement over time [4], considering all the joint data-object assignments and picking the best. JPDAF runs faster than MHT [255], but it requires a fixed number of targets. Chen *et al.* [42] proposed the use of a JPDAF to compute hidden Markov models transition probabilities for a contour-based human tracking method performing in real-time. Liu *et al.* [139] proposed a person tracking method combining JPDAF and multi-sensor fusion. [106] implemented a tracking method based on JPDAF and capable of tracking about 400 persons in real-time. Rezatofighi *et al.* [183] presented a JPDAF-based tracker for challenging conditions, such as observations from fluorescence microscopy sequences or surveillance cameras.

Probabilistic Hypothesis Density (PHD) This filter was introduced by [49]. It can track a variable number of tracks, estimating their number and their locations at each time step. There are different types of PHD filters, such as the Sequential Monte Carlo PHD filter (SMC-PHD) [187], the Gaussian Mixture PHD filter (GM-PHD) [244] and the Gaussian Inverse Wishart PHD filter (GIW-PHD) [99]. Zhang *et al.* [246] used a GMM-PHD (Gaussian Mixture Measurement PHD) tracker to tackle problems with bearing measurements. Khazaei *et al.* [119] developed a PHD filter in distributed camera network where each camera fuses its track estimates with its neighbors. Feng *et al.* [81] proposed a variational Bayesian PHD filter with deep learning update to track multiple persons. In [57], a PHD filter is used to track in real-time multiple people in a crowded environment. Yoon *et al.* [241] used hybrid (i.e. local and global) observations in a PHD filter, where the filter observations are combined with local observations generated by on-line trained detectors. This method allows to handle missed detections and it assigns an identity to each person.

Markov Chain Monte Carlo Data Association (MCMCDA) Introduced first by [168], this filter is an approximation of the Bayesian filter, derived from MCMC, which draws a set of samples and builds Markov chains over the target state space. A sampler moves from its current state to the next following the proposal distribution. The new state is accepted with an acceptance probability, otherwise the sampler stays at its current state. Oh *et al.* [159, 158] proposed an MCMCDA algorithm known as Metropolis-Hastings, where single-scan and multi-scan MCMCDA algorithms are used for known and unknown number of targets, respectively. A bipartite graph is used to represent possible associations between observations and targets. Their simulation results show a better performance than MHT algorithms and their method has been tested on tracking people from video sequences. Yu *et al.* [243] proposed a data-driven MCMC (DD-MCMC) approach for sampling and incorporating a person's motion and appearance information, using a joint probability model. Their method was tested in simulations and on real videos.

Bipartite Graph Matching This uses two sets of graph nodes representing existing trajectories and new detections in online tracking, or two sets of tracklets (components of tracks) in offline tracking. The weights of nodes model affinities between trajectories and detections. The Bipartite assignment algorithm or optimal Hungarian algorithm is used to find matching nodes in the two sets. A review on graph matching is presented in [53]. Chen *et al.* [109] used a dynamical graph matching method to track multiple people in order to dynamically change the graph nodes with the tracks movements.

Dynamic Programming This method solves the data association problem by linking several detections over time. Pirsiavash *et al.* [170] used a greedy algorithm based on dynamic programming to find the global solution in a network flow. Another method is presented in [18] which can follow up to six people over several frames.

Min-Cost Max-Flow Network Flow This is a popular method, which models the network flow as a directed graph. A trajectory is represented by a start node and an end node (sink), and it corresponds to one flow path in the graph. The global optimal solution is obtained with the push-relabel algorithm. Zhang *et al.* [245] used a min-cost flow algorithm combined with a recursive occlusion model to deal with occluded people. Their method does not require pruning. Chari *et al.* [38] proposed a new approach to the min-cost max-flow network flow optimization using pair-wise costs, which can deal with occluded people.

Conditional Random Field (CRF) A graph $G = (V, E)$ is defined as a set of nodes V and a set of edges E. Nodes represent observations and tracklets. A label is used to predict which track observations are linked to. Sutton and McCallum [211] presented a CRF tutorial. Taycher *et al.* [216] proposed a person tracking method learning from data, based on a CRF state-space estimation and a grid-filter with real-time capabilities. Milan *et al.* [148] developed a CRF-based MTT, detecting people using a HOG-SVM detector, and defining two unary potentials for detection and superpixel nodes. Milan *et al.* [147] proposed a CRF-based multiple person tracker using discrete-continuous energy minimization, whose goal is to assign a unique trajectory to each detection.

Maximum-Weight Independent Set (MWIS) The MWIS graph is defined as $G = (V, E, w)$. As in the CRF, the nodes V represent the pairs of tracklets in successive frames, which are given a weight w indicating the affinity of the tracklet pair. If two tracklets share the same detection, then their edges E are connected together. Brendel *et al.* [25] proposed a multi-target tracker based on MWIS data association algorithm. Their approach is as follows: (1) detection of multiple targets in all frames using different object detectors; (2) detections are considered as distinct tracks, with the assumption that one detection can only be one track; (3) a graph is built to match tracks over two consecutive frames; (4) an MWIS algorithm is used to perform the data association with guaranteed optimal solution; (5) statistical and contextual properties of objects are learnt online for their similarity measurement using Mahalanobis distances; steps (2) to (5) are repeated over the frames to handle long-term occlusions by merging or splitting tracks. In [105], a multi-person tracker is used with data association modelled as a Connected Component Model (CCM) based on MWIS. A divide-and-conquer strategy is used to solve the Multi-Dimensional Assignment (MDA) problem.

5.4.4 Advanced MTT Techniques

Here *Advanced MTT* refers to multi-target tracking that is performed at a higher-level, simultaneously with other tasks.

MTT with Video Adaptation MTT approaches rely on an object detector that is trained offline, so its performance can be totally different from a video to another. A possible solution is to create a generic detector adapted for a specific video by tuning some parameters. Previous works for multiple people tracking include [91, 39].

MTT with Deep Learning Deep learning has proven to be a high performance method for classification, detection and many computer visions tasks. Applied to MTT, deep learning could provide a stronger observation model which could increase the tracking accuracy [242, 131]. In [161], Ondruska *et al.* introduced deep tracking, an end-to-end human tracking approach, based on recurrent neural network, using unsupervised learning on simulated data without dealing with the data association problem. In [66], Dequaire *et al.* used a similar method for static and dynamic person tracking in real-world environments. In [149], Milan *et al.* proposed a complete online multiple people tracking method based on recurrent neural networks.

MTT under Multiple Cameras Also called Multi-Target Multi-Camera (MTMC), this type of systems can be used to improve large tracking problems. Wang *et al.* [228] presented a survey on the challenges of MTMC. One problem would be overlapping cameras, in which case it is necessary to find a good way to fuse multiple information. But if the camera angles do not overlap, then the data association problem becomes an identification problem. In [184], Ristani *et al.* proposed different performance measures to test MTMC methods. In [185], they used neural networks to learn features from MTMC systems and for re-identification. In [140], Generalized Maximum Multi-Clique optimization – a graph-based method – is used for the MTMC problem. Munaro *et al.* [155] developed an open-source software, called OpenPTrack, for multi-camera calibration and people tracking using RGB-D data.

Multiple 3D Object Tracking This method could provide better position accuracy, size estimation and occlusion handling. The major problem for this technique is the camera calibration. Park *et al.* [167] applied 3D object tracking from a monocular camera for augmented reality applications. Some other works on 3D visual tracking include [59, 166, 194], which used a single camera with a multi-Bernoulli mixture tracking filter. Some works with 3D lidar sensors include [114, 207, 234], which proposed online classification of humans for 3D lidar tracking. In [175], both camera and lidar data are used to improve people tracking.

MTT with Scene Understanding Scene understanding can provide contextual information and scene structure for the tracking algorithm, especially in crowded scenes. Leal-Taixé *et al.* [128] developed a model that decomposes an image and extracts features from the observed scene called 'interaction feature strings'. These features are then used in a Random Forest framework to track human targets [64].

MTT with Other Computer Vision Tasks Information from image segmentation or human pose estimation could not only improve the performance of multiple-people tracking but also the computation of the tracking algorithm. For example, in [148], tracking is done with image segmentation and in [47] people are tracked for group activity recognition.

5.5 Discussion

Single pedestrian tracking is now a fully mature area with widely available open-source and commercial implementations. Body pose tracking has made strong recent progress, likely to soon bring it to maturity, through the use of larger data sets and computer power.

Tracking multiple pedestrians requires additional algorithms which were major research areas until recently, but have largely matured in the last few years with methods such as MHT becoming standard. Tracking multiple pedestrians in the presence of occlusion by one another or by other objects remains a serious research problem, which requires the use of other data or prior information to compensate for the lack of purely visual data. We suggest that the higher-level models from psychology and sociology discussed in the Part II of this review [29] should be used to provide such priors. Traditionally, tracking was a clearly separate task from both lower (detection) and higher (behaviour modelling) layers of pedestrian modelling, but a current trend is to merge it with nearby layers through neural network and probabilistic methods in this fashion to improve performance.

Practical implementation of tracking algorithms may be found in the Bayes Tracking library[1] which provides open-source implementation of EKF, UKF and SIR Particle Filters with NN and JPDA data association algorithms. In addition, a detection and tracking pipeline[2]

[1] https://github.com/LCAS/bayestracking
[2] https://github.com/sbreuers/detta

contains an implementation of MHT. Choi *et al.* [46] proposed a fast tracker TRACA[1] with a deep feature compression approach for single target tracking.

In terms of computational efficiency, Bellotto and Hu [14] have shown that Kalman-based people tracking is much faster than particle-based, and in particular that UKF was faster and still almost as reliable as particle filter. Linder *et al.* [136] proposed a comparison (computation speed and other metrics) of various people tracking methods, including NN trackers, MHT and others. A common heuristic for some mobile robots is to run at 10Hz or more, i.e. if the robot moves at 1m/s, a people tracker running at 10Hz will estimate the position of humans every 10cm, which is usually considered safe. But with cars moving much faster such as 10m/s (36km/h), the computational requirements would be greater, such as operating 100Hz to obtain the same 10cm accuracy.

6 Conclusions

Autonomous vehicles must interact with pedestrians in order to drive safely and to make progress. It is not enough to simply stop whenever a pedestrian is in the way as this leads to the freezing robot problem and to the vehicle making no progress. Rather, AVs must develop similar interaction methods as used by human drivers, which include understanding the behaviour and predicting the future behaviour of pedestrians, predicting how pedestrians will react to the AVs movements, and choosing those motions to efficiently control the interaction.

This Part I review has surveyed the state of the art in the lower levels of machine perception and intelligence needed to enable such interaction control, namely: sensing, detection, recognition, and tracking of pedestrians. It has found that the level of maturity of these fields is high at the lowest levels, but fades into current research areas at the higher-levels. Sensing technology has progressed to maturity over the last decade so that lidars and stereo cameras are now reliable and cheap enough for use in research and even by hobbyist systems. Similarly, GPUs have fallen in price to enable both stereo camera processing and deep learning recognition to be run in these systems. Deep learning recognition has largely replaced classical feature-based methods for detection. Open-source software is mature and freely available for these tasks. Beyond detection are areas with successful, open-source, partial implementations but which require further research to become fully mature. Recognition of body pose and head direction are almost mature, including via deep learning methods.

[1] https://github.com/jongwon20000/TRACA

But recognition of higher-level states, such as gestures used for explicit signalling, body language used as implicit signalling, actions as sequences of poses, and recognition of underlying emotional state, remain research areas.

Tracking is mature for single pedestrians, but remains challenging for multiple pedestrians in the presence of occlusion. Algorithms to solve this task are known but require the use of extensive prior knowledge to predict behaviour in the absence of sensory information, which is not yet fully available. This includes information from recognition of poses, gestures, actions, and emotions, but also feedback information from very high-level models of behaviour and psychology which will be studied in Part II of this review [29].

Acknowledgements

This project has received funding from EU H2020 project interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

References

[1] A. Adam, E. Rivlin, and I. Shimshoni. Robust fragments-based tracking using the integral histogram. In *Proc. of IEEE CVPR*, volume 1, pages 798–805, 2006.

[2] Ejaz Ahmed, Michael Jones, and Tim K Marks. An improved deep learning architecture for person re-identification. In *Proc. of CVPR*, pages 3908–3916, 2015.

[3] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[4] Angelos Amditis, George Thomaidis, Pantelis Maroudis, Panagiotis Lytrivis, and Giannis Karaseitanidis. *Multiple Hypothesis Tracking Implementation*, chapter 10. InTech, Rijeka, 2012.

[5] Anelia Angelova, Alex Krizhevsky, Vincent Vanhoucke, Abhijit Ogale, and Dave Ferguson. Real-time pedestrian detection with deep network cascades. In *Proc. of BMVC*, 2015.

[6] Kai O. Arras, Óscar Martínez Mozos, and Wolfram Burgard. Using boosted features for the detection of people in 2d range data. In *Proc. of IEEE ICRA*, pages 3402–3407, 2007.

[7] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Sig. Proc.*, 50(2):174–188, 2002.

[8] Mohammad Azari, Ahamd Seyfi, and Amir Hossein Rezaie. Real time multiple object tracking and occlusion reasoning using adaptive kalman filters. In *Proc. of 7th Iranian Machine Vision and Image Processing (MVIP)*, pages 1–5. IEEE, 2011.

[9] T. BaltruÅ¡aitis, P. Robinson, and L. Morency. Openface: An open source facial behavior analysis toolkit. In *Proc. of IEEE Winter Conference on Applications of Computer Vision*, pages 1–10, 2016.

[10] Y. Bar-Shalom and X.R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques.* Y. Bar-Shalom, 1995.

[11] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software.* John Wiley & Sons, 2004.

[12] Benjamin K. Barton, Thomas A. Ulrich, and Roger Lew. Auditory detection and localization of approaching vehicles. *Accident Analysis & Prevention*, 49:347 – 353, 2012. PTW + Cognitive impairment and Driving Safety.

[13] A. M. Baumberg and D. C. Hogg. An efficient method for contour tracking using active shape models. In *Proc. of IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pages 194–199, 1994.

[14] N. Bellotto and H. Hu. Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of Bayesian filters. *Autonomous Robots*, 28(4):425–438, 2010.

[15] Nicola Bellotto, Serhan Cosar, and Zhi Yan. *Human Detection and Tracking.* Springer edition, 2019.

[16] Rodrigo Benenson, Markus Mathias, Tinne Tuytelaars, and Luc Van Gool. Seeking the strongest rigid detector. In *Proc. of IEEE CVPR*, pages 3666–3673, 2013.

[17] Ben Benfold and Ian D Reid. Guiding visual surveillance by tracking human attention. In *Proc. of BMVC*, volume 2, page 7, 2009.

[18] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In *Proc. of IEEE CVPR*, volume 1, pages 744–750, 2006.

[19] M Bigas, Enric Cabruja, Josep Forest, and Joaquim Salvi. Review of CMOS image sensors. *Microelectronics journal*, 37(5):433–451, 2006.

[20] Osameh Biglari, Reza Ahsan, and Majid Rahi. Human detection using SURF and SIFT feature extraction methods in different color spaces. *J Math Comput Sci*, 11(111):274, 2014.

[21] S. S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5–18, 2004.

[22] Samuel S Blackman and (author.) Popoli, Robert. *Design and analysis of modern tracking systems.* Boston : Artech House, 1999.

[23] A. F. Bobick and A. D. Wilson. Parametric hidden Markov models for gesture recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 21:884–900, 1999.

[24] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision in C++ with the OpenCV Library.* O'Reilly Media, Inc., 2nd edition, 2013.

[25] W. Brendel, M. Amer, and S. Todorovic. Multiobject tracking as maximum weight independent set. In *Proc. of CVPR*, pages 1273–1280, 2011.

[26] Richard R Brooks and Sundararaja S Iyengar. *Multi-sensor fusion: fundamentals and applications with software.* Prentice-Hall, Inc., 1998.

[27] Aurélie Bugeau and Patrick Pérez. Detection and segmentation of moving objects in complex scenes. *Computer Vision and Image Understanding*, 113(4):459–476, 2009.

[28] F. Camara, O. Giles, R. Madigan, M. Rothmüller, P. H. Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox. Predicting

pedestrian road-crossing assertiveness for autonomous vehicle control. In *Proc. of IEEE ITSC*, pages 2098–2103, 2018.

[29] Fanta Camara, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Anna Schieben, Gustav Markkula, Fabio Tango, Natasha Merat, and Charles W. Fox. Pedestrian models for autonomous driving Part II: high-level models of human behaviour. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[30] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proc. of the International Conference on Intelligent Autonomous Systems (IAS-15) Workshops*, 2018.

[31] Erik Cambria. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107, 2016.

[32] Erik Cambria, Andrew Livingstone, and Amir Hussain. The hourglass of emotions. In *Cognitive behavioural systems*, pages 144–157. Springer, 2012.

[33] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.

[34] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2019.

[35] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proc. of IEEE CVPR*, pages 1302–1310, 2017.

[36] Alessio Carullo and Marco Parvis. An ultrasonic sensor for distance measurement in automotive applications. *IEEE Sensors journal*, 1(2):143, 2001.

[37] Alexandros Andre Chaaraoui, Pau Climent-Pèrez, and Francisco Flòrez-Revuelta. Silhouette-based human action recognition using sequences of key poses. *Pattern Recognition Letters*, 34(15):1799 – 1807, 2013. Smart Approaches for Human Action Recognition.

[38] Visesh Chari, Simon Lacoste-Julien, Ivan Laptev, and Josef Sivic. On pairwise costs for network flow multi-object tracking. In *Proc. of IEEE CVPR*, 2015.

[39] Duc Phu Chau, Monique Thonnat, and François Brémond. Automatic parameter adaptation for multi-object tracking. In *Proc. of International Conference on Computer Vision Systems (ICVS)*, pages 244–253, 2013.

[40] Feng-Sheng Chen, Chih-Ming Fu, and Chung-Lin Huang. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing*, 21(8):745 – 758, 2003.

[41] Yu-Ting Chen and Chu-Song Chen. A cascade of feed-forward classifiers for fast pedestrian detection. In *Asian Conference on Computer Vision*, pages 905–914. Springer, 2007.

[42] Yunqiang Chen, Yong Rui, and Thomas S Huang. Jpdaf based hmm or real-time contour tracking. In *null*, page 543. IEEE, 2001.

[43] Zhe Chen et al. Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69, 2003.

[44] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[45] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath. Millimeter-wave vehicular communication to support massive automotive sensing. *IEEE Communications Magazine*, 54(12):160–167, 2016.

[46] Jongwon Choi, Hyung Jin Chang, Tobias Fischer, Sangdoo Yun, Kyuewang Lee, Jiyeoup Jeong, Yiannis Demiris, and Jin Young Choi. Context-aware Deep Feature Compression for High-speed Visual Tracking. In *Proc. of IEEE CVPR*, pages 479–488, 2018.

[47] Wongun Choi and Silvio Savarese. A unified framework for multi-target tracking and collective activity recognition. In *Proc. of ECCV*, pages 215–230. Springer, 2012.

[48] Grzegorz Cielniak, Tom Duckett, and Achim J. Lilienthal. Data association and occlusion handling for vision-based people tracking by mobile robots. *Robotics and Autonomous Systems*, 58(5):435–443, 2010.

[49] Daniel Edward Clark. *Multiple target tracking with the probability hypothesis density filter*. PhD thesis, Heriot-Watt University, 2006.

[50] R. T. Collins. Mean-shift blob tracking through scale space. In *Proc. of IEEE CVPR*, volume 2, pages II–234, 2003.

[51] R. T. Collins, R. Gross, and Jianbo Shi. Silhouette-based human identification from body shape and gait. In *Proc. of 5th IEEE International Conference on Automatic Face Gesture Recognition*, pages 366–371, 2002.

[52] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. of CVPR*, volume 2, pages 142–149 vol.2, 2000.

[53] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(03):265–298, 2004.

[54] Claudio Coppola, Serhan Cosar, Diego R. Faria, and Nicola Bellotto. Social activity recognition on continuous RGB-D video sequences. *International Journal of Social Robotics*, 2019.

[55] J. Y. R. Cornejo and H. Pedrini. Recognition of occluded facial expressions based on centrist features. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1298–1302, 2016.

[56] Jadisha Y Ramírez Cornejo, Helio Pedrini, and Francisco Flórez-Revuelta. Facial expression recognition with occlusions based on geometric representation. In *Iberoamerican Congress on Pattern Recognition*, pages 263–270. Springer, 2015.

[57] Javier Correa, Jindong Liu, and Guang-Zhong Yang. Real time people tracking in crowded environments with range measurements. In *International Conference on Social Robotics*, pages 471–480. Springer, 2013.

[58] S. Cosar, F. Fernandez-Carmona, R. Agrigoroaie, F J. Pages, Ferland, F. Zhao, S. Yue, N. Bellotto, and A. Tapus. Enrichme: Perception and interaction of an assistive robot for the elderly at home. *International Journal of Social Robotics*, 2019.

[59] A. Crivellaro, M. Rad, Y. Verdie, K. M. Yi, P. Fua, and V. Lepetit. Robust 3d object tracking from monocular images using stable parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1465–1479, 2018.

[60] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE CVPR*, volume 1, pages 886–893 vol. 1, 2005.

[61] T. Darrell, A. P. Pentland, A. Azarbayejani, and C. R. Wren. Pfinder : Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 19:780–785, 1997.

[62] Fabian de Ponte Müller. Survey on ranging sensors and cooperative techniques for relative positioning of vehicles. *Sensors*, 17(2):271, 2017.

[63] Iván del Pino, Víctor Vaquero, Beatrice Masini, Joan Solà, Francesc Moreno-Noguer, Alberto Sanfeliu, and Juan Andrade-Cetto. Low resolution lidar-based multi-object tracking for driving applications. In *Proc. of Iberian Robotics conference*, pages 287–298, 2017.

[64] Vincent Delaitre. *Modeling and Recognizing Interactions between People, Objects and Scenes.* Theses, ENS Paris - Ecole Normale Supérieure de Paris, 2015.

[65] MJ Den Uyl and H Van Kuilenburg. The facereader: Online facial expression recognition. In *Proc. of Measuring Behavior*, volume 30, pages 589–590. Citeseer, 2005.

[66] Julie Dequaire, Peter Ondrúška, Dushyant Rao, Dominic Wang, and Ingmar Posner. Deep tracking in the wild: End-to-end tracking using recurrent neural networks. *International Journal of Robotics Research*, 37(4-5):492–512, 2018.

[67] Ayush Dewan, Tim Caselitz, Gian Diego Tipaldi, and Wolfram Burgard. Motion-based detection and tracking in 3D LiDAR scans. In *Proc. of IEEE ICRA*, pages 4508–4513, 2016.

[68] Mariella Dimiccoli, Alejandro Cartas, and Petia Radeva. Chapter 6 - activity recognition from visual lifelogs: State of the art and future challenges. In Xavier Alameda-Pineda, Elisa Ricci, and Nicu Sebe, editors, *Multimodal Behavior Analysis in the Wild*, pages 121 – 134. Academic Press, 2019.

[69] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *Proc. of IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 65–72, 2005.

[70] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Proc. of CVPR*, 2009.

[71] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, 2012.

[72] Alexander Ekimov and James M Sabatier. Human detection range by active Doppler and passive ultrasonic methods. In *Proc. of Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VII*, volume 6943, page 69430R, 2008.

[73] Oron Eliezer, Kumar Anil, and Bar-Shalom Yaakov. Precision tracking with segmentation for imaging sensors. *IEEE Transactions on Aerospace and Electronic Systems*, 29(3):977–986, 1993.

[74] Hadi Elzein, Sridhar Lakshmanan, and Paul Watta. A motion and shape-based pedestrian detection algorithm. In *Proc. of IEEE IV*, pages 500–504, 2003.

[75] Markus Enzweiler and Dariu M Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, 2008.

[76] Andreas Ess, Bastian Leibe, Konrad Schindler, and Luc J. van Gool. Moving obstacle detection in highly dynamic scenes. In *Proc. of IEEE ICRA*, pages 56–63, 2009.

[77] M. Everingham, S. M. A. Eslami, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015.

[78] L. Fan, Z. Wang, B. Cail, C. Tao, Z. Zhang, Y. Wang, S. Li, F. Huang, S. Fu, and F. Zhang. A survey on multiple object tracking algorithm. In *Proc. of IEEE International Conference on Information and Automation (ICIA)*, pages 1855–1862, 2016.

[79] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[80] X. Fen and G. Ming. Pedestrian tracking using particle filter algorithm. In *Proc. of International Conference on Electrical and Control Engineering*, pages 1478–1481, 2010.

[81] P. Feng, W. Wang, S. M. Naqvi, and J. Chambers. Variational Bayesian PHD Filter with deep learning network updating for multiple human tracking. In *Proc. of Sensor Signal Processing for Defence*, pages 1–5, 2015.

[82] Fengliang Xu, Xia Liu, and K. Fujimura. Pedestrian detection and tracking with night vision. *IEEE Transactions on Intelligent Transportation Systems*, 6(1):63–71, 2005.

[83] Antonio Fernández-Caballero, José Carlos Castillo, Javier Martínez-Cantos, and Rafael Martínez-Toms. Optical flow or image subtraction in human detection from infrared camera on mobile robot. *Robotics and Autonomous Systems*, 58(12):1273–1281, 2010.

[84] David Fernandez-Llorca, Raul Quintero Minguez, Ignacio Parra Alonso, Carlos Fernandez Lopez, Ivan Garcia Daza, Miguel Angel Sotelo, and Cristina Alen Cordero. Assistive intelligent transportation systems: The need for user localization and anonymous disability identification. *IEEE Intelligent Transportation Systems Magazine*, 9(2):25–40, 2017.

[85] William J Fleming. New automotive sensors – a review. *IEEE Sensors Journal*, 8(11):1900–1921, 2008.

[86] Fabian Flohr, Madalin Dumitru-Guzu, Julian FP Kooij, and Dariu M Gavrila. Joint probabilistic pedestrian head and body orientation estimation. In *Proc. of IEEE IV*, pages 617–622, 2014.

[87] T. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8(3):173–184, 1983.

[88] Charles Fox. *Data Science for Transport: A Self-Study Guide with Computer Exercises*. 2018.

[89] Katerina Fragkiadaki, Sergey Levine, Panna Felsen, and Jitendra Malik. Recurrent network models for human dynamics. In *Proc. of IEEE ICCV*, pages 4346–4354, 2015.

[90] William T Freeman and Michal Roth. Orientation histograms for hand gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, volume 12, pages 296–301, 1995.

[91] Adrien Gaidon and Eleonora Vig. Online domain adaptation for multi-object tracking. In *Proc. of BMVC*, pages 3.1–3.13, 2015.

[92] J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn, and H. Seidel. Motion capture using joint skeleton tracking and surface estimation. In *Proc. of IEEE CVPR*, pages 1746–1753, 2009.

[93] Davi Geiger, Alok Gupta, Luiz A Costa, and John Vlontzos. Dynamic programming for detecting, tracking, and matching deformable contours. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (3):294–302, 1995.

[94] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat. Vision and RFID data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding*, 114(6):641–651, 2010.

[95] Partha Ghosh, Jie Song, Emre Aksan, and Otmar Hilliges. Learning human motion models for long-term predictions. In *Proc. of the International Conference on 3D Vision (3DV)*, pages 458–466, 2017.

[96] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. of IEEE CVPR*, pages 580–587, 2014.

[97] Alejandro González, Gabriel Villalonga, Jiaolong Xu, David Vázquez, Jaume Amores, and Antonio M. López. Multiview random forest of local experts combining RGB and LIDAR data for pedestrian detection. In *Proc. of IEEE IV*, pages 356–361, 2015.

[98] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[99] K. Granstrom and U. Orguner. A phd filter for tracking multiple extended targets using random matrices. *IEEE Transactions on Signal Processing*, 60(11):5657–5671, 2012.

[100] Ian Gresham, Alan Jenkins, Robert Egri, Channabasappa Eswarappa, Noyan Kinayman, Nitin Jain, Richard Anderson, Frank Kolak, Ratana Wohlert, Shawn P Bawell, et al. Ultra-wideband radar sensors for short-range vehicular applications. *IEEE Transactions on Microwave Theory and Techniques*, 52(9):2105–2122, 2004.

[101] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with Microsoft Kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5):1318–1334, 2013.

[102] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge University Press, 2003.

[103] Jürgen Hasch, Eray Topak, Raik Schnabel, Thomas Zwick, Robert Weigel, and Christian Waldschmidt. Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. *IEEE Transactions on Microwave Theory and Techniques*, 60(3):845–860, 2012.

[104] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proc. of IEEE ICCV*, pages 2980–2988, 2017.

[105] Z. He, X. Li, X. You, D. Tao, and Y. Y. Tang. Connected component model for multi-object tracking. *IEEE Transactions on Image Processing*, 25(8):3698–3711, 2016.

[106] Paul Horridge and Simon Maskell. Real-time tracking of hundreds of targets with efficient exact JPDAF implementation. In *Proc. of IEEE International Conference on Information Fusion*, pages 1–8, 2006.

[107] Wenbo Huang, K Kim, Yong Yang, and Yoo Sung Kim. Automatic shadow removal by illuminance in HSV color space. *Comput. Sci. Inf. Technol*, 3(3):70–75, 2015.

[108] Patrick Hurney, Peter Waldron, Fearghal Morgan, Edward Jones, and Martin Glavin. Review of pedestrian detection techniques in automotive far-infrared video. *IET Intelligent Transport Systems*, 9(8):824–832, 2015.

[109] Hwann-Tzong Chen, Horng-Horng Lin, and Tyng-Luh Liu. Multi-object tracking using dynamical graph matching. In *Proc. of IEEE CVPR*, volume 2, pages II–II, 2001.

[110] Umar Iqbal, Anton Milan, and Juergen Gall. Posetrack: Joint multi-person pose estimation and tracking. In *Proc. of IEEE CVPR*, pages 2011–2020, 2017.

[111] Omid Hosseini Jafari, Dennis Mitzel, and Bastian Leibe. Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras. In *Proc. of IEEE ICRA*, pages 5636–5643. IEEE, 2014.

[112] Seunghun Jin, Junguk Cho, Xuan Dai Pham, Kyoung Mu Lee, Sung-Kee Park, Munsang Kim, and Jae Wook Jeon. Fpga design and implementation of a real-time stereo vision system. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(1):15–26, 2010.

[113] Frédéric Jurie and Michel Dhome. Real time robust template matching. In *Proc. of BMVC*, pages 1–10, 2002.

[114] Achim Kampker, Mohsen Sefati, Arya Abdul Rachman, Kai Kreisköther, and Pascual Campoy. Towards multi-object detection and tracking in urban scenario under uncertainties. In *Proc. of VEHITS*, 2018.

[115] T. Kaneko and O. Hori. Feature selection for reliable tracking using template matching. In *Proc. of CVPR*, volume 1, pages I–I, 2003.

[116] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada. An open approach to autonomous vehicles. *IEEE Micro*, 35(6):60–68, 2015.

[117] Christoph G Keller, Markus Enzweiler, Marcus Rohrbach, David Fernández Llorca, Christoph Schnorr, and Dariu M Gavrila. The benefits of dense stereo for pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1096–1106, 2011.

[118] Bahman Yari Saeed Khanloo, Ferdinand Stefanus, Mani Ranjbar, Ze-Nian Li, Nicolas Saunier, Tarek Sayed, and Greg Mori. A large margin framework for single camera offline tracking with hybrid cues. *Computer Vision and Image Understanding*, 116(6):676–689, 2012.

[119] M. Khazaei and M. Jamzad. Multiple human tracking using PHD filter in distributed camera network. In *Proc. of International Conference on Computer and Knowledge Engineering*, pages 569–574, 2014.

[120] Hilke Kieritz, Stefan Becker, Wolfgang Hübner, and Michael Arens. Online multi-person tracking using integral channel features. In *Proc. of IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 122–130, 2016.

[121] Chanho Kim, Fuxin Li, Arridhana Ciptadi, and James M Rehg. Multiple hypothesis tracking revisited. In *Proc. of IEEE ICCV*, pages 4696–4704, 2015.

[122] Y. Kim, S. Ha, and J. Kwon. Human detection using Doppler radar based on physical characteristics of targets. *IEEE Geoscience and Remote Sensing Letters*, 12(2):289–293, 2015.

[123] Yoshiki Kohari, Jun Miura, and Shuji Oishi. Cnn-based human body orientation estimation for robotic attendant. In *Proc. of IAS-15 Workshop on Robot Perception of Humans*, 2018.

[124] Pavlina Konstantinova, Alexander Udvarev, and Tzvetan Semerdjiev. A study of a target tracking algorithm using global nearest neighbor approach. In *Proc. of CompSysTech*, pages 290–295, 2003.

[125] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. of NIPS*, volume 1, pages 1097–1105, 2012.

[126] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[127] Anil Kumar, Yaakov Bar-Shalom, and Eliezer Oron. Precision tracking based on segmentation with optimal layering for imaging sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):182–188, 1995.

[128] Laura Leal-Taixé, Michele Fenzi, Alina Kuznetsova, Bodo Rosenhahn, and Silvio Savarese. Learning an image-based motion context for multiple people tracking. In *Proc. of IEEE CVPR*, 2014.

[129] Laura Leal-Taixé, Anton Milan, Konrad Schindler, Daniel Cremers, Ian Reid, and Stefan Roth. Tracking the trackers: An analysis of the state of the art in multiple object tracking. *arXiv:1704.02781 [cs]*, 2017. arXiv: 1704.02781.

[130] Yann Lecun, LÃ©on Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proc. of the IEEE*, pages 2278–2324, 1998.

[131] Byungjae Lee, Enkhbayar Erdenee, Songguo Jin, Mi Young Nam, Young Giu Jung, and Phill Kyu Rhee. Multi-class multi-object tracking using changing point detection. In *Proc. of ECCV Workshops*, pages 68–83, 2016.

[132] Hyeon-Kyu Lee and J. H. Kim. An HMM-based threshold model approach for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):961–973, 1999.

[133] Stéphanie Lefèvre, Dizan Vasquez, and Christian Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, 1(1):1, 2014.

[134] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proc. of IEEE CVPR*, pages 152–159, 2014.

[135] Liang Lin, Yongyi Lu, Chenglong Li, Hui Cheng, and Wangmeng Zuo. Detection-free multiobject tracking by reconfigurable inference with bundle representations. *IEEE Transactions on Cybernetics*, 46(11):2447–2458, 2016.

[136] T. Linder, S. Breuers, B. Leibe, and K. O. Arras. On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In *Proc. of IEEE ICRA*, pages 5512–5519, 2016.

[137] A. J. Lipton, H. Fujiyoshi, and R. S. Patil. Moving target classification and tracking from real-time video. In *Proc. of IEEE Workshop on Applications of Computer Vision*, pages 8–14, 1998.

[138] An-An Liu, Wei-Zhi Nie, Yu-Ting Su, Li Ma, Tong Hao, and Zhao-Xuan Yang. Coupled hidden conditional random fields for RGB-D human action recognition. *Signal Processing*, 112:74 – 82, 2015. Signal Processing and Learning Methods for 3D Semantic Analysis.

[139] Nanyang Liu, Rong Xiong, Qianshan Li, and Yue Wang. Human tracking using improved sample-based joint probabilistic data association filter. In *Proc. of Intelligent Autonomous Systems*, pages 293–302, 2013.

[140] WenQian Liu, Octavia I. Camps, and Mario Sznaier. Multi-camera multi-object tracking. *CoRR*, abs/1709.07065, 2017.

[141] Matthias Luber, Johannes Andreas Stork, Gian Diego Tipaldi, and Kai O. Arras. People tracking with human motion predictions from social forces. In *Proc. of IEEE ICRA*, 2010.

[142] Wenhan Luo, Xiaowei Zhao, and Tae-Kyun Kim. Multiple object tracking: A review. *CoRR*, abs/1409.7618, 2014.

[143] Mario Marchetti, Vincent Boucher, Jean Dumoulin, and Michèle Colomb. Retrieving visibility distance in fog combining infrared thermography, principal components analysis and partial least-square regression. *Infrared Physics & Technology*, 71:289–297, 2015.

[144] Julieta Martinez, Michael J. Black, and Javier Romero. On human motion prediction using recurrent neural networks. *Proc. of IEEE CVPR*, pages 4674–4683, 2017.

[145] S. V. Martnez, J. F. Knebel, and J. P. Thiran. Multi-object tracking using the particle filter algorithm on the top-view plan. In *Proc. of the 12th European Signal Processing Conference*, pages 285–288, 2004.

[146] François Meyer and Patrick Bouthemy. Region-based tracking in an image sequence. In *Proc. of ECCV*, pages 476–484, 1992.

[147] A. Milan, K. Schindler, and S. Roth. Multi-target tracking by discrete-continuous energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):2054–2068, 2016.

[148] Anton Milan, Laura Leal-Taixe, Konrad Schindler, and Ian Reid. Joint tracking and segmentation of multiple targets. In *Proc. of IEEE CVPR*, 2015.

[149] Anton Milan, Seyed Hamid Rezatofighi, Anthony R. Dick, Konrad Schindler, and Ian D. Reid. Online multi-target tracking using recurrent neural networks. In *Proc. of AAAI*, 2017.

[150] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3):311–324, 2007.

[151] Dennis Mitzel and Bastian Leibe. Close-range human detection and tracking for head-mounted cameras. In *Proc. of BMVC*, 2012.

[152] Thomas B Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, 2006.

[153] Sungphill Moon, Youngbin Park, Dong Wook Ko, and Il Hong Suh. Multiple Kinect sensor fusion for human skeleton tracking using Kalman filtering. *International Journal of Advanced Robotic Systems*, 13(2):65, 2016.

[154] B. T. Morris and M. M. Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1114–1127, 2008.

[155] Matteo Munaro, Filippo Basso, and Emanuele Menegatti. Openptrack: Open source multi-camera calibration and people tracking for RGB-D camera networks. *Robotics and Autonomous Systems*, 75:525–538, 2016.

[156] Matteo Munaro and Emanuele Menegatti. Fast RGB-D people tracking for service robots. *Autonomous Robots*, 37(3):227–242, 2014.

[157] Willie Neiswanger, Frank Wood, and Eric Xing. The dependent dirichlet process mixture of objects for detection-free tracking and object modeling. In *Proc. of the 7th International Conference on Artificial Intelligence and Statistics*, pages 660–668, 2014.

[158] Songhwai Oh, S. Russell, and S. Sastry. Markov chain Monte Carlo data association for general multiple-target tracking problems. In *Proc. of the 43rd IEEE Conference on Decision and Control*, volume 1, pages 735–742, 2004.

[159] Songhwai Oh, Stuart Russell, and Shankar Sastry. Markov chain Monte Carlo data association for multi-target tracking. *IEEE Transactions on Automatic Control*, 54(3):481–497, 2009.

[160] Katsuyuki Ohguchi, Masayoshi Shono, and Masayuki Kishida. 79 GHz band ultra-wideband automotive radar. *Fujitsu Ten Tech. J.*, 39:9–14, 2013.

[161] Peter Ondruska and Ingmar Posner. Deep tracking: Seeing beyond seeing using recurrent neural networks. In *Proc. of AAAI*, 2016.

[162] Dan Oneata, Jérôme Revaud, Jakob Verbeek, and Cordelia Schmid. Spatio-temporal object detection proposals. In *Proc. of ECCV*, pages 737–752, 2014.

[163] Sakrapee Paisitkriangkrai, Chunhua Shen, and Anton van Den Hengel. Strengthening the effectiveness of pedestrian detection with spatially pooled features. In *Proc. of ECCV*, pages 546–561. Springer, 2014.

[164] Himani S Parekh, Darshak G Thakore, and Udesang K Jaliya. A survey on object detection and tracking methods. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(2):2970–2978, 2014.

[165] Dennis Park, C Lawrence Zitnick, Deva Ramanan, and Piotr Dollár. Exploring weak stabilization for motion feature extraction. In *Proc. of IEEE CVPR*, pages 2882–2889, 2013.

[166] Jin-hyung Park, Seungmin Rho, and Chang-sung Jeong. Real-time robust 3d object tracking and estimation for surveillance system. *Security and Communication Networks*, 7(10):1599–1611, 2013.

[167] Y. Park, V. Lepetit, and Woontack Woo. Multiple 3d object tracking for augmented reality. In *Proc. of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 117–120, 2008.

[168] Hanna Pasula, Stuart Russell, Michael Ostland, and Yaacov Ritov. Tracking many objects with many sensors. In *Proc. of IJCAI*, volume 99, pages 1160–1171, 1999.

[169] Peter Pinggera, David Pfeiffer, Uwe Franke, and Rudolf Mester. Know your limits: Accuracy of long range stereoscopic object measurements in practice. In *Proc. of ECCV*, pages 96–111, 2014.

[170] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *Proc. of CVPR*, pages 1201–1208, 2011.

[171] Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Amir Hussain. Convolutional mkl based multimodal emotion recognition and sentiment analysis. In *Proc. of IEEE International Conference on Data Mining*, pages 439–448, 2016.

[172] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on Lie algebra. In *Proc. of IEEE CVPR*, volume 1, pages 728–735, 2006.

[173] Danil V Prokhorov. Road obstacle classification with attention windows. In *Proc. of IEEE IV*, pages 889–895, 2010.

[174] R. Quintero, I. Parra, J. Lorenzo, D. Fernández-Llorca, and M. A. Sotelo. Pedestrian intention recognition by means of a hidden Markov model and body language. In *Proc. of IEEE ITSC*, pages 1–7, 2017.

[175] A. Rangesh and M. M. Trivedi. No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras and lidars. *IEEE Transactions on Intelligent Vehicles*, 4(4):588–599, 2019.

[176] Siddharth S. Rautaray and Anupam Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.

[177] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In *Proc. of IEEE CVPR*, pages 6517–6525, 2017.

[178] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. of IEEE CVPR*, pages 779–788, 2016.

[179] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.

[180] Donald Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843–854, 1979.

[181] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[182] Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. Robust part-based hand gesture recognition using Kinect sensor. *IEEE Transactions on Multimedia*, 15(5):1110–1120, 2013.

[183] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid. Joint probabilistic data association revisited. In *Proc. of IEEE ICCV*, pages 3047–3055, 2015.

[184] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set forÂ multi-target, multi-camera tracking. In *Proc. of ECCV Workshops*, pages 17–35, 2016.

[185] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification. In *Proc. of IEEE CVPR*, 2018.

[186] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, 2004.

[187] B. Ristic, D. Clark, and B. Vo. Improved smc implementation of the PHD filter. In *Proc. of the 13th International Conference on Information Fusion*, pages 1–8, 2010.

[188] Mikel D Rodriguez and Mubarak Shah. Detecting and segmenting humans in crowded scenes. In *Proc. of the 15th ACM international conference on Multimedia*, pages 353–356. ACM, 2007.

[189] M. S. Ryoo and J. K. Aggarwal. Observe-and-explain: A new approach for multiple hypotheses tracking of humans and objects. In *Proc. of IEEE CVPR*, pages 1–8, 2008.

[190] SAE International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2018.

[191] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[192] Cameron Schaeffer. A comparison of keypoint descriptors in the context of pedestrian detection: FREAK vs. SURF vs. BRISK. Technical report, 2013.

[193] Fredrik Schalling, Sebastian Ljungberg, and Naveen Mohan. Benchmarking lidar sensors for development and evaluation of automotive perception. In *Proc. of IEEE International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–6, 2019.

[194] S. Scheidegger, J. Benjaminsson, E. Rosenberg, A. Krishnan, and K. Granström. Mono-camera 3d multi-object tracking using deep learning detections and pmbm filtering. In *Proc. of IEEE IV*, pages 433–440, 2018.

[195] Thomas Schlegl, Thomas Bretterklieber, Markus Neumayer, and Hubert Zangl. Combined capacitive and ultrasonic distance measurement for automotive applications. *IEEE Sensors Journal*, 11(11):2636–2642, 2011.

[196] Tobias Schubert, Alexis Gkogkidis, Tonio Ball, and Wolfram Burgard. Automatic initialization for skeleton tracking in optical motion capture. In *Proc. of IEEE ICRA*, pages 734–739. IEEE, 2015.

[197] A. T. Schulz and R. Stiefelhagen. Pedestrian intention recognition using latent-dynamic conditional random fields. In *Proc. of IEEE IV*, pages 622–627, 2015.

[198] Dirk Schulz, Dieter Fox, and Jeffrey Hightower. People tracking with anonymous and id-sensors using rao-blackwellised particle filters. In *Proc. of IJCAI*, 2003.

[199] Loren Arthur Schwarz, Artashes Mkhitaryan, Diana Mateus, and Nassir Navab. Human skeleton tracking from depth data using geodesic distances and optical flow. *Image and Vision Computing*, 30(3):217–226, 2012.

[200] A. Shashua, Y. Gdalyahu, and G. Hayun. Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In *Proc. of IEEE IV*, pages 1–6, 2004.

[201] Yaser Sheikh, Omar Javed, and Takeo Kanade. Background subtraction for freely moving cameras. In *Proc. of IEEE ICCV*, pages 1219–1225, 2009.

[202] Jianbo Shi and Carlo Tomasi. Good features to track. Technical report, Cornell University, 1993.

[203] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *Proc. of IEEE CVPR*, pages 1297–1304, 2011.

[204] Additi Mrinal Singh, Soumyasree Bera, and Rabindranath Bera. Review on vehicular radar for road safety. In *Advances in Communication, Cloud, and Big Data*, pages 41–47, 2019.

[205] Chanjira Sinthanayothin, Nonlapas Wongwaen, and Wisarut Bholsithi. Skeleton tracking using Kinect sensor & displaying in 3 d virtual scene. *International Journal of Advancements in Computing Technology*, 4(11), 2012.

[206] Arnold WM Smeulders, Dung M Chu, Rita Cucchiara, Simone Calderara, Afshin Deghghan, and Mubarak Shah. Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1442–1468, 2013.

[207] S. Song, Z. Xiang, and J. Liu. Object tracking with 3d lidar via multi-task sparse learning. In *Proc. of IEEE International Conference on Mechatronics and Automation*, pages 2603–2608, 2015.

[208] James V. Stone. *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis.* Sebtel Press, 2013.

[209] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *Proc. of IEEE IV*, pages 206–212, 2006.

[210] Daisuke Sugimura, Kris M Kitani, Takahiro Okabe, Yoichi Sato, and Akihiro Sugimoto. Using individuality to track individuals: Clustering individual trajectories in crowds using local appearance and frequency trait. In *Proc. of IEEE ICCV*, pages 1467–1474, 2009.

[211] Charles Sutton and Andrew McCallum. An introduction to conditional random fields. *Foundations and Trends in Machine Learning*, 4(4):267–373, 2012.

[212] Richard Szeliski. *Computer Vision: Algorithms and Applications.* Springer-Verlag, 1st edition, 2010.

[213] Siyu Tang, Mykhaylo Andriluka, and Bernt Schiele. Detection and tracking of occluded people. *International Journal of Computer Vision*, 110(1):58–69, 2014.

[214] Y. Tang, L. Ma, W. Liu, and W. Zheng. Long-Term Human Motion Prediction by Modeling Motion Context and Enhancing Motion Dynamic. In *Proc. of IJCAI-ECAI*, 2018.

[215] Hai Tao, H. S. Sawhney, and R. Kumar. Object tracking with Bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):75–89, 2002.

[216] L. Taycher, D. Demirdjian, T. Darrell, and G. Shakhnarovich. Conditional random people: Tracking humans with CRFs and grid filters. In *Proc. of IEEE CVPR*, volume 1, pages 222–229, 2006.

[217] A. Techmer. Contour-based motion estimation and object tracking for real-time applications. In *Proc. of International Conference on Image Processing*, volume 3, pages 648–651, 2001.

[218] Thiago Teixeira, Gershon Dublon, and Andreas Savvides. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. *ACM Computing Surveys*, 5(1):59–69, 2010.

[219] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[220] Jonathan J Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Proc. of NIPS*, pages 1799–1807, 2014.

[221] R. Trichet and F. Bremond. Dataset optimization for real-time pedestrian detection. *IEEE Access*, 6:7719–7727, 2018.

[222] Douglas L. Vail, Manuela M. Veloso, and John D. Lafferty. Conditional random fields for activity recognition. In *Proc. of ACM International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 235:1–235:8, 2007.

[223] Mathias Versichele, Tijs Neutens, Matthias Delafontaine, and Nico Van de Weghe. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography*, 32(2):208 – 220, 2012.

[224] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE CVPR*, volume 1, pages I–511–I–518, 2001.

[225] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. In *Proc. of IEEE CVPR*, pages 1030–1037, 2010.

[226] Junqiu Wang and Yasushi Yagi. Shadow extraction and application in pedestrian detection. *EURASIP Journal on Image and Video Processing*, 2014(1):12, 2014.

[227] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[228] Yong Wang, Rui Zhai, and Ke Lu. Challenge of multi-camera tracking. In *Proc. of the 7th International Congress on Image and Signal Processing*, 2014.

[229] P. J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974.

[230] Jingshu Wu, Rory Austin, and Chou-Lin Chen. Incidence rates of pedestrian and bicyclist crashes by hybrid electric passenger vehicles: An update. Technical report, 2011.

[231] Fen Xu and Ming Gao. Human detection and tracking based on HOG and particle filter. In *Proc. of the 3rd International Congress on Image and Signal Processing*, volume 3, 2010.

[232] Hulya Yalcin, Michael J Black, and Ronan Fablet. The dense estimation of motion and appearance in layers. In *Proc. of IEEE CVPR Workshops*, pages 165–165, 2004.

[233] J. Yan, Z. Lei, L. Wen, and S. Z. Li. The fastest deformable part model for object detection. In *Proc. of IEEE CVPR*, pages 2497–2504, 2014.

[234] Z. Yan, T. Duckett, and N. Bellotto. Online learning for 3d lidar-based human detection: Experimental analysis of point cloud clustering and classification methods. *Autonomous Robots*, 2019.

[235] Z. Yan, L. Sun, T. Duckett, and N. Bellotto. Multisensor online transfer learning for 3d lidar-based human detection with a mobile robot. In *Proc. of IEEE/RSJ IROS*, pages 7635–7640, 2018.

[236] Bo Yang and Ram Nevatia. Online learned discriminative part-based appearance models for multi-human tracking. In *Proc. of ECCV*, pages 484–498, 2012.

[237] Ehwa Yang, Jeonghwan Gwak, and Moongu Jeon. Multi-human tracking using part-based appearance modelling and grouping-based tracklet association for visual surveillance applications. *Multimedia Tools and Applications*, 76(5):6731–6754, 2017.

[238] Kwok-Yun Yeung, Tsz-Ho Kwok, and Charlie CL Wang. Improved skeleton tracking by duplex Kinects: A practical approach for real-time applications. *Journal of Computing and Information Science in Engineering*, 13(4), 2013.

[239] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4):13–es, 2006.

[240] Alper Yilmaz, Xin Li, and Mubarak Shah. Object contour tracking using level sets. In *Proc. of the Asian Conference on Computer Vision*, 2004.

[241] J. H. Yoon, K. J. Yoon, and D. Y. Kim. Multi-object tracking using hybrid observation in PHD filter. In *Proc. of IEEE International Conference on Image Processing*, pages 3890–3894, 2013.

[242] Fengwei Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan. Poi: Multiple object tracking with high performance detection and appearance feature. In *Proc. of ECCV Workshops*, pages 36–42, 2016.

[243] Q. Yu, G. Medioni, and I. Cohen. Multiple target tracking using spatio-temporal Markov chain Monte Carlo data association. In *Proc. of IEEE CVPR*, pages 1–8, 2007.

[244] H. Zhang, Jinlong Yang, Hongwei Ge, and Le Yang. An improved GM-PHD tracker with track management for multiple target tracking. In *Proc. of International Conference on Control, Automation and Information Sciences*, pages 185–190, 2015.

[245] Li Zhang, Yuan Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *Proc. of IEEE CVPR*, pages 1–8, 2008.

[246] Qian Zhang and Taek Lyul Song. Improved bearings-only multi-target tracking with GM-PHD filtering. *Sensors*, 16(9):1469, 2016.

[247] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele. How far are we from solving pedestrian detection? In *Proc. of IEEE CVPR*, pages 1259–1267, 2016.

[248] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele. Towards reaching human performance in pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):973–986, 2018.

[249] Zhang Zhang, Kaiqi Huang, and Tieniu Tan. Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In *Proc. of IEEE International Conference on Pattern Recognition*, volume 3, pages 1135–1138, 2006.

[250] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19:4–12, 2012.

[251] Fuquan Zhao, Hao Jiang, and Zongwei Liu. Recent development of automotive LiDAR technology, industry and trends. In *Proc. of the 11th International Conference on Digital Image Processing*, volume 11179, pages 1132 – 1139, 2019.

[252] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proc. of IEEE ICCV*, 2015.

[253] Lingxiang Zheng, Xiaoyang Ruan, Yunbiao Chen, and Minzheng Huang. Shadow removal for pedestrian detection and tracking in indoor environments. *Multimedia Tools and Applications*, 76(18):18321–18337, 2017.

[254] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In *Proc. of IEEE ICCV*, pages 3774–3782, 2017.

[255] Bin Zhou and NK Bose. Multitarget tracking in clutter: Fast algorithms for data association. *IEEE Transactions on Aerospace and Electronic Systems*, 29(2):352–363, 1993.

[256] Junyi Zhou and Jing Shi. RFID localization algorithms and applications – a review. *Journal of Intelligent Manufacturing*, 20(6):695, 2009.

[257] Yue Zhou and Hai Tao. A background layer model for object tracking through occlusion. In *Proc. of IEEE ICCV*, pages 1079–1085, 2003.

[258] Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Proc. of IEEE CVPR*, volume 2, pages 1491–1498, 2006.

Chapter 3

Pedestrian Models for Autonomous Driving Part II: High-Level Models of Human Behavior

Abstract

Autonomous vehicles (AVs) must share space with pedestrians, both in carriageway cases such as cars at pedestrian crossings and off-carriageway cases such as delivery vehicles navigating through crowds on pedestrianized high-streets. Unlike static obstacles, pedestrians are active agents with complex, interactive motions. Planning AV actions in the presence of pedestrians thus requires modelling of their probable future behaviour as well as detecting and tracking them. This narrative review article is Part II of a pair, together surveying the current technology stack involved in this process, organising recent research into a hierarchical taxonomy ranging from low-level image detection to high-level psychological models, from the perspective of an AV designer. This self-contained Part II covers the higher levels of this stack, consisting of models of pedestrian behaviour, from prediction of individual pedestrians' likely destinations and paths, to game-theoretic models of interactions between pedestrians and autonomous vehicles. This survey clearly shows that, although there are good models for optimal walking behaviour, high-level psychological and social modelling of pedestrian behaviour still remains an open research question that requires many conceptual issues to be clarified. Early work has been done on descriptive and qualitative models of behaviour, but much work is still needed to translate them into quantitative algorithms for practical AV control.

1 Introduction

To operate successfully in the presence of pedestrians, autonomous vehicles require input from a huge variety of models that have to work seamlessly together. These models range from simple visual models for detection of pedestrians, to predicting their future movements using psychological and sociological methods. Part I of this two-part survey [35] covered models for sensing, detection, recognition, and tracking of pedestrians. Part II here reviews models

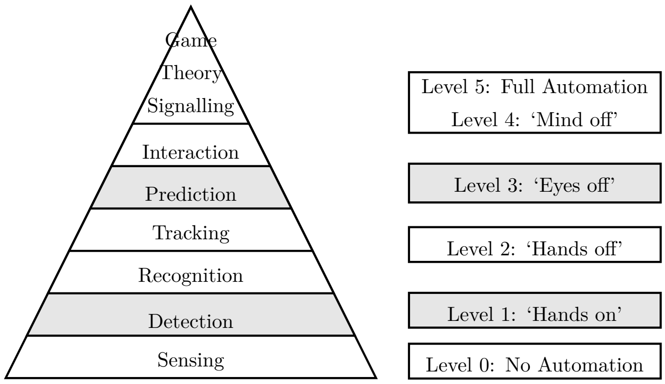


Figure 3.1: Main structure of the review.

for pedestrian trajectory prediction, interaction of pedestrians, and behavioral modelling of pedestrians, and also experimental resources to validate all the types of models. Interacting with pedestrians is a particular type of social intelligence. Autonomous vehicles will need to utilize many different levels of models of pedestrians, each addressing different aspects of perception and action. Each of these models can be based on empirical science results or obtained via machine learning. In contrast to the models of Part I, Part II requires models from higher levels of the technology stack, as researched by psychologists and taught in advanced driver training programmes. For instance, drivers often try to infer the personality of other humans, predict their likely behaviours, and interact with them to communicate mutual intentions [102]. Between the high level surveyed in this Part II and the low levels of Part I, researchers infer psychological information from perceptual information. As an example, researchers build systems to recognize the body language, gestures, and demographics information of pedestrians to better predict their likely goals and behaviours. Despite the importance of bridging the research between the higher and lower levels, their connection is still thin, both conceptually and in terms of actual implementations.

While prediction of likely future pedestrian trajectories is becoming increasingly well understood, models for actively controlling pedestrian interactions – including game-theoretic models – are still in their infancy. Active control here means that the vehicle's own future actions are taken into account in predicting how the pedestrian will respond, and vice versa. One reason is that sufficient data to rigorously study interaction between pedestrians has only recently become available as presented in Sec. 5 on experimental resources. Another reason is that one first has to be able to reliably sense, detect, recognize, and track pedestrians in order to gather enough data for modelling interaction and game-theoretic models. A third reason is

Table 3.1: Proposed mapping from SAE levels to pedestrian model requirements.

SAE LEVEL	DESCRIPTION	MODEL REQUIREMENTS	SECTION
0	No Automation. Automated system issues warnings and may momentarily intervene, but has no sustained vehicle control.	Sensing	Part I Sec. 2
1	Hands on. The driver and the automated system share control of the vehicle. For example, adaptive cruise control (ACC), where the driver controls steering and the automated system controls speed. The driver must be ready to resume full control when needed.	+Detection	Part I Sec. 3
2	Hands off. The automated system takes full control of the vehicle (steering and speed). The driver must monitor and be prepared to intervene immediately. Occasional contact between hand and wheel is often mandatory to confirm that the driver is ready to intervene.	+Recognition +Tracking	Part I Sec. 4 Part I Sec. 5
3	Eyes off. Driver can safely turn attention away from the driving tasks, e.g. use a phone or watch a movie. Vehicle will handle situations that call for an immediate response, like emergency braking. The driver must still be prepared to intervene within some limited time.	+Unobstructed Walking Models, Known Goals +Behaviour Prediction, Known Goals +Behaviour Prediction, Unknown Goals	Sec. 2.1 Sec. 2.2 Sec. 2.3
4	Mind off. No driver attention is required for safety, except in limited spatial areas (geofenced) or under special circumstances, like traffic jams.Outside of these areas or circumstances, the vehicle must be able to safely abort or transfer control to the human.	+Event/Activity Models +Effects of Pedestrian Class on Trajectory +Pedestrian Interaction Models +Game Theoretic and Signalling Models	Sec. 2.4 Sec. 2.5 Sec. 3 Sec. 4
5	Full automation. No human intervention is required at all, fully automated driving.	+Extreme Robustness and Reliability	

Note: '+X' means that 'X' is required in addition to the requirements of the previous level.

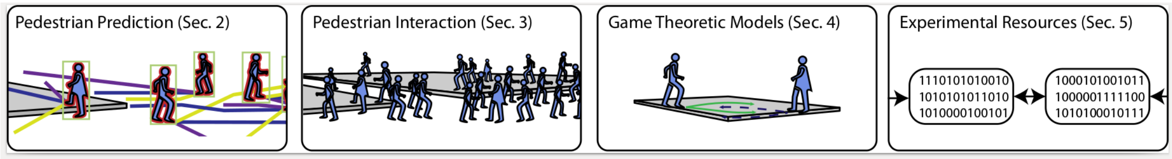


Figure 3.2: Structure of the paper.

that interaction and game-theoretic models are only relevant in crowded environments, while many situations do not require much interaction. However, crowded environments are those that are typically most relevant for autonomous driving. Fig. 3.1 shows the review structure.

To assess the maturity of the methods presented, the level of autonomy is used, as defined by the Society of Automotive Engineers (SAE) – the same measure has already been used in Part I [35]. For the convenience of the reader, the five SAE levels are briefly presented, ranging from simple driver assistance tools to full self-driving [183]. Requirements for pedestrian modelling increase with each level, with lower levels typically requiring lower and more mature levels of pedestrian models, such as detection and tracking, while higher levels require models for psychological and social understanding to fully interact with pedestrians in a human-like way [30]. Table 3.1 gives an overview of SAE levels and requirements mappings.

While many papers propose pedestrian models at various levels, no unifying theory has yet been produced which would make it possible to easily transfer results across all levels from detection to prediction. This review uncovers bottlenecks in transferring results to facilitate closing existing research gaps. Also, many existing studies only consider results from empirical science or those obtained via machine learning. This survey provides an overview considering both possibilities. While machine learning results work particularly well for detection and recognition, they are not yet performing so well for prediction. Some reasons are that prediction is a more high-dimensional problem, with dimensions including goals, obstacles, various state variables of pedestrians, and road geometry. A further reason is that less labelled data is available for training prediction models. A promising future direction is to combine empirical science results with machine learning to better safeguard techniques using machine learning and to avoid over-fitting.

While similar concepts apply to modelling human drivers and their vehicles for interactions with AVs, this article presents a review of the state of the art specifically in modelling human pedestrians for social decision-making. In some cases it goes beyond modelling aspects to also cover more conceptual aspects or empirical psychological findings, when the studies in question are judged to have very direct applicability to mathematical models. Results from human driving cannot be directly translated to pedestrians due to the variability in locomotion, the differences in shape, the changes in postures and the less-structured environment.

Pedestrians are defined as humans moving on and near public highways including roads and pedestrianised areas, who walk using their own locomotive power. This excludes, for example, humans moving on cycles, wheelchairs and other mobility devices, skates and skateboards, or those transported by other humans. This review does not cover interactions of traffic participants without pedestrians: a survey on trajectory prediction of on-road vehicles is provided in [123] and a survey on vision-based trajectory learning is provided in [146].

This Part II is organized as shown in Fig. 3.2. In Sec. 2, methods for predicting the movements of pedestrians are reviewed. In particular, we consider models and methods for unstructured environments, for prediction around obstacles, to estimate destinations, and for the prediction of events such as crossing the road. These methods are enhanced in Sec. 3 for groups of pedestrians interacting with each other. This section considers the complete variety of researched models from macroscopic models only considering flow of people to microscopic models that consider individual pedestrians. In many situations, interaction models do not require game theory, because pedestrians often have different goals. However, there are also many situations, where pedestrians have competing goals, e.g., when several pedestrians have to pass a narrow passage. In such situations, the game theoretic models presented in Sec. 4 can be very useful. Finally, Sec. 5 surveys available resources: datasets and simulators, both for pedestrians and vehicles.

2 Behaviour Models without Interaction

The tracking models reviewed in Part I are *kinematic* in that they assume that pedestrians move in physical and/or pose space in motion described by kinematic models. This is a very basic assumption – human drivers typically have much more complex understandings and hence predictions of pedestrian behavior which they use to drive safely in their presence [102]. These range from slightly more advanced kinematic understandings such as ‘pedestrians tend to walk in straight lines’ to models of how they are likely to interact with static objects in their environment, and predictions of pedestrians’ likely destinations from reading the street scene.

This section reviews such models starting from simple unobstructed path models to uncer-

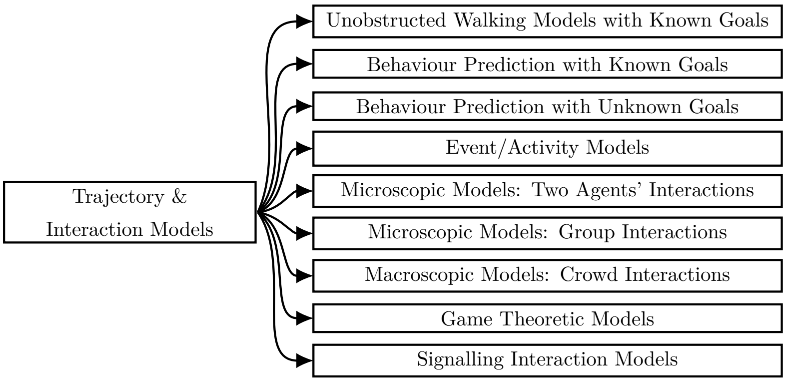


Figure 3.3: Pedestrian behaviour prediction and interaction models.

tain destination models and more advanced event/activity models. These models do not yet consider interaction with other agents. Figure 3.3 summarizes the classes of models presented in this section. A previous review was proposed by Ridel *et al.* [172], which mainly considered pedestrian crossing intent and offered a restricted view of the different models developed for trajectory prediction.

2.1 Unobstructed Walking Models with Known Goals

Given a start location and orientation, and a goal location, humans do not typically turn towards the goal on the spot (which would waste time) and then walk in a straight line, but rather set off walking in their initial heading and adjust their orientation gradually as they walk, resulting in smooth, curved trajectories from origin to destination [72]. Models from optimal control theory as also used in robotics [50] define cost functions for travel time, speed, and accelerations, to reproduce these characteristic curved trajectories. The model in [72] instead achieves curved trajectories by modelling the rate of turning of the pedestrian as a function of the visual angle and distance to the goal. A simple kinematic model consists in considering human locomotion as a nonholonomic motion [161], using the unicycle model (3.1) where the pedestrian walking trajectory is represented by the trajectory of their center of gravity, 2D coordinates (x, y) and by the angle θ,

$$\begin{aligned} \dot{x} &= u_1 \cos\theta \\ \dot{y} &= u_1 \sin\theta \\ \dot{\theta} &= u_2 \end{aligned} \tag{3.1}$$

where u_1 is the forward velocity and u_2 is the angular velocity. Assuming known origin and destination with inverse optimal control, one can reliably predict human walking paths using this model [9] [155].

2.2 Behaviour Prediction with Known Goals

Here, the likely behaviour of a pedestrian in a static environment is considered, given a map. Pedestrians are likely to route around obstacles, and to stop at the edges of roads before crossing. This section does not consider social effects of other agents – this is presented later in Sec. 3.

2.2.1 Dynamic Graphical Models

Dynamic Graphical Models (DGM) are Graphical Models of a particular topology, containing some Markovian sequence of variables over time. DGMs include simple Markov and Hidden Markov Models and also more complex models. The method in [145] used tracking in a DGM based on particle filter approximation to infer beliefs over future pedestrian trajectories and combined this with a GNSS (Global Navigation Satellite System) module that provides information about the hazardous areas and people.

2.2.2 Gaussian Process Methods

Habibi *et al.* [88] proposed a context-based approach to pedestrian trajectory prediction using Gaussian Processes [166]. This model incorporates context features such as the pedestrian's distance to the traffic light, the distance to the curbside, and the curbside orientation in the transition learning phase to improve the prediction. A context-based augmented semi non-negative sparse coding (CASNSC) algorithm is used to predict pedestrian trajectories.

2.2.3 Deep Learning Methods

Bock *et al.* [24] developed a Recurrent Neural Network (LSTM) model to learn pedestrian behaviour patterns at intelligent intersections using camera data from the onboard vehicle and the infrastructure. The model can predict trajectories for a horizon of 5s.

2.2.4 Other Methods

Kruse *et al.* [121] was one of the first attempts to statistically infer human motion patterns from data and incorporate them in a robot motion planner for obstacle avoidance. Garzón *et al.* [77] presented a pedestrian trajectory prediction model based on two path planning algorithms that require a set of possible goals, a map and the initial position. It then computes similarities between the obtained and observed trajectories into probabilities. This model is run along with a pedestrian detector and tracker. Tamura *et al.* [198] proposed a pedestrian behaviour model that is based on social forces and takes into account the intention of the pedestrian in the trajectory prediction by defining a set of subgoals. In [171] the uncertain goals are used as latent variables to guide the motion prediction of pedestrians. Their positions are predicted by combining forward propagation of a physical model with local a priori information (e.g., obstacles and different road types) from the start position, and by planning the trajectory from a goal position. The distribution over the destinations is modeled with a particle filter.

In [209], Vasishta *et al.* presented a model based on the principle of natural vision that incorporates contextual information extracted from the environment to the pedestrian behavior and it especially tries to predict hazardous behavior such as crossing in non authorized areas. The aforementioned model in [72] considers goals and obstacles as distance-dependent attractors and repellers in heading angle space. The contributions from the goal and obstacles are linearly combined, yielding a momentary rate of acceleration of heading, which results in human-like trajectories for simultaneous goal-seeking and obstacle avoidance. In [57], Dias *et al.* developed a model simulating pedestrian behaviour around corners, using minimum jerk theory and one-thirds power law concept. Their model uses Monte Carlo simulation to generate pedestrian trajectories with turning maneuvers, which were comparable to empirical trajectories.

2.3 Behaviour Prediction with Unknown Goals

Many of the above models assume known probable destinations for pedestrians, which enable routing to act not just around local obstacles, but to predict entire long-term trajectories, such as for pedestrians intending to cross the road. However, in reality a pedestrian's destination is rarely given.

2.3.1 Dynamic Graphical Models

Ziebart *et al.* [233] presented a pedestrian trajectory prediction model that takes into account hindrance due to robot motion, as is required in off-carriageway interactions such as last mile AVs in pedestrianized areas. A maximum entropy inverse optimal control technique, introduced in [232], is used and is equivalent to a soft-maximum version of Markov decision process (MDP) that accounts for decision uncertainty into the trajectories distribution. The cost function is a linear combination of the features (e.g obstacles) in the environment. People's motion can be modeled by an MDP and by choosing a certain path, there is an immediate reward. The model is conditioned on a known destination location but the model reasons about all possible destinations and the real destination is not known at the prediction time. The destination is inferred in a Bayesian way, by computing the prior distributions over destinations using previous observed trajectories. When there is no previous data, features (door, chair etc.) in the environment are used to model the destination. In [113], Kitani *et al.* extended [232, 233] by incorporating visual features to forecast future activities and destinations. The observations provided by the vision system (e.g. tracking algorithm) are assumed to be noisy and uncertain therefore they used a hidden variable Markov decision process (hMDP) where the agent knows its own states, action and reward but observes only noisy measurements. Negative Log-Loss (NLL) is used as a probabilistic metric and Modified Hausdorff Distance (MHD) as a physical measure of the distance between two trajectories. Vasquez [210] extends the work of Ziebart [233] and Kitani [113] while reducing computational costs.

Bennewitz *et al.* [18, 17] proposed a learning method for human motion recognition using the expectation maximization (EM) and a hidden Markov model (HMM) for clustering and predicting human trajectories and incorporating them into a robot path planner. In [221], Wu *et al.* presented a model that uses Markov chains for pedestrian motion prediction (able to deal with non-Gaussian distribution and several constraints). A heuristic method is proposed to automatically infer the positions of several potential goals on a generic semantic map. It also incorporates policies to predict the pedestrian motion direction and takes into account other traffic participants by incorporating a collision checking approach. Borgers *et al.* [29] presented a model that predicts pedestrians' route choice based on Markov chains. Similarly, Bai *et al.* [11] presented a real-time approximate POMDP (Partially Observable Markov Decision Process) controller, DESPOT, for use in high-street type environments. The method is intention-aware in the sense of inferring pedestrian destinations and route plans from their observed motion over time, and accounting for the value of this information against the value of making progress while planning a robot's own route around them. Karasev *et al.* [110] presented a long-term prediction model that incorporates environmental constraints with the intent modeled by a policy in a MDP framework. The pedestrian state is estimated using a Rao-Blackwellized filter and pedestrian intent by planning according to a stochastic policy. This model assumes that pedestrians behave rationally.

2.3.2 Deep Learning Methods

Hug *et al.* [98] proposed a LSTM-MDL model combined with a particle filter method for multi-modal trajectory prediction, and tested on Stanford Drone Dataset (SDD) [176]. Rehder *et al.* [170] proposed a method to infer pedestrian destinations. The trajectory prediction is computed as a goal-oriented motion planning. The whole system is based on deep-learning and trained via inverse reinforcement learning. A general introduction on reinforcement learning in robotics can be found in [115]. Deo *et al.* [56] presented a framework for multi-modal pedestrian trajectory forecasting in structured environments. They used a convolutional neural network to compute both the reward maps of the path states and the possible goal states for MDPs. The derived policy information is then fed into a recurrent neural network, combined with track history, to generate possible future trajectories. Goldhammer *et al.* [81] developed a Multilayer Perceptron (MLP) neural network with polynomial least square approximation to predict pedestrian trajectories based on camera data. A long-term prediction model using RNNs is proposed in [22].

2.3.3 Other Methods

Cosgun *et al.* [52] presented a person-following service robot with a task dependent motion planner. The robot can track and predict the future trajectory of the person by maximizing its reward at future steps while avoiding entering into the human's personal space. Koschi *et al.* [117] proposed a set-based method to predict all possible behaviours of pedestrians using reachability analysis [5] for pedestrian occupancy. Pedestrians are described as point mass with a certain maximum velocity and maximum acceleration. A rule-based occupancy is applied that does not allow a pedestrian to obstruct traffic, e.g. pedestrians are given priority at crosswalks and their trajectory is assumed to be evasive.

2.4 Event/Activity Models

Pedestrian event models consider stereotypical sequences of behaviours of individual pedestrians. These may give additional information about route choice, beyond that available from static classification of the pedestrian. For example, a commuter, or class of commuters, who engage in similar actions every day, such as road crossing in a certain way then checking their phone, may reveal information about their identity to enable re-identification[1] which is in turn predictive of their future destinations. These models look for features predictive of route choice in static environments and do not consider social factors.

2.4.1 Dynamic Graphical Methods

Duckworth *et al.* [65] [64] developed on a mobile robot an unsupervised qualitative spatio-temporal relations (QSR) model to learn motion patterns using a graph representation and is able to predict people's future behaviour. Dondrup *et al.* [63] presented a ROS-based real-time human perception framework for mobile robots using laser and RGB-D data and tracking people with a Kalman filter approach. Human trajectories are converted into QSR (Qualitative Spatial Relations) and used for a Hidden Markov Model (HMM) to classify the behaviour of the different people encountered [62]. In [187], Schneider and Gavrila presented a comparative study on Bayesian filters (EKF and IMM) for short-term (<2s) pedestrian trajectory prediction, in particular they used stereo camera images to apply these methods to four different types of behaviour: crossing, stopping, bending in and starting.

Body heading is used above in basic path planning models, but head-turning events are distinct from body heading, and are discrete events which occur when a pedestrian turns their head to look around rather than to orient their body. Such an activity model is used in [116] to enhance path prediction of pedestrians while intending to cross a street. For low-level occupancy prediction, a dynamic Bayesian network (DFBN) is used on top of a switching linear dynamic system (SLDS) anticipating the changes of pedestrian dynamics. As in [116],

[1] Identity here is distinct from "personal information" as defined by privacy laws such as the EU General Data Protection Regulation (GDPR).

studies [189, 190] also model head orientation by an event/activity model to enhance the underlying prediction approach.

2.4.2 Gaussian Process Methods

Quintero *et al.* [162] [163] proposed a pedestrian path prediction method up to 1s ahead based on balanced Gaussian Process dynamical models (B-GPDMs) and naïve Bayes classifiers. GPDM is used to transform a sequence of timed feature vectors into a low dimensional latent space and it can predict the next position based on the current one. The naïve Bayes classifiers are used to classify pedestrian actions based on 3D joint positions.

2.4.3 Feature Selection Methods

Bonnin *et al.* [27] proposed a generic context-based model to predict pedestrians behavior according to features describing their local urban environment. To learn about interactions between autonomous vehicles and pedestrian interactions, in [33], Camara *et al.* collected data from real-world pedestrian-vehicle interactions at an unsignalized intersection. The actions of pedestrians and vehicles were ordered into sequences of events comprising descriptive features and the study revealed the most predictive features in a crossing scenario such as the head direction, the position on the pavement, hand gestures etc. In [40], these features were filtered over time to predict whether the pedestrian would first cross the intersection or not. Völz *et al.* [213] [214] proposed a model that can predict whether or not a pedestrian will cross the street with a set of features learnt from a database of LIDAR pedestrian trajectories that are used as inputs for a support vector machine (SVM).

2.5 Effects of Pedestrian Class on Trajectory

The models reviewed so far consider all pedestrians to be alike, but human drivers interacting with pedestrians may consider their attributes as members of stereotypical classes. Membership of various demographic and psychological state classes may be predictive of their behaviour. This section first reviews findings from the psychological literature suggesting what such classes could be usefully predictive of behaviour, if it was possible to classify them automatically from autonomous vehicles. Rasouli and Tsotsos reviewed pedestrian demographics for interactions with autonomous vehicles and argued that knowing such information could help AVs, cf. Sec. III. 1. in [167]. Figure 3.4 presents a set of pedestrian attributes used for behaviour modelling.

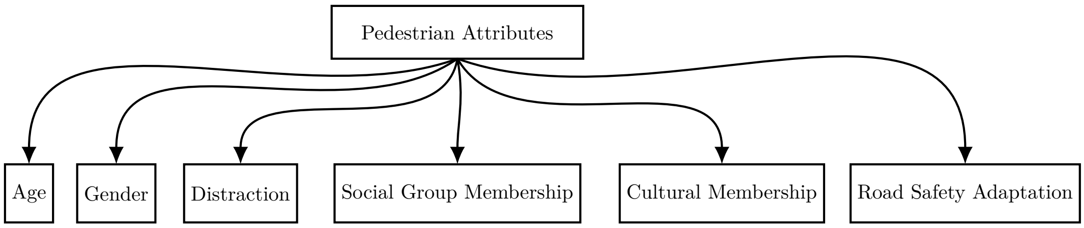


Figure 3.4: Pedestrian behaviour attributes.

Effects of Age and Gender Wilson *et al.* [219] performed a large-scale study on adult pedestrian crossing behavior and concluded that elderly people take more time and have more head movements during the crossing. Evans *et al.* [71] used the Theory of Planned Behavior (TPB) [1] via a questionnaire to predict adolescents' intentions during a hazardous road-crossing scenario. Their results show that older and male adolescents had stronger intentions to cross and that moral norms do not have any influence on crossing decisions. Pedestrians who considered themselves as safe pedestrians were less likely to cross and the anticipated affective reactions were important. Bernhoft and Carstensen [21] compared the crossing preferences and behaviour of elderly pedestrians and cyclists (age 70+) to younger people aged 40-49. It was found that elderly people have a preference for road facilities that they consider to be safer such as pavements, pedestrian crossings, signalized intersections, cycle paths. The differences between the two groups are said to be related to health and physical abilities of the people rather than their differences in age and gender.

Several studies have shown that older pedestrians have a larger accident rate than younger people [219]. Gorrini et al. [83] also found differences in adults and elderly people crossing behaviour. The study of Oxley *et al* [153] showed that older pedestrians have more risky crossing behavior in complex traffic environments than younger people. Not surprisingly, many authors have found decreasing crossing speeds with age [10] [130], [202], compensated for by requiring larger time gaps in traffic before commencing crossing [130]. In addition, Avineri *et al.* [10] found lower crossing speeds for female than male pedestrians, and that the fear of falling in elderly pedestrians has an effect on the number of downward head pitches during crossing. Holland and Hill [96] used the TPB for pedestrians' intention analysis while crossing the road. The results showed that women perceived more risk and were less likely to cross than men. In [95], they also studied the effect of gender on pedestrian crossing behaviour and showed that men with a driving experience make safer crossings than non-drivers and that older women were found to make more unsafe crossing decisions than younger women.

Distraction Distraction of pedestrians from traffic environments would ideally be defined via their mental state i.e., thinking about a problem unrelated to their environment; or approximated in practice via observable proxies. While it is possible that mental distraction might be measurable via hard-to-observe proxies such as gaze direction or high-level body language, it may be more practical to look instead for known *causes* of distraction. Schwebel *et al.* [191] performed a study in a semi-immersive virtual pedestrian street with college students, finding an impact of talking on mobile phones on crossing behaviour. Walker *et al.* [215] showed that male pedestrians using a personal music device were more cautious in crossing than those who were not distracted. In [200], the effects of personal electronic device usage on crossing behavior is studied. The results show a third of the observed pedestrians were distracted by their mobile phone and that distracted pedestrians are more likely to have unsafe crossing behaviour and walk much faster than undistracted pedestrians.

Social Group Membership Group membership can affect road crossing. Three strangers in a group are less likely to assert in a crossing than three friends. In particular, group size influences a lot crossing behavior [167]. Zeedyk *et al.* [226] performed a study with adult-child pairs while crossing the road at a pedestrian crossing. They found that adults were more likely to hold girls' hands than boys'.

Cultural Membership In contrast to the above membership of short-term, physically present groups, it is also possible to consider 'cultural membership' of a pedestrian to any long-term, non-physically present group that may be usefully predictive of behaviour. For example, it might be possible for a human driver or autonomous vehicle to classify pedestrians as members of religious, sporting, or musical (sub)cultures as a probabilistic function of features of their clothing such as shape and colour of garments or symbols displayed on them; and that members of such groups show statistically significant differences in assertiveness, politeness, and other road interaction behaviours (cf. [167]). In Sociology, classifications of individuals into cultures is notoriously problematic and politicised. But for the purpose of predicting road interactions, *any* classification derived from observable features may be usefully considered if it improves predictions.

Road Safety Adaptation Related to the possible predictiveness of cultural clothing is the effect of road safety clothing on behaviour. Human drivers are more likely to yield to pedestrians wearing high-visibility clothing [92], so it is also possible that knowing this fact will make a pedestrian wearing such clothing more likely to behave assertively. This is an example of risk compensation *adaptation*, a well-known effect in road safety in which the owners of safety improvements make economic decisions whether to use them to reduce accidents or alternatively to gain some other advantage at the cost of retaining the original accident rates [180].

2.6 Discussion

Single pedestrian unobstructed walking path and behaviour prediction around obstacles for known origins and destinations has well-established solutions. Their main strength lies in their simplicity and ease of implementation but their applicability to solve real AV problems is very limited due to the strong assumptions (e.g static obstacles, known origin-destination of pedestrians) which are not easily verified in the real world. But when – as is usual in real-time systems – the destinations of pedestrians are not known in advance, trajectory prediction is harder and remains an open research area.

Uncertain destination models may use known destination models as a subcomponent and average over them weighted by predictions about what the destination is. To predict what a pedestrian's destination will be, many medium and high-level sources of information may be relevant and useful, if suitable models can be found. These models split roughly into short-term models for prediction horizons around 1-2s and long-term models predicting for a horizon of around 5-6s. Event-based models of activity assume that behaviour often contains repeated stereotypical chunks of behavior, which once recognised in early stages can predict their later stages. The major emerging long-term prediction methods rely on neural network ('deep learning') methods. There is a need to verify how the data-driven methods such as [6] can be actually applied online for real-time systems. These models can help AVs to more accurately predict single pedestrian behaviour for shorter or longer time horizons, e.g. to know precisely whether a pedestrian's trajectory would interfere with the AV's own path. But their main challenges lie in their computational cost, which increases significantly with the number of destination guesses, with longer time horizons and the amount of data needed for learning pedestrian motion patterns. Moreover, deep learning models are sometimes referred to as 'black-box models', in the sense that AI developers cannot fully explain some decisions (e.g. feature selection) made by the neural networks, rendering them potentially problematic for investigating the causes of incidents involving AVs and for determining their liabilities [43][85].

Single pedestrians' destinations and behaviours may be informed by their class memberships, including their demographics and other visible features, such as clothing types. There are many recent sociological studies giving evidence of these effects, but they have not yet been translated into algorithms suitable for autonomous vehicle use, which would be a promising new research area. It is conjectured that additional information about pedestrians' emotion states would be similarly informative (e.g. angry pedestrians more likely to assert themselves in competitions for road space), but no studies were found in this area. Traditionally, emotional state has been difficult to capture and record, so that manually annotation of data sets are too small for machine learning to use. But as machine vision for face and body language recognition continues to improve (cf. Part I [35] Sect. 4), they are expected to produce big data sets which will enable machine learning to operate and inform destination and behaviour predictions.

3 Pedestrian Interaction Models

So far, only path prediction models for single pedestrians in static environments ignoring interactions with other pedestrians have been reviewed. This section will consider models of interaction between pedestrians. In Social Science, pedestrian behavior models have been studied for a long time: a survey is provided in [42] [201]. These models can be classified in two categories, namely microscopic models and macroscopic models, as reviewed in [211]. Microscopic models model only each pedestrian individually. Macroscopic models do not model individual pedestrians and instead model the behaviour of a single aggregate entity such as a "crowd" or a "flow". Papadimitriou *et al.* [154] presented a review on pedestrian behavior models and a study on pedestrian and crowd dynamics was proposed by Vizzari and Bandini in [212]. Bellomo *et al.* [15] reviewed mathematical models of vehicular traffic and crowds while Duives *et al.* [66] surveyed pedestrian crowd simulation models. Figure 3.5 presents a summary of pedestrian microscopic and macroscopic models.

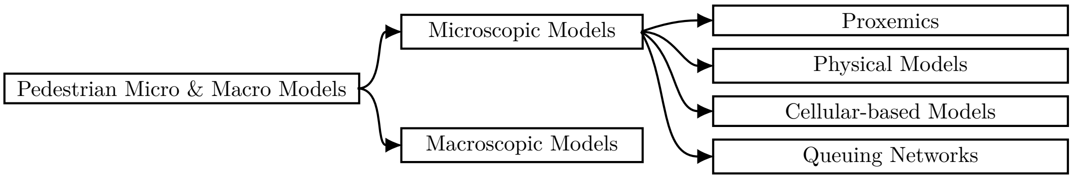


Figure 3.5: Pedestrian microscopic and macroscopic models.

3.1 Microscopic Models

This section first describes pedestrian behaviour models at the microscopic level. It then presents pedestrian interaction models using these behaviour models for two agents' interactions and group behaviour modelling.

3.1.1 Behaviour Models

Microscopic models are divided into three main groups: physical models, cellular-based models and queuing network models. Each model is generally structured by two terms: one term that represents the attractive effects of pedestrians toward their goal and the other repulsive effects among and between pedestrians and the obstacles [42]. Proxemics is first described in this section.

Proxemics The Psychology theory of Proxemics [91] studies human preferences (utilities) for having other humans in their proximity. Proxemics typically identifies four radial comfort zones, whose radii differ between cultures, for intimate, personal, social, and public space. These zones can be described by eight dimensions [91]:

1. postural-sex identifiers
2. sociofugal-sociopetal orientation (SFP axis)
3. kinesthetic factors
4. touch code
5. retinal combinations
6. thermal code
7. olfaction code

8. voice loudness scale

This model has been empirically tested with participants [217]. The theory is of great interest to pedestrian interaction models because it provides a possibly hard-wired negative utility not just for actual collisions with pedestrians but also for simply feeling too close to them. In particular, this provides a method for an AV to inflict a real negative utility on a pedestrian without touching them or risking their physical harm. Binary proxemics is the simplest case used in simple models, in which a negative utility is assigned to actually hitting someone, and zero utility is assigned to not hitting anyone. Zonal proxemics is more subtle, it relies on the eight proxemic dimensions defined above. It assigns different utilities to the presence of a person in four different zones around an individual which are defined as the *intimate distance*, the *personal distance*, *social distance* and the *public distance* [90]. Gorrini *et al.* [82] studied the proxemics behaviour of groups of pedestrians in interaction and showed that it has negative effects in walking speed for evacuation scenarios. Manenti *et al.* [139] presented an agent-based pedestrian behaviour model that takes into account proxemics and group behaviour. Their model was tested with groups of people and in a simulated environment. A detailed review on proxemics models for robot navigation among humans is proposed in [173].

Physical Models These are splitted into three sub-categories. The *utility maximization model*, as used in [114], assumes that pedestrian behaves such as to maximize their utility, for example their speed of motion and approach or avoidance of some objects or persons. In the *magnetic force model* proposed by [151], the pedestrian behavior is determined by the equation of motion of the magnetic field. Pedestrians are positive poles and their destinations are negative poles. In the *social force model*, introduced by Helbing [93], each pedestrian has a desired velocity, a target time and a target destination which are affected by social forces such as the interaction with other pedestrians and the effects of the environment. In [134] social forces are described as individual forces (fidelity, constancy) and group forces (attraction, repulsion, coherence). Most of the time, social forces are modeled such that to minimize an energy objective which include terms for individual and group forces.

Cellular-based Models These represent a cost model such as Blue and Adler's cellular automata model [23] and used for motion prediction. Cellular Automata (CA) is a discrete, time based modelling formalism on a regular cell grid. It describes the walk of a pedestrian according to rules of a cell occupancy, e.g. a cell can be occupied only if it is free and a pedestrian can have three possible movements: lateral, longitudinal or mitigation of the conflicts. The benefit cost model, developed by Gipps and Marksjo [79], is a discrete and deterministic model where the space is divided into a grid of cells and each agent is described as a particle in a cell. A benefit value, equivalent to the pedestrian utility, is arbitrarily assigned to each cell. In [60] a cellular automata model simulates multi-agent interactions.

Queuing Network Models They have been developed for studies of evacuation dynamics [131]. These are evaluated by Monte Carlo simulation methods for discrete events. Each pedestrian is represented as an individual flow entity interacting with other objects, facilities are modeled as a network of arches for openings and of nodes for rooms. In [13], a queuing network model is compared to a social force model for pedestrian crossing movement prediction.

3.1.2 Two Agents' Interaction

These models are those involving only two agents with mutually influencing behaviours, rather than larger groups of agents. They may be simpler than larger group models but sometimes provide a foundation for extension to larger group models, hence they are here presented first.

Dynamic Graphical Models The method in [31] uses POMDPs (Partially Observable Markov Decision Processes) with a time-indexed state space to model interactions and they used the example of an elevator-riding task to test the model. In [179], Rudenko *et al.* proposed a method that uses MDPs with a joint random walk stochastic policy sampling algorithm to predict motion and social forces to model interactions. The model in [120] learns features from observed pedestrian behaviors using a Markov Chain Monte Carlo (MCMC) sampling and performs a Turing test with human participants to validate the human-like behavior of the model. Chen *et al.* [48] used an extended Kalman filter to predict future motions of pedestrians and estimate the time-to-collision range (TTCR) for collision risk level identification.

Gaussian Process Methods Kawamoto *et al.* [112] proposed a method to learn pedestrian dynamics with kriging, the most traditional form of Gaussian Processes. Their work can predict pedestrian movement using spatial kriging and spatio-temporal kriging. Social interaction is modeled by spatio-temporal correlation of pedestrian dynamics and correlation is estimated by kriging.

Deep Learning Methods Alahi *et al.* [2] predicted pedestrian trajectories in crowded spaces using a social LSTM, a variant of recurrent neural network model that can learn human movement (velocity, acceleration, gait...) taking into account social human motion conventions and predict their future trajectories. This technique is opposed to traditional social forces methods and outperforms most the state-of-art methods on public datasets (ETH and UCY). Long Short-Term Memory (LSTM) can learn and reproduce long sequences, it is a data-driven technique. One LSTM is used for each person and the interaction among people is modeled by a social pooling layer which allows the share of states between neighboring LSTMs. Although group behavior is not modeled, the social LSTM can predict it very well. Similarly to the previous method, Chen *et al.* [49] developed a long-term pedestrian prediction model using RNNs for pedestrian trajectory prediction.

Road Crossing Models This section extends the event-activity models from section 2.4 by adding interaction between pedestrians and vehicles. When microscopic models of pedestrian movement are included in larger-scale traffic simulations together with vehicles, they are typically extended with specific provisions to account for pedestrian's decisions on where and when to initiate road crossing, when this is needed for the pedestrians to reach their goals. Other, so called *gap acceptance* models, have instead described probabilities of pedestrians crossing in a certain gap between vehicles, using generalised linear models, with predictors including both the available gap itself, as well as other factors such as age and gender of the pedestrians, number of pedestrians waiting to cross, and time spent waiting [195, 188].

Markkula *et al.* [140] proposed another type of model for pedestrian's road crossing decision, modelled as the result of a number of perceptual decisions concerning the available gap, but also car yielding, explicit communicative signals from the car, and eye contact with the driver. These decisions were described as several interconnected evidence accumulation processes, and it was shown that empirically observed bimodal distributions of pedestrian waiting time were qualitatively reproduced by the model. In [34], Camara *et al.* proposed a heuristic model for pedestrian crossing intention estimation. Their method is based on a distance ratio model that computes the pedestrian crossing probability over time until the curbside. Their results showed that this heuristic model is sufficient for most of the crossing scenarios present in the dataset used and that the remaining scenarios would require higher level models such as game theory.

Other Methods Discrete choice models [8, 28] offer a framework to model pedestrian walking along link levels, where their paths are composed of a sequence of straight lines in absence of obstacles. For example, the model in [28] predicts pedestrian behaviour in the presence of other people in shopping street areas.

3.1.3 Group Interaction

A *group* is here considered to be a collection of more than two pedestrians, but smaller and more cohesive than a crowd. These models are developed primarily for use by non-carriageway autonomous vehicles, such as delivery robots, navigating through crowded pedestrianized areas, needing to cut their way between groups.

Dynamic Graphical Models In [19] a real-time pedestrian path prediction is performed in cluttered environments without making any assumption on pedestrian motion or pedestrian density. Pedestrian motion and movements patterns are learnt from 2D trajectories. Bera *et al.* used sparse and noisy trajectories data from indoor and outdoor crowd videos. By combining local movements (microscopic and macroscopic motion models) and global movements (movement flow), the patterns help improve the accuracy of the long-term prediction. An ensemble Kalman filter (EnKF) was used to predict the next state based on current observation and EM algorithm to maximize the likelihood of the state. Pedestrian clusters are computed based on their positions, velocities, inter-pedestrian-distances, orientations etc. Global movement patterns are the past movement and intended velocity of pedestrians. Local movement patterns are obtained by fitting the best motion model to pedestrian clusters and individual motions. In [20], the same authors implemented a tracking algorithm built on top of [19]. Deo *et al.* in [55] uses VGMMs to model pedestrian trajectory using pedestrian origins and destinations. Their model is tested on a dataset of a crowded unsignalized intersection in a university campus. Pellegrini *et al.* [156] introduced a linear trajectory avoidance (LTA) model which has similarities with the social force model. In [157], the same authors extended the LTA model with a stochastic version taking into account group behavior and allows multiple hypotheses about the pedestrian position. Zhou *et al.* [231] proposed a mixture model of dynamic pedestrian-agents (MDA) for pedestrian trajectory prediction in crowds.

Gaussian Process Methods Henry *et al.* [94] used inverse reinforcement learning (IRL) to learn human-like navigation behavior in crowds. The model estimates environmental features using Gaussian Processes and extends Maximum Entropy Inverse Reinforcement Learning (MaxEnt IRL) of [232] by assuming that features in the environment are partially observable and dynamic. The proposed approach was developed for mobile robot motion planning, but it could be used for human motion prediction. In [203], Trautman and Krause proposed to solve the freezing robot problem, where a robot motion planner gets stuck and cannot find any proper move to perform, by a model based on Gaussian Processes, a statistical model that is able to estimate crowd interaction.

Deep Learning Methods The subsequent models may not explicitly consider interaction, but they learn interaction implicitly through machine learning techniques. The model in [196] implemented a real-time Temporal 3DOF-Pose Long-Short-Term Memory using 3D lidar data from a mobile robot. Shi *et al.* [193] developed a long-term pedestrian trajectory prediction model for crowded environments using LSTM. In [224], Yi et al. proposed a deep neural network model called behavior-CNN that is trained with crowded scenes video data. A pedestrian behavior model is encoded from the previous frames and used as an input for the CNN model to predict their future walking path and destination as well as a predictor for a tracking system. Radwan *et al.* [164] presented an interaction-aware TCNN, a convolutional neural network model that can predict interactive motion of multiple pedestrians in urban areas.

Amirian *et al.* [6] predicted the motion of pedestrians over a few seconds, given a set of observations of their own past motion and of those of the pedestrians sharing the same space, using a Generative Adversarial Network (GAN)-based trajectory sampler. The reason for this choice is that such a method naturally encompasses the uncertainty and the potential multi-modality of the pedestrian steering decision, which is of great importance when using this predictive distribution as a belief in higher level decision-making processes. Lee *et al.* [122] developed *DESIRE* a trajectory prediction framework for multiple interacting agents based on deep neural networks. A conditional variational auto-encoder is used to generate hypothetical future trajectories. An RNN is then used to score and rank those features in an inverse optimal control manner and taking into account the scene context. Gupta *et al.* [86] proposed a socially-aware GAN with RNNs for pedestrian motion sequence prediction in dynamic environments. However, their model assumes that people influence each other uniformly. A detailed analysis and improvement of this GAN method is proposed in [118]. With a similar method, called SoPhie, Sadeghian *et al.* [181] developed a GAN-based trajectory prediction model that focuses on the most important agents' for each interacting agent.

Other Methods Moussaid *et al.* [148] presented a heuristics-based model to predict pedestrian behavior in crowded environments. Based on the idea that visual information is very important for pedestrians [12, 204], they found that two simple heuristics can model the interaction among people: the desired walking direction and speed of pedestrians are sufficient. Bonneaud and Warren [26] proposed a related type of model, extending the behavioral dynamics model by [72] to goal-seeking and obstacle avoidance in crowds, and found that the model was able to reproduce qualitative crowd phenomena like lane formation. The model in [101] learns behavioral patterns from pedestrian trajectories in a mall. It assumes that a robot can model interactions using social forces and segment pedestrian trajectories into sub-goals to estimate their future positions.

3.2 Macroscopic Models

In macroscopic models, the crowd is modeled as a single ontological object, replacing and simplifying the representation of multiple microscopic pedestrians. The crowd behaves as a continuous fluid with a flow average speed [199].

The first macroscopic models of pedestrians are due to Hughes and Henderson [99]. The fluid dynamic model classifies pedestrians into groups which are characterized by average features, their position, speed and intended velocity. In [14], pedestrian flows are modeled in simulations for crowded environments. Crowd modelling has also an established community focused on models for evacuation, as reviewed in [184]. In [4] Ali *et al.* used Lagrangian Particle Dynamics to segment high density crowd flows. This method, based on Lagrangian Coherent Structures (LCS) from fluid dynamics and particle advection, is capable of detecting instabilities in the crowd.

Smooth Particle Hydrodynamics (SPH) is a hybrid of microscopic and macroscopic models. Pedestrians are considered individually, but at each time they are aggregated into a density where each particle is moved according to the macroscopic velocity. Etikyala *et al.* [70] reviewed smooth particle hydrodynamics pedestrian flow models while [225] proposed a generic SPH framework for modeling pedestrian flow.

3.3 Discussion

The theory of proxemics has been well studied in psychology and now being more and more used for VR experiments [152] [58] and computer scientists are just beginning to apply it to make more detailed models of the utility of pedestrian's personal space than simply collisions and non-collisions. In general, microscopic models are preferred to macroscopic models, in particular the social force model is very popular for pedestrian interaction modelling, while macroscopic models are more suited for crowd behaviour modeling, especially in the specialised domain of emergency evacuation modeling. Physical models bring interesting results when there are a lot of interactions, e.g. modelling pedestrian movement in cities [177]. Cellular-based models are useful for modelling pedestrians with minimal movement choices and when representing their collisions is not required. Two agents' and group interaction models offer more precise pedestrian models but they require more computational resources, in particular dynamic graphical, Gaussian Process and deep learning models. More computational research is needed in interaction modelling: psychology/human factors studies and theories are more mature, but their results have not yet been quantified to the extent of enabling translation into algorithms for AVs.

4 Game Theoretic and Signalling Models

4.1 Game Theory Interaction Models

The models in section 2 predict the behavior of a single pedestrian X from the point of view of an external observer O (i.e. the experimenter), when no other pedestrians are present. We call this a first-order model of pedestrian behaviour.

The models in section 3 all further allow O to also model X's own first-order model of another pedestrian Y's behaviour, which X can use to plan to avoid Y. We call this a second-order model of behaviour.

We could then imagine third and higher order models. For example, O might model X's belief about Y's belief about X's belief about Y's belief, as both agents try to 'out-think' each other during their planning. This would lead to an infinite computational regress.

Game theory provides an alternative and stronger framework which can compute the infinite limit of these higher order models directly, via analytic solutions.

Isaacs [106] introduced vehicle-pedestrian interactions as the famous 'homicidal taxi driver problem' which considered the inverse of the modern AV interaction problem: how an AV controller should act in order to *hit* a pedestrian[1]. Game theory is in common use in descriptive road user modelling as reviewed in [67], where applications include modelling of lane changes and merging onto motorways, route selection and departure time in congested networks, and socio-economic choices such as purchasing large vehicles or using conventions such as headlight dipping. It has been applied to AV-vehicle interactions in [165] though here only pedestrian models are considered.

The use of game theory for active control of AVs is less common. Descriptive models may be incomplete as active controllers, in particular by allowing for multiple Nash Equilibria to exist without selecting between them. A Nash equilibrium is a set of probabilistic strategies to be played by each of the players, such that no player would change their strategy if they knew the strategies of the other players. It is generally agreed in Game Theory that it is not optimal for players to employ strategies which are not Nash equilibria, though there is still philosophical debate over what strategy is optimal when multiple equilibria exist.

4.1.1 Two Agents' Game Theory Interactions

Hoogendoorn and Bovy [97] give a purely theoretic construction (left as an exercise to the reader) for a continuous ('differential') game theory solution to pedestrian interactions, based on similar control theory models to those reviewed in Sec. 2.1. They also provide an implementation of a second-order truncation of this model which is found to be sufficient to produce flows of pedestrians in crowded environments similar to those observed in some Japanese crossings.

The methods in [141] and [205] predict selection of pedestrian trajectories from a finite set as a higher-order model. For a small set of known origins and destinations, optimal free space trajectories are computed from control theory, and actual trajectories from a video set are compared to them and assigned costs according to their deviations from them. These models assume that the choice of the entire continuous trajectory is drawn from a finite set of previously observed and costed trajectories as a single decision at the start of the interaction and does not model responses to the other agent during the interaction. They are used only as descriptive models rather than as real-time control because they require each pedestrian's final goal location to be known in advance to form the cost matrix – which is only obtainable by looking ahead in the data to see what happened *post hoc.* The authors state that (in

[1] The application to pedestrians was accidental as the taxi scenario was used initially as a declassification technique to publish missile-defence algorithms, requiring control of one missile to hit another

the context of AV control), 'few researchers have considered interaction between (pedestrian) objects, thus neglecting that humans give way to each other'. Turnwald *et al.* [206] add an alternative model where one player chooses their trajectory first then the second chooses theirs in response to seeing their initial motion.

Ma *et al.* [135] proposed a long-term game-theoretic prediction of interacting pedestrian trajectories from a single starting image. For each future time in the prediction sequence, fictitious play is used to converge the probabilities of each pedestrian's actions to one (of possibly many) Nash equilibrium. The fictitious play assumes that each pedestrian has a known destination goal, some known visual features (age, gender, initial body heading etc) and a known utility function. The utility function scores vectors of word-state features which contain all of (1) the pedestrian's own future trajectory (which may include control theory style costs); (2) probabilistic beliefs about the other agents' trajectories; (3) the pedestrian's own visual features (age, heading etc); (4) proximity to static obstacles; (5) the pedestrian's distance to their goal. Unusually, the utility functions are learned entirely automatically from video data of actualized trajectories, rather than set by theories. Where theory-like behaviours such as proxemics and social forces are observed in simulations, they arise entirely from this learning process. The functions are assumed to be a weighted linear function of the features and a reinforcement-learning-style model is used to obtain per-state values from the full trajectories during learning. A (deep learning) classifier is used to obtain the visual demographic and heading features from annotated training examples. Performance is degraded when the pedestrian's goal locations are not known and are set to be completely uncertain in the feature vectors.

In [75], Fox *et al.* presented a version of the game-theoretic 'game of chicken' for autonomous vehicle-pedestrian interactions at unsignalized intersections. The obtained discrete model called the 'sequential chicken' model allows two players to choose a set of two speeds: decelerate or continue. A new method to compute Nash equilibria is presented, called 'meta-strategy convergence', used for equilibrium selection. Camara *et al.* [41, 36] evaluated the model [75] by fitting one parameter θ to controlled laboratory experiments where pedestrians were asked to play sequential chicken. This behavioural parameter θ was found to be a ratio between the utility of avoiding a collision and the utility of saving time. A summary of the work using the sequential chicken model is provided in [39].

4.1.2 Small Group Game Theory Models

Vascon *et al.* [208] proposed a game theory model for detecting conversational groups of pedestrians from video data, based on the socio-psychological concept of an F-formation and the empirical geometries of these formations. Johora and Müller [109] proposed a three-layer trajectory prediction model composed of a trajectory planner, a force-based (social force) model and a game theoretic decision model. The game theory model is based on Stackelberg games, a sequential leader-follower game where pedestrians have three different possible actions: continue, decelerate and deviate and the car has two possible actions: continue and decelerate. This model is able to handle several interactions at the same time.

4.1.3 Crowd Game Theory Models

Mesmer *et al.* [143] modelled pedestrians' decision-making and interactions during evacuations with game theory. In [192] a model of pedestrian behavior in an evacuation used game theory and showed that pedestrians get greater benefits by cooperating.

4.2 Signalling Interaction Models

Signalling models extend interaction models by allowing both the pedestrian and the AV to model and predict each other's actions of giving and receiving pure information, rather than communicating only through their physical poses.

Nathanael *et al.* [149] has proposed a stratified model of mutual awareness between pedestrians and vehicles including AVs. The actor's awareness is divided into three levels, i.e., (1) unaware of the others, (2) factually aware of the other, or (3) aware and actively attending to the other. When one of the two agents is unaware of the other, the interaction may be as simple as collision avoidance by the one aware, relying only on bodily and kinematic cues. When both agents are aware of each other, the interaction takes the form of mutual coordination through implicit cues, whereas when both agents are attentive to each other (as evidenced through eye contact between human actors), the interaction may involve direct communication through explicit signals, such as gestures, nodding etc. In addition attentiveness, as opposed to mere awareness, designates that any physical action from an attentive agent is a response explicitly addressed to the agent at the focus of attention (i.e. it also has a signalling function).

This line of research raises an epistemological question about signalling-based interaction. Some of the models above involve the concepts not just of an agent (1) knowing that the other agent is there, and (2) acting to show the other agent that they are present; but also higher-order knowing and showing these facts. This includes (3) knowing that the other knows they are there and (4) showing the other that they know that the other knows they are there. But also includes arbitrarily higher orders, such as 'knowing that the other has showed that they know that they know that the other knows' and so on. There appears to be a potentially infinite regress here, though intuitively most humans find it difficult to comprehend many more levels than the four mentioned here. But it is difficult to argue for why any cut-off should occur at this or other specific level. Intuitively: when two agents make eye contact, they assume that they both then come to know the infinite stack of such statements about each other.

4.2.1 Signals from Pedestrian to Vehicle

The need for precise eye contact as opposed to simple head direction or gaze towards the vehicle is controversial. Considering gaze or head orientation towards vehicles, there is evidence that pedestrians who initiate crossings without looking at the oncoming vehicle tend to make drivers more attentive to them by keeping larger safety margins [111]. On the other hand, eye contact between pedestrian and driver tends to increase the probability of the vehicle yielding for pedestrians [84]. The apparent controversy between these findings may be attributed to profound differences in the function of these two behavioural traits. While head orientation towards vehicles typically signifies pedestrian situational awareness to drivers, eye contact most probably signifies driver awareness of the pedestrian to the latter [169]. In addition, eye contact is reported to play a non-trivial role in the social dynamics between the two. Nathanael et al. [149] in a naturalistic study of driver pedestrian interaction reported that pedestrian head turning towards a vehicle was sufficient for drivers to confidently infer pedestrians intent in 52% of interaction cases observed. In retrospective think-aloud sessions of their interaction with pedestrians, drivers mentioned pedestrian active head movement and orientation as an important indication of pedestrian awareness of their vehicle. Mutual eye contact between driver and pedestrian was observed only in 13% of interaction cases, accompanied by explicit signalling in 2% of total cases. This is consistent with recent research [168] that reported head orientation/gaze towards vehicles as the most prominent cues for predicting pedestrian intent. In addition, computational models have shown that head direction is a useful trait for pedestrian path prediction and state of situation awareness such as in [33] which argued that if a pedestrian looks at the vehicle, they are less likely to cross the road.

Matthews *et al.* [142] studied pedestrians' behavior with an autonomous goal car equipped with an Intent Communication System (ICS) based on Decentralized MDP to model the uncertainty associated with pedestrian's behavior. Another important factor to take into account is the poor pedestrian signal settings. It has been proven that signal indication and timing affect significantly pedestrian behavior and their crossing decisions [3] [104] [105]. Pedestrians can have sudden speed change while crossing, and such sudden behavioral changes may not be expected by conflicting vehicles, which may lead to hazardous situations. In [105], Iryo-Asano and Alhajyaseen proposed a discrete choice model and Monte Carlo simulation for generating pedestrian speed profiles at crosswalks. In [103], the same authors modelled pedestrian behaviour after the onset of pedestrian flashing green (PFG) via a Monte Carlo simulation. Their results showed a higher probability of pedestrian stopping at longer crosswalks and a significant difference in pedestrian speeds.

Some early steps have however been taken towards modelling at least some levels of explicit knowing and showing of beliefs about each other via signalling behaviour.

4.2.2 Signals from Vehicle to Pedestrian

Beyond understanding pedestrian's signalling behavior, game theoretic models may also enable the AV to give signals to the pedestrians, creating a higher level information game with both players communicating through both their physical actions and also their signals. The full game theory of such interactions has not yet been worked out, and will form part of a complex socio-technical system [175], but there has been notable activity – especially via company patents – in researching displays and other mechanisms for the signalling itself.

Lundgren et al. [133] showed that the lack of two-way communication between driver and pedestrian may reduce pedestrians' confidence to cross the street and their perceived feeling of safety, when crossing. Lichtenthäler *et al.* [129] reviewed robot trajectories among humans, including identifying needs for additional gestures or motion information such as gaze to communicate intention, which is relevant for last mile delivery. Researchers are currently conducting studies to better understand exactly which information needs to be transferred when interacting with an AV. Schieben et al. [185] propose the following information to be considered by the design team.

- Information about the vehicle automation status
- Information about next manoeuvres

- Information about perception of environment
- Information about cooperation capabilities

To transfer the relevant information, two means of communications can be used for shaping the communication language of an AV. First, pedestrians might benefit from direct communication through the means of external human machine interfaces (eHMIs) [133], [178]. Secondly, also careful design of vehicle movement can be used to explicitly communicate. Risto et al. introduced the term 'movement gestures' and found 'advancing', 'slowing early' and 'stopping short' as commonly used gestures [175]. Consistent with this, Portouli et al. [160] in the context of driver-driver interaction have shown that 'edging' was explicitly used by drivers trying to enter a two-way street as a sign of their intent to inform oncoming cars. Studies of human robot interaction have shown that allowing humans to anticipate robot movements by explicit communication through movements of the robot's head raises perceived intelligence of the robot even if it did not succeed completing it's intended tasks [197], thus overcoming potential machine error through the means of explicit communication. These studies might suggest similar devices such as head-like and eye-like displays for AVs.

While Clamann et al. [51] found mixed influences of explicit communication through novel eHMI on crossing behavior in dynamic traffic situations and argued that pedestrians will largely rely on legacy behavior and not on eHMIs, Habibovic *et al.* [89] found that traffic participants feel calmer, more in control and safer when an eHMI was present on an AV. Petzoldt, Schleinitz, and Banse [158] found that an eHMI can help to convey the intention of a vehicle to give priority to a pedestrian. They also observed that pedestrians needed more time to understand the intention of a vehicle without eHMI in mixed traffic situations [158]. Communicating the intent and awareness of automated vehicles has been considered in a positive way [137] [138]. Habibovic et al [89], [7] argued that, for safety reasons, communication should never be command-based. The vehicle should communicate solely its intentions.

Communication can be directed or undirected. Pedestrians usually assume that any AV's communication is referring to themselves, hence using eHMIs with multiple pedestrians present has to be carried out in a way that minimizes miscommunication (i.e. either letting all pedestrians pass or not displaying a signal at all). Directed signalling minimizes this risk as other road users do not visually perceive the signal of the eHMI. Dietrich *et al.* [59] found that pedestrians were not able to distinguish whether an undirected light signal was addressed to themselves or other traffic participants.Therefore, AVs should either use directed communication in ambiguous situations involving multiple pedestrians or no communication at all, as pedestrians will base their crossing decision on the approaching vehicle's kinematics if no eHMI is present. The color of the visual eHMI stimulus may be of importance [218].

The most common eHMI display types are projection, high resolution displays and direct light. Semantics used include animations, concrete iconography, or text. For instance, Habibovic *et al.* developed a communication concept based on external light signals on the top of the windshield [89]. Using various light animations, the intention of the AV as well as the current driving mode such as 'I'm about to yield', 'I'm resting', and 'I'm about to start' are displayed on the LED light bar. Clamann *et al.* [51] empirically examines similar models' efficacy for giving signals to pedestrians. Further eHMI concepts include mimicking eye contacts by adding visible 'eyes' to AVs – based on the well-known tendency for humans to perceive and design faces in cars– which can communicate detection and awareness of pedestrians through eye contact [44], as well as a virtual driver's mimicking furthermore facial expressions or hand signals. In addition to the pure visual-based communication between AVs and other TPs, some concepts also consider a combination of light and audio signals, as in the Google, Uber concepts and Mercedes-Benz concept car F015.

4.3 Discussion

Game Theory has a long history of use in V2V (vehicle to vehicle) interactions in classical transport studies, as microscopic models underlying simulations of traffic flows and infrastructure design. Also multi-robot game theory systems are quite mature in robotics. These two streams have not generally been unified or applied to AV-pedestrian modelling, though this is beginning to emerge as an early research area. Like other sophisticated methods, game theoretic models can be computationally expensive and it remains unclear which of their theoretical solutions will have computationally tractable algorithms.

Signalling models remain a distant research frontier. Physical actuators for eHMI signalling are currently being investigated by car manufacturers and recent years have seen much patent activity in the area. But how to best use them to transmit information is not understood. There are currently no game-theoretic models using knowing and showing with explicit signalling but this would appear to be a fruitful area for future research. Eye contact is a particular form of signalling, but even in high level psychology research there remains an ongoing and lively debate about whether it is relevant or useful. The signalling methods reviewed here are mainly from qualitative studies, some work is still needed to implement their findings in algorithms for AVs.

Most of the eHMI concepts presented here do not yet include detailed user studies and thus there remains a need for thorough evaluation including the behavioral and emotional responses of pedestrians in realistic environments. Different findings might be due to different eHMI concepts, diverse traffic scenarios, as well as different communication strategies. While research is still lacking in full understanding of the effects of eHMI on traffic, a large number of conceptual solutions have been proposed. Their influence on pedestrians, regarding their safety, experience and acceptance remains unclear. Most of these conceptual solutions are proposed by industry and involve some form of visual communication as the visual channel is the currently most used channel of communication in traffic as well as the best suited for communication at larger distances in busy environments.

5 Experimental Resources

5.1 Pedestrian Datasets

Large data sets are important resources for training and testing models at all levels, especially when they are annotated with 'ground truth' information by humans. Their use has been common for low-level models such as detection and tracking, though there is currently a shortage of high quality annotated data for the higher-level models such as social interactions.

Major visual pedestrian datasets include the Caltech Pedestrian Benchmark [61], ETH [69], TUD-Brussels [220], Daimler [68], Stanford Drone Dataset [176], UCY Zara pedestrian dataset [125] and INRIA [53]. CityPersons [228] is a large dataset for pedestrian detection. Town Center Dataset [16] is a video dataset composed of 71.5k annotations.

Datasets used for pedestrian re-identification, i.e. having many images of the same people with identifiers include for example CUHK01 [128], CUHK02 [127] and CUHK03 [126], collected at a university campus and composed of thousands of bounding boxes of unique people. DUKEMTMC [174] and DUKEMTMC-reID [230] datasets have been developed in the Duke university campus and are used for tracking and re-identifying multiple people with multi-camera systems. MARKET-1501 [229] dataset provides 35k images of 1500 individuals but also comes with a 500k dataset of non-pedestrian street window distractors for training classifiers. Multi-Object Tracking Benchmark [144] collects diverse datasets and publishes new data. Several releases have already appeared: MOT15, MOT16 and MOT17.

PETA benchmark [54] is a mixture of several public datasets (e.g VIPER, SARC3D,

PRID, MIT, I-LID, GRID, CAVIAR4REID, 3DPES), which has been used to recognize pedestrian attributes at far distance. The benchmark has been tested with an SVM method. Social ground truth annotations are much rarer. [40] and [33] collected high quality human annotations of physical and social events during pedestrian-vehicle interactions, including the presence and timings of the agents communicating with each other via eye contact, hand gestures, positions and speeds, and the final 'winners' of interactions which compete for road space during crossings.

Yang *et al.* [223] pointed out that in mixed urban scenarios, intelligent vehicles (IVs) have to cope with a certain number of surrounding pedestrians. Therefore, it is necessary to understand how vehicles and pedestrians interact with each other. They proposed a novel pedestrian trajectory dataset composed of CITR dataset and DUT dataset, so that the pedestrian motion models can be further calibrated and verified, especially when the vehicle's influence on pedestrians plays an important role. In particular, the final trajectories of pedestrians and vehicles were refined by Kalman filters with linear point-mass model and nonlinear bicycle model, respectively, in which xy-velocity of pedestrians and longitudinal speed and orientation of vehicles were estimated.

Zhan et al. proposed INTERACTION dataset [227] which contains naturalistic motions of various traffic participants in a variety of highly interactive driving scenarios. Trajectory data was collected using drones and traffic cameras, containing data from multiple countries (USA, China, Germany and Bulgaria). There are four different driving scenarios, with their semantic maps provided: roundabouts, un-signalized intersection, signalized intersection, merging and lane changing. Chang *et al.* proposed Argoverse [45] containing two datasets and HD maps recorded from a self-driving car. Argoverse 3D Tracking is for 3D object annotations, it contains a collection of 11,052 tracks, and Argoverse Motion Forecasting is a curated collection of 324,557 scenarios, each 5 seconds long, for trajectory prediction. Each scenario contains the 2D, birds-eye-view centroid of each tracked object. ApolloScape dataset [216] was recorded in urban areas in China using various sensors. The dataset contains different road road users (vehicles, pedestrians, bicycles). The ApolloScape LeaderBoard shows the ranking and performance of the models tested on the dataset for different tasks, such as scene parsing, detection/tracking, trajectory prediction, self-localisation. The Intersection Drone (InD) dataset [25] contains naturalistic vehicle trajectories recorded using a drone at four German intersections. It provides the trajectories for thousands of road users and their types (e.g car, pedestrian, bicycle, truck), and can be used for example for road user prediction.

Person detection in off-road agricultural vehicle environments has become popular in recent years. Results from these studies are not well known in transport research but may transfer to on-carriageway and on-pavement AVs as they deal with similar types of pedestrian interactions. The National Robotics Engineering Center (NREC) Agricultural Person Detection Dataset [159] consists of labeled stereo video of people in orange and apple orchards taken from a tractor and a pickup truck, along with vehicle position data. The dataset combines a total of 76k labeled person images and 19k sampled person-free images. Gabriel *et al.* [76] present a dataset that focuses on action/intention recognition problems for human interactions with small robots in agriculture, including ten actors performing nine gestures and four activities. Stereo camera images, thermal camera images and Lidar point cloud data are recorded on grassland, under varying lighting conditions and distances. Kragh *et al.* [119] presents a multi-modal dataset for obstacle detection in agriculture containing 2h of raw sensor data from a tractor-mounted sensor system in a grass mowing scenario, including moving humans scattered in the field.

A summary of pedestrian datasets is given in the supplementary material Sect. 3 Table B.2.

5.2 Vehicle Datasets

To train and test models of pedestrians interacting with vehicles, it is most likely useful to provide similar big data about vehicles as well as about pedestrians. This may include ground truth information on vehicle location and motion, but also high level social annotations to use in studies of interaction with pedestrians. Visual data available includes the Berkeley DeepDrive Video (BDDV) dataset [222], currently the largest vehicle dataset publicly available with 10k hours of driving videos around the world. KITTI dataset [78] provides a one hour video of a vehicle driving in an urban environment. Caesar *et al.* [32] presented *nuScenes* a dataset for autonomous driving composed of multiple sensor data (RGB, LIDAR, RADAR) from two cities and containing 1k scenes. A summary of vehicle datasets is given in the supplementary material Sec. 3 Table B.3.

5.3 Pedestrian and Driving Simulators

Three types of relevant simulation research work exist: pedestrian, vehicle, and combined pedestrian-vehicle. Hardware designs and source code for commonly used simulators are often not made public, making it difficult for others researchers to investigate and replicate experiments. So there remains a clear need for more open-source simulators. The open source Godot game and VR engine[1] has recently matured so may soon be used for this purpose. A summary of the simulators is included in the supplementary material Sec. 3 Table B.4.

Pedestrian Simulators Pedestrian simulators are VR (Virtual Reality) based environments where pedestrian participants encounter virtual vehicles in order to study pedestrian perception and decision making subject to various oncoming vehicle behaviors [186]. For example, Camara *et al.* [37, 38] used a HTC Vive VR headset for pedestrians interacting with a game theoretic autonomous vehicle. Results showed that VR is a reliable setup for measuring human behaviour for the development and testing of AV technology. Mahadevan *et al.* [136] presented OnFoot, a VR pedestrian simulator that studies pedestrian interactions with autonomous vehicles in a mixed traffic environment. The Technical University of Munich also developed a pedestrian simulator [73] composed of a head-mounted display, a motion capture system and a driving simulator software. This setup could be connected to a driving simulator enabling multi-agent studies while extracting the participant's gait during the crossing process. The current setup utilizes Unity (with a VIVE HMD) and is sometimes coupled with VIVE Trackers for a virtual self-representation to create an immersive virtual environment enabling fast implementations and evaluation of eHMI concepts [59]. PedSim [80] is a free crowd simulation software.

Vehicle Simulators Vehicle simulators are physical platforms where drivers encounter virtual pedestrians (dummies) in order to study driver yielding behaviors in specific interaction scenarios. Simulators such as [150] studied driver-pedestrian interactions in mixed traffic environments using a driving simulator (DriveSafety's DS-600c Research Simulator). JARI-ARV (Augmented Reality Vehicles) [108] is a road running driving simulator and JARI-OVDS (Omnidirectional View Driving Simulator is a driving simulator with 360-degree spherical screen and a rocking device. The University of Iowa [207] has developed a driving simulator. A previous review on driving simulators is presented in [194].

Pedestrian-Vehicle Simulators Micro or macro simulations model both pedestrian and vehicle behavior. Most of these simulations rely on sets of behavioral rules for both agents. These simulators are primarily used for road design purposes and for policy decisions such as

[1] www.godotengine.org

the cellular automata-based simulators proposed in [74] and [132] where vehicle-pedestrian crossing behaviour is studied at crosswalks. Feliciani *et al.* [74] further evaluated the necessity of introducing a new crosswalk and/or switching to a traffic light. Chao *et al.* [46] developed a microscopic-based traffic simulator based on a force model to represent the behaviour and interactions between the road users, and aimed for autonomous vehicle development and testing. Chen *et al.* [47] proposed a simulation platform composed of several behaviour models at crosswalks for vehicle-pedestrian conflicts assessment. Gupta *et al.* [87] developed a simulation model, using Matlab and the open-source SUMO (Simulation of Urban Mobility), for autonomous vehicle-pedestrian negotiations at unmarked intersections, considering different pedestrian behaviours. Commercial products include STEPS [147] software for and Legion [124] simulating pedestrian dynamics. VirtuoCity is an example of physical vehicle-pedestrian simulators. It is composed of a pedestrian simulator, HIKER [182], which is a virtual reality 'CAVE-based' environment for pedestrian behavior analysis, a driving simulator [107] and a truck simulator for driver behaviour understanding. IFSTTAR [100] also possesses a pedestrian simulator and developed a driving simulator for driver behavior analysis and human-machine interactions, an immersive simulator for cars, motorcycles and pedestrians behavior simulation, a driving simulator with human assistive devices and a bicycle simulator.

6 Conclusions

Pedestrian sensing, detection and kinematic tracking are now well understood and have mature models as reviewed in Part I [35]. Moving from simple kinematic tracking and prediction of pedestrian motions can however depend on extremely high-level models of the state transition required by tracking and prediction. Going far beyond simple random velocity walk models, the present review has shown that there is much scope here to integrate models of pedestrians as intelligent, goal-based, psychological, active, and interactive agents at several levels.

Unlike the more mature methods reviewed in Part I, this review does not recommend particular software implementations for algorithms at these levels, because they remain active research areas rather than completed and standardizable tools. This review finds that many conceptual issues first need to be cleared, before mathematical interfaces – such as probabilities – can be created to link models at these layers, and only then standardized software development can become a reality. (The only exception to this would be for entirely end-to-end machine learning systems, which are not generally considered to be safe or practical due to their lack of transparency.)

At the level of single pedestrian modelling, there now exist good control theoretic models of optimal walking behaviour from known origin to known destination. Here, pedestrians do not usually walk in straight lines, but optimise gradual turning during walking to move in smooth curves. There has been some recent research success in inferring likely destinations from historical data and partial trajectories.

When interaction with other agents is included, models of pedestrians rapidly become more complex and much less well understood. Suboptimal models include only finite orders of epistemological models of pedestrians beliefs, raising the open question of how to handle higher order beliefs about beliefs. Recent game theory approaches have just begun to find optimal behaviours in these higher-order belief cases but only under various simplifying assumptions.

There has been a general shift away from psychology-informed models, using empirical findings such as demographics predicting behaviours, to purely big-data-driven models which learn aspects of such theories internally as black boxes, usually aiming only to predict the behaviour rather than give theoretical explanations of it.

The role of signalling between pedestrians and vehicles during interactions has been studied qualitatively, but is not yet understood at the algorithmic level. Psychologists and road safety designers have evaluated and commercialised many signalling mechanisms, such as flashing of headlights, use of horns, and custom communication light signals. Finding algorithmic strategies to make optimal use of them, and to process information from receiving signals from others, suitable for real-time AV control, remains an open and important question.

Acknowledgements

This project has received funding from EU H2020 project interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

References

[1] I. Ajzen. Attitudes and personality traits. *Attitudes, Personality, and Behavior*, pages 2–24, 1988.

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proc. of IEEE CVPR*, pages 961–971, 2016.

[3] Wael K.M. Alhajyaseen and Miho Iryo-Asano. Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. *Safety Science*, 91:351 – 360, 2017.

[4] S. Ali and M. Shah. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. of IEEE CVPR*, pages 1–6, 2007.

[5] Matthias Althoff. *Reachability analysis and its application to the safety assessment of autonomous cars*. PhD thesis, Technische Universität München, 2010.

[6] Javad Amirian, Jean-Bernard Hayet, and Julien Pettré. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs. In *Proc. of IEEE CVPR Workshops*, 2019.

[7] Jonas Andersson, Azra Habibovic, Maria Klingegård, Cristofer Englund, and Victor Malmsten-Lundgren. Hello Human, can you read my mind? *ERCIM News*, (109):36–37, 2017.

[8] Gianluca Antonini, Michel Bierlaire, and Mats Weber. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667 – 687, 2006.

[9] Gustavo Arechavaleta, Jean-Paul Laumond, Halim Hicheur, and Alain Berthoz. On the nonholonomic nature of human locomotion. *Autonomous Robots*, 25(1-2):25–35, 2008.

[10] Erel Avineri, David Shinar, and Yusak O. Susilo. Pedestrians' behaviour in cross walks: The effects of fear of falling and age. *Accident Analysis & Prevention*, 44(1):30 – 34, 2012.

[11] Haoyu Bai, Shaojun Cai, Nan Ye, David F. C. Hsu, and Wee Sun Lee. Intention-aware online pomdp planning for autonomous driving in a crowd. *Proc. of IEEE ICRA*, pages 454–460, 2015.

[12] M Batty. Predicting where we walk. *Nature*, 388(6637):19–20, 1997.

[13] Dietmar Bauer. Comparing pedestrian movement simulation models for a crossing area based on real world data. In *Pedestrian and Evacuation Dynamics*, pages 547–556. Springer, 2011.

[14] Dietmar Bauer, Stefan Seer, and Norbert Brändle. Macroscopic pedestrian flow simulation for designing crowd control measures in public transport after special events. In *Proc. of the Summer Computer Simulation Conference*, pages 1035–1042, 2007.

[15] Nicola Bellomo and Christian Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM review*, 53(3):409–463, 2011.

[16] Ben Benfold and Ian Reid. Stable multi-target tracking in real-time surveillance video. In *Proc. of IEEE CVPR*, pages 3457–3464, 2011.

[17] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. Learning motion patterns of people for compliant robot motion. *International Journal of Robotics Research*, 24(1):31–48, 2005.

[18] M. Bennewitz, W. Burgard, and S. Thrun. Learning motion patterns of persons for mobile service robots. In *Proc. of IEEE ICRA*, volume 4, pages 3601–3606, 2002.

[19] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha. GLMP - realtime pedestrian path prediction using global and local movement patterns. In *Proc. of IEEE ICRA*, pages 5528–5535, 2016.

[20] Aniket Bera and Dinesh Manocha. Pedlearn: Realtime pedestrian tracking, behavior learning, and navigation for autonomous vehicles. In *Proc. of IROS 9th International workshop on Planning, Perception and Navigation for Intelligent Vehicles*, 2017.

[21] Inger Marie Bernhoft and Gitte Carstensen. Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(2):83 – 95, 2008.

[22] Apratim Bhattacharyya, Mario Fritz, and Bernt Schiele. Long-term on-board prediction of pedestrians in traffic scenes. In *Proc. of 1st Conference on Robot Learning*, 2017.

[23] Victor J. Blue and Jeffrey L. Adler. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B: Methodological*, 35(3):293 – 312, 2001.

[24] Julian Bock, Till Beemelmanns, Markus Klösges, and Jens Kotte. Self-learning trajectory prediction with recurrent neural networks at intelligent intersections. In *Proc. of VEHITS*, 2017.

[25] Julian Bock, Robert Krajewski, Tobias Moers, Lennart Vater, Steffen Runde, and Lutz Eckstein. The ind dataset: A drone dataset of naturalistic vehicle trajectories at german intersections. *arXiv preprint arXiv:1911.07602*, 2019.

[26] Stephane Bonneaud and William H Warren. A behavioral dynamics approach to modeling realistic pedestrian behavior. In *Proc. of 6th International Conference on Pedestrian and Evacuation Dynamics*, pages 1–14, 2012.

[27] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuedderich. Pedestrian crossing prediction using multiple context-based models. In *Proc. of IEEE ITSC*, pages 378–385, 2014.

[28] Aloys Borgers, Astrid Kemperman, and Harry Timmermans. *Modeling Pedestrian Movement in Shopping Street Segments*, chapter Chapter 5, pages 87–111. 2009.

[29] Aloys Borgers and Harry Timmermans. A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical Analysis*, 18(2):115–128, 2010.

[30] Rodney Brooks. The big problem with self-driving cars is people and we'll go out of our way to make the problem worse. *IEEE Spectrum*, 2017.

[31] Frank Broz, Illah Nourbakhsh, and Reid Simmons. Planning for human-robot interaction using time-state aggregated pomdps. In *Proc. of the 23rd National Conference on Artificial Intelligence*, volume 3, pages 1339–1344, 2008.

[32] 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 multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[33] F. Camara, O. Giles, R. Madigan, M. RothmÃ¼ller, P. H. Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox. Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. In *Proc. of IEEE ITSC*, pages 2098–2103, 2018.

[34] F. Camara, N. Merat, and C. W. Fox. A heuristic model for pedestrian intention estimation. In *Proc. of IEEE ITSC*, pages 3708–3713, 2019.

[35] Fanta Camara, Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, and Charles W. Fox. Pedestrian models for autonomous driving Part I: low-level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[36] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. Towards pedestrian-AV interaction: method for elucidating pedestrian preferences. In *Proc. of IEEE/RSJ IROS Workshops*, 2018.

[37] Fanta Camara, Patrick Dickinson, Natasha Merat, and Charles W. Fox. Towards game theoretic AV controllers: measuring pedestrian behaviour in virtual reality. In *Proc. of IEEE/RSJ IROS Workshops*, 2019.

[38] Fanta Camara, Patrick Dickinson, Natasha Merat, and Charles W. Fox. Examining pedestrian behaviour in virtual reality. In *Transport Research Arena (TRA) (Conference canceled)*, 2020.

[39] Fanta Camara and Charles W. Fox. Game theory for self-driving cars. In *UK-RAS Conference*, 2020.

[40] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proc. of the International Conference on Intelligent Autonomous Systems (IAS-15) Workshops*, 2018.

[41] Fanta Camara, Richard Romano, Gustav Markkula, Ruth Madigan, Natasha Merat, and Charles Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Proc. of Measuring Behavior*, 2018.

[42] C. Caramuta, G. Collodel, C. Giacomini, C. Gruden, G. Longo, and P. Piccolotto. Survey of detection techniques, mathematical models and simulation software in pedestrian dynamics. *Transportation Research Procedia*, 25(Supplement C):551 – 567, 2017.

[43] Davide Castelvecchi. Can we open the black box of AI? *Nature News*, 538(7623):20, 2016.

[44] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. Eyes on a Car: An Interface Design for Communication Between an Autonomous Car and a Pedestrian. In *Proc. of ACM AutomotiveUI*, pages 65–73, 2017.

[45] Ming-Fang Chang, John W Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *Proc. of IEEE CVPR*, 2019.

[46] Qianwen Chao, Xiaogang Jin, Hen-Wei Huang, Shaohui Foong, Lap-Fai Yu, and Sai-Kit Yeung. Force-based heterogeneous traffic simulation for autonomous vehicle testing. In *Proc. of ICRA*, pages 8298–8304. IEEE, 2019.

[47] Peng Chen, Weiliang Zeng, and Guizhen Yu. Assessing right-turning vehicle-pedestrian conflicts at intersections using an integrated microscopic simulation model. *Accident Analysis & Prevention*, 129:211–224, 2019.

[48] Zhijun Chen, Chaozhong Wu, Nengchao Lyu, Gang Liu, and Yi He. Pedestrian-vehicular collision avoidance based on vision system. In *Proc. of IEEE ITSC*, pages 11–15, 2014.

[49] B. Cheng, X. Xu, Y. Zeng, J. Ren, and S. Jung. Pedestrian trajectory prediction via the social-grid LSTM model. *The Journal of Engineering*, 2018(16):1468–1474, 2018.

[50] Howie Choset, Kevin M. Lynch, Seth Hutchinson, George Kantor, Wolfram Burgard, Lydia Kavraki, and Sebastian Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, 2005.

[51] Michael Clamann, Miles Aubert, and Mary L Cummings. Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles. In *Proc. of TRB*, 2017.

[52] A. Cosgun, D. A. Florencio, and H. I. Christensen. Autonomous person following for telepresence robots. In *Proc. of IEEE ICRA*, pages 4335–4342, 2013.

[53] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE CVPR*, volume 1, pages 886–893 vol. 1, 2005.

[54] Yubin Deng, Ping Luo, Chen Change Loy, and Xiaoou Tang. Pedestrian attribute recognition at far distance. In *Proc. of the 22nd ACM international Conference on Multimedia*, pages 789–792. ACM, 2014.

[55] N. Deo and M. M. Trivedi. Learning and predicting on-road pedestrian behavior around vehicles. In *Proc. of IEEE ITSC*, pages 1–6, 2017.

[56] Nachiket Deo and Mohan M Trivedi. Trajectory forecasts in unknown environments conditioned on grid-based plans. *arXiv preprint arXiv:2001.00735*, 2020.

[57] Charitha Dias, Muhammad Abdullah, Majid Sarvi, Ruggiero Lovreglio, and Wael Alhajyaseen. Modeling and simulation of pedestrian movement planning around corners. *Sustainability*, 11(19):5501, 2019.

[58] Patrick Dickinson, Kathrin Gerling, Kieran Hicks, John Murray, John Shearer, and Jacob Greenwood. Virtual reality crowd simulation: effects of agent density on user experience and behaviour. *Virtual Reality*, 2018.

[59] A. Dietrich, J-H. Willrodt, K. Wagner, and K. Bengler. Projection-based external human machine interfaces â€“ enabling interaction between automated vehicles and pedestrians. In *Driving Simulation Conference*, 2018.

[60] Jan Dijkstra, Harry JP Timmermans, and AJ Jessurun. A multi-agent cellular automata system for visualising simulated pedestrian activity. In *Theory and Practical Issues on Cellular Automata*, pages 29–36. Springer, 2001.

[61] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Proc. of CVPR*, 2009.

[62] C. Dondrup, N. Bellotto, M. Hanheide, K. Eder, and U. Leonards. A computational model of human-robot spatial interactions based on a qualitative trajectory calculus. *Robotics*, 4(1):63–102, 2015.

[63] Christian Dondrup, Nicola Bellotto, Ferdian Jovan, and Marc Hanheide. Real-time multisensor people tracking for human-robot spatial interaction. In *Proc. of ICRA Workshop on Machine Learning for Social Robotics*, 2015.

[64] Paul Duckworth, Muhannad Al-Omari, James Charles, David C. Hogg, and Anthony G. Cohn. Latent dirichlet allocation for unsupervised activity analysis on an autonomous mobile robot. In *Proc. of AAAI*, 2017.

[65] Paul Duckworth, Yiannis Gatsoulis, Ferdian Jovan, Nick Hawes, David C. Hogg, and Anthony G. Cohn. Unsupervised learning of qualitative motion behaviours by a mobile robot. In *Proc. of AAMAS*, pages 1043–1051, 2016.

[66] Dorine C Duives, Winnie Daamen, and Serge P Hoogendoorn. State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies*, 37:193–209, 2013.

[67] Rune Elvik. A review of game-theoretic models of road user behaviour. *Accident Analysis & Prevention*, 62:388 – 396, 2014.

[68] Markus Enzweiler and Dariu M Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, 2008.

[69] A. Ess, B. Leibe, and L. Van Gool. Depth and appearance for mobile scene analysis. In *Proc. of IEEE ICCV*, pages 1–8, 2007.

[70] Raghavender Etikyala, Simone Göttlich, Axel Klar, and Sudarshan Tiwari. Particle methods for pedestrian flow models: from microscopic to nonlocal continuum models. *Mathematical Models and Methods in Applied Sciences*, 24(12):2503–2523, 2014.

[71] Daphne Evans and Paul Norman. Predicting adolescent pedestrians' road-crossing intentions: an application and extension of the theory of planned behaviour. *Health Education Research*, 18(3):267–277, 2003.

[72] Brett R Fajen and William H Warren. Behavioral dynamics of steering, obstable avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):343, 2003.

[73] Ilja Feldstein, André Dietrich, Sasha Milinkovic, and Klaus Bengler. A pedestrian simulator for urban crossing scenarios. volume 49, pages 239 – 244, 2016. Proc. of IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems.

[74] Claudio Feliciani, Luca Crociani, Andrea Gorrini, Giuseppe Vizzari, Stefania Bandini, and Katsuhiro Nishinari. A simulation model for non-signalized pedestrian crosswalks based on evidence from on field observation. *Intelligenza Artificiale*, 11(2):117–138, 2017.

[75] Charles W. Fox, Fanta Camara, Gustav Markkula, Richard Romano, Ruth Madigan, and Natasha Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *Proc. of VEHITS*, 2018.

[76] A. Gabriel, S. Cosar, N. Bellotto, and P. Baxter. A dataset for action recognition in the wild. In *Proc. of TAROS*, pages 362–374, 2019.

[77] Mario Garzón, David Garzón-Ramos, Antonio Barrientos, and Jaime del Cerro. Pedestrian trajectory prediction in large infrastructures. In *Proc. of International Conference on Informatics in Control, Automation and Robotics*, pages 381–389, 2016.

[78] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[79] P.G. Gipps and B. Marksjö. A micro-simulation model for pedestrian flows. *Mathematics and Computers in Simulation*, 27(2):95 – 105, 1985.

[80] Christian Gloor. Pedsim : Pedestrian crowd simulator, 2012.

[81] M. Goldhammer, S. Köhler, K. Doll, and B. Sick. Camera based pedestrian path prediction by means of polynomial least-squares approximation and multilayer perceptron neural networks. In *Proc. of SAI Intelligent Systems Conference (IntelliSys)*, pages 390–399, 2015.

[82] Andrea Gorrini, Stefania Bandini, and Majid Sarvi. Group dynamics in pedestrian crowds: Estimating proxemic behavior. *Transportation Research Record*, 2421(1):51–56, 2014.

[83] Andrea Gorrini, Luca Crociani, Giuseppe Vizzari, and Stefania Bandini. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59:269–285, 2018.

[84] Nicolas Guéguen, Sébastien Meineri, and Chloé Eyssartier. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety Science*, 75:87 – 89, 2015.

[85] David Gunning. Explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA)*, 2, 2017.

[86] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proc. of IEEE CVPR*, pages 2255–2264, 2018.

[87] Surabhi Gupta, Maria Vasardani, and Stephan Winter. Negotiation between vehicles and pedestrians for the right of way at intersections. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):888–899, 2018.

[88] Golnaz Habibi, Nikita Jaipuria, and Jonathan P How. Context-aware pedestrian motion prediction in urban intersections. *arXiv preprint arXiv:1806.09453*, 2018.

[89] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, Dennis Saluäär, and Pontus Larsson. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology*, 9:1336, 2018.

[90] Edward T Hall. A system for the notation of proxemic behavior. *American Anthropologist*, 65(5):1003–1026, 1963.

[91] Edward T. Hall. The hidden dimension. 1969.

[92] W. Andrew Harrell. The impact of pedestrian visibility and assertiveness on motorist yielding. *The Journal of Social Psychology*, 133(3):353–360, 1993.

[93] Helbing and Molnár. Social force model for pedestrian dynamics. *Physical Review E*, 51:4282–4286, 1995.

[94] P. Henry, C. Vollmer, B. Ferris, and D. Fox. Learning to navigate through crowded environments. In *Proc. of IEEE ICRA*, pages 981–986, 2010.

[95] Carol Holland and Ros Hill. Gender differences in factors predicting unsafe crossing decisions in adult pedestrians across the lifespan: A simulation study. *Accident Analysis and Prevention*, 42(4):1097 – 1106, 2010.

[96] Carol Holland and Roslyn Hill. The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations. *Accident Analysis & Prevention*, 39(2):224 – 237, 2007.

[97] Serge Hoogendoorn and Piet HL Bovy. Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods*, 24(3):153–172, 2003.

[98] Ronny Hug, Stefan Becker, Wolfgang Hübner, and Michael Arens. Particle-based pedestrian path prediction using LSTM-MDL models. *CoRR*, abs/1804.05546, 2018.

[99] Roger L. Hughes. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, 36(6):507 – 535, 2002.

[100] IFSTARR. Simulators, 2018. http://www.ifsttar.fr/en/exceptional-facilities/simulators/.

[101] Tetsushi Ikeda, Yoshihiro Chigodo, Daniel Rea, Francesco Zanlungo, Masahiro Shiomi, and Takayuki Kanda. Modeling and prediction of pedestrian behavior based on the sub-goal concept. In *Proc. of Robotics: Science and Systems*, 2012.

[102] I Institute for Advanced Motorists. *How to Be A Better Driver: Advanced Driving the Essential Guide*. Motorbooks, 2007.

[103] M. Iryo-Asano, W. K. M. Alhajyaseen, and H. Nakamura. Analysis and modeling of pedestrian crossing behavior during the pedestrian flashing green interval. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):958–969, 2015.

[104] Miho Iryo-Asano and Wael K.M. Alhajyaseen. Analysis of pedestrian clearance time at signalized crosswalks in Japan. *Procedia Computer Science*, 32:301 – 308, 2014.

[105] Miho Iryo-Asano and Wael K.M. Alhajyaseen. Modeling pedestrian crossing speed profiles considering speed change behavior for the safety assessment of signalized intersections. *Accident Analysis & Prevention*, 108:332 – 342, 2017.

[106] Rufus Isaacs. Games of pursuit. 1951.

[107] A. Jamson, Anthony Horrobin, and Robin Auckland. Whatever happened to the lads? design and development of the new University of Leeds driving simulator. In *Proc. of Driving Simulator Conference*, 2007.

[108] JARI. Simulators, 2018. http://www.jari.or.jp/.

[109] Fatema Tuj Johora and Jörg Müller. Modeling interactions of multimodal road users in shared spaces. In *Proc. of IEEE ITSC*, 2018.

[110] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto. Intent-aware long-term prediction of pedestrian motion. In *Proc. of IEEE ICRA*, pages 2543–2549, 2016.

[111] A. Katz, D. Zaidel, and A. Elgrishi. An experimental study of driver and pedestrian interaction during the crossing conflict. *Human Factors*, 17(5):514–527, 1975.

[112] K. Kawamoto, Y. Tomura, and K. Okamoto. Learning pedestrian dynamics with kriging. In *Proceedings IEEE/ACIS International Conference on Computer and Information Science*, pages 1–4, 2016.

[113] Kris M. Kitani, Brian D. Ziebart, James Andrew Bagnell, and Martial Hebert. Activity forecasting. In *Proc. of ECCV*, pages 201–214, 2012.

[114] Kay Kitazawa and Michal Batty. Pedestrian behaviour modelling. In *Proc. of the International Conference on Developments in Design and Decision Support Systems in Architecture and Urban Planning*, 2004.

[115] Jens Kober, J. Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[116] Julian Francisco Pieter Kooij, Nicolas Schneider, Fabian Flohr, and Dariu M. Gavrila. Context-based pedestrian path prediction. In *Proc. of ECCV*, pages 618–633, 2014.

[117] Markus Koschi, Christian Pek, Mona Beikirch, and Matthias Althoff. Set-based prediction of pedestrians in urban environments considering formalized traffic rules. In *Proc. of IEEE ITSC*, 2018.

[118] Parth Kothari and Alexandre Alahi. Adversarial loss for human trajectory prediction. In *Proc. of hEART*, 2019.

[119] Mikkel Kragh, Peter Christiansen, Morten Laursen, Morten Larsen, Kim Steen, Ole Green, Henrik Karstoft, and Rasmus Jørgensen. Fieldsafe: Dataset for obstacle detection in agriculture. *Sensors*, 17, 2017.

[120] H. Kretzschmar, M. Kuderer, and W. Burgard. Learning to predict trajectories of cooperatively navigating agents. In *Proc. of IEEE ICRA*, pages 4015–4020, 2014.

[121] E. Kruse, R. Gutsche, and F. M. Wahl. Acquisition of statistical motion patterns in dynamic environments and their application to mobile robot motion planning. In *Proc. of IEEE/RSJ IROS*, volume 2, pages 712–717 vol.2, 1997.

[122] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B Choy, Philip HS Torr, and Manmohan Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *Proc. of IEEE CVPR*, pages 336–345, 2017.

[123] Stéphanie Lefèvre, Dizan Vasquez, and Christian Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, 1(1):1, 2014.

[124] Legion. Legion for aisun, 1997. http://www.legion.com/legion-for-aimsun-%E2%80%93-vehicle-pedestrian-simulation.

[125] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer Graphics Forum*, volume 26, pages 655–664. Wiley Online Library, 2007.

[126] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proc. of IEEE CVPR*, pages 152–159, 2014.

[127] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proc. of IEEE CVPR*, pages 3594–3601, 2013.

[128] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Proc. of Asian Conference on Computer Vision*, pages 31–44. Springer, 2012.

[129] Christina Lichtenthäler and Alexandra Kirsch. Towards legible robot navigation - how to increase the intend expressiveness of robot navigation behavior. In *Int. Conf. Soc. Robot. Embodied Commun. Goals Intention*, 2013.

[130] Régis Lobjois and Viola Cavallo. The effects of aging on street-crossing behavior: from estimation to actual crossing. *Accident Analysis & Prevention*, 41(2):259–267, 2009.

[131] Gunnar G. Løvås. Modeling and simulation of pedestrian traffic flow. *Transportation Research Part B: Methodological*, 28(6):429 – 443, 1994.

[132] Lili Lu, Gang Ren, Wei Wang, Ching-Yao Chan, and Jian Wang. A cellular automaton simulation model for pedestrian and vehicle interaction behaviors at unsignalized mid-block crosswalks. *Accident Analysis & Prevention*, 95:425–437, 2016.

[133] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlönn, Rikard Fredriksson, Claes Edgren, Stas Krupenia, and Dennis Saluäär. Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context? In *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation*, pages 485–497. Springer International Publishing, 2017.

[134] Wenhan Luo, Xiaowei Zhao, and Tae-Kyun Kim. Multiple object tracking: A review. *CoRR*, abs/1409.7618, 2014.

[135] Wei-Chiu Ma, De-An Huang, Namhoon Lee, and Kris M. Kitani. Forecasting interactive dynamics of pedestrians with fictitious play. In *Proc. of IEEE CVPR*, 2017.

[136] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E Young, and Ehud Sharlin. AV-Pedestrian interaction design using a pedestrian mixed traffic simulator. In *Proc. of the Conference on Designing Interactive Systems*, pages 475–486. ACM, 2019.

[137] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. Can Interfaces Facilitate Communication in Autonomous Vehicle-Pedestrian Interaction? In *Proc. of ACM/IEEE HRI*, 2018.

[138] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proc. of the CHI Conference on Human Factors in Computing Systems*, page 429. ACM, 2018.

[139] Lorenza Manenti, Sara Manzoni, Giuseppe Vizzari, Kazumichi Ohtsuka, and Kenichiro Shimura. Towards an agent-based proxemic model for pedestrian and group dynamic. In *Proc. of the 11th WOA Workshop*, volume 621, 2010.

[140] Gustav Markkula, Richard Romano, Ruth Madigan, Charles W Fox, Oscar T Giles, and Natasha Merat. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. *Transportation Research Record*, 2018.

[141] Andrea Martin. Interactive motion prediction using game theory. Master's thesis, 2013. Technical University of Munich.

[142] Milecia Matthews, Girish Chowdhary, and Emily Kieson. Intent communication between autonomous vehicles and pedestrians. *CoRR*, abs/1708.07123, 2017.

[143] Bryan L. Mesmer and Christina L. Bloebaum. Modeling decision and game theory based pedestrian velocity vector decisions with interacting individuals. *Safety Science*, 87:116 – 130, 2016.

[144] Anton Milan, Laura Leal-Taixé, Ian D. Reid, Stefan Roth, and Konrad Schindler. MOT16: A benchmark for multi-object tracking. *CoRR*, abs/1603.00831, 2016.

[145] A. Mögelmose, M. M. Trivedi, and T. B. Moeslund. Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. In *Proc. of IEEE IV*, pages 330–335, 2015.

[146] B. T. Morris and M. M. Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1114–1127, 2008.

[147] Mott-Macdonald. Steps, simulation pedestrian dynamics, 1997. https://www.stepsmottmac.com/.

[148] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proc. of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[149] Dimitris Nathanael, Evangelia Portouli, Vassilis Papakostopoulos, Kostas Gkikas, and Angelos Amditis. Naturalistic observation of interactions between car drivers and pedestrians in high density urban settings. In *Proc. of the 20th Congress of the International Ergonomics Association*, pages 389–397, 2019.

[150] Hassan Obeid, Hoseb Abkarian, Maya Abou-Zeid, and Isam Kaysi. Analyzing driver-pedestrian interaction in a mixed-street environment using a driving simulator. *Accident Analysis & Prevention*, 108:56–65, 2017.

[151] Shigeyuki Okazaki. A study of pedestrian movement in architectural space, part 1: Pedestrian movement by the application on of magnetic models. *Trans. AIJ*, 283:111–119, 1979.

[152] A. Olivier, J. Bruneau, R. Kulpa, and J. PettrÃ©. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments. *IEEE Transactions on Visualization and Computer Graphics*, 24(7):2251–2263, 2018.

[153] Jennie Oxley, Brian Fildes, Elfriede Ihsen, Ross Day, and Judith Charlton. An investigation of road crossing behaviour of older pedestrians. *Transportation Research Board*, 1995.

[154] Eleonora Papadimitriou, George Yannis, and John Golias. A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3):242 – 255, 2009.

[155] A. V. Papadopoulos, L. Bascetta, and G. Ferretti. Generation of human walking paths. In *Proc. of IEEE/RSJ IROS*, pages 1676–1681, 2013.

[156] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. *Proc. of IEEE ICCV*, pages 261–268, 2009.

[157] Stefano Pellegrini, Andreas Ess, and Luc Van Gool. *Predicting Pedestrian Trajectories*, pages 473–491. Springer London, 2011.

[158] T. Petzoldt, K. Schleinitz, and R. Banse. Laboruntersuchung zur potenziellen Sicherheitswirkung einer vorderen Bremsleuchte in PKW. *Zeitschrift für Verkehrssicherheit*, 63(1):19–24, 2017.

[159] Zachary Pezzementi, Trenton Tabor, Peiyun Hu, Jonathan K. Chang, Deva Ramanan, Carl Wellington, Benzun P. Wisely Babu, and Herman Herman. Comparing apples and oranges: Off-road pedestrian detection on the national robotics engineering center agricultural person-detection dataset. *Journal of Field Robotics*, 35(4):545–563, 2018.

[160] Evangelia Portouli, Dimitris Nathanael, Kostas Gkikas, Angelos Amditis, and Psarakis Loizos. Field observations of interactions among drivers at unsignalized urban intersections. In *Proc. of the 20th Congress of the International Ergonomics Association*, 2019.

[161] A. Puydupin-Jamin, M. Johnson, and T. Bretl. A convex approach to inverse optimal control and its application to modeling human locomotion. In *Poc. of IEEE ICRA*, pages 531–536, 2012.

[162] R. Quintero, I. Parra Alonso, D. Fernández-Llorca, and M. Sotelo. Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2018.

[163] Raúl Quintero, Ignacio Parra, David Fernández Llorca, and MA Sotelo. Pedestrian intention and pose prediction through dynamical models and behaviour classification. In *Proc. of IEEE ITSC*, pages 83–88, 2015.

[164] Noha Radwan, Abhinav Valada, and Wolfram Burgard. Multimodal interaction-aware motion prediction for autonomous street crossing. *arXiv preprint arXiv:1808.06887*, 2018.

[165] H. A. Rakha, I. Zohdy, and R. K. Kamalanathsharma. Agent-based game theory modeling for driverless vehicles at intersections. 2013.

[166] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[167] A. Rasouli and J. K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, 21(3):900–918, 2020.

[168] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. Agreeing to cross: How drivers and pedestrians communicate. *Proc. of IEEE IV*, pages 264–269, 2017.

[169] Amir Rasouli and John K. Tsotsos. Joint attention in driver-pedestrian interaction: from theory to practice. *CoRR*, abs/1802.02522, 2018.

[170] E. Rehder, F. Wirth, M. Lauer, and C. Stiller. Pedestrian prediction by planning using deep neural networks. In *Proc. of IEEE ICRA*, pages 5903–5908, 2018.

[171] Eike Rehder and Horst Kloeden. Goal-directed pedestrian prediction. *Proc. of IEEE ICCV Workshop*, pages 139–147, 2015.

[172] Daniela Ridel, Eike Rehder, Martin Lauer, Christoph Stiller, and Denis Wolf. A literature review on the prediction of pedestrian behavior in urban scenarios. In *Proc. of IEEE ITSC*, pages 3105–3112, 2018.

[173] J. Rios-Martinez, A. Spalanzani, and C. Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

[174] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set forÂ multi-target, multi-camera tracking. In *Proc. of ECCV Workshops*, pages 17–35, 2016.

[175] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. In *Proc. of the 9th International Driving Symposium on Human Factors in Driver Assessment*, pages 186–192, 2017.

[176] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Proc. of ECCV*, pages 549–565, 2016.

[177] Nicole Ronald, Leon Sterling, and Michael Kirley. A conceptual framework for specifying and developing pedestrian models. 2020.

[178] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *Proc. of IEEE RO-MAN*, pages 795–802, 2016.

[179] Andrey Rudenko, Luigi Palmieri, and Kai Oliver Arras. Joint long-term prediction of human motion using a planning-based social force approach. In *Proc. of IEEE ICRA*, 2018.

[180] Christina Rudin-Brown and Samantha Jamson. *Behavioural adaptation and road safety: Theory, evidence and action.* CRC Press, 2013.

[181] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. Sophie: An attentive GAN for predicting paths compliant to social and physical constraints. In *Proc. of IEEE CVPR*, pages 1349–1358, 2019.

[182] Ehsan Sadraei, Richard Romano, Natasha Merat, Jorge Pedro, Yee Mun Lee, Ruth Madigan, Chinebuli Uzondu, Wei Lyu, and Andrew Tomlinson. Vehicle-pedestrian interaction: A distributed simulation study. In *Proc. of Driving Simulation Conference*, 2020.

[183] SAE International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2018.

[184] Andreas Schadschneider, Wolfram Klingsch, Hubert Klüpfel, Tobias Kretz, Christian Rogsch, and Armin Seyfried. *Evacuation Dynamics: Empirical Results, Modeling and Applications*, pages 3142–3176. Springer New York, 2009.

[185] Anna Schieben, Marc Wilbrink, Carmen Kettwich, Ruth Madigan, Tyron Louw, and Natasha Merat. Designing the interaction of automated vehicles with other traffic participants: A design framework based on human needs and expectations. *Cognition, Technology and Work*, 2018.

[186] Henri Schmidt, Jack Terwilliger, Dina AlAdawy, and Lex Fridman. Hacking nonverbal communication between pedestrians and vehicles in virtual reality. In *Proc. of Driving Assessment Conference*, 2019.

[187] Nicolas Schneider and Dariu M. Gavrila. Pedestrian path prediction with recursive Bayesian filters: A comparative study. In *Proc. of German Conference on Pattern Recognition*, pages 174–183, 2013.

[188] Bastian Schroeder, Nagui Rouphail, Katy Salamati, Elizabeth Hunter, Briana Phillips, Lily Elefteriadou, Thomas Chase, Yinan Zheng, Virginia P Sisiopiku, Shrikanth Mamidipalli, et al. Empirically-based performance assessment & simulation of pedestrian behavior at unsignalized crossings. Technical report, Southeastern Transportation Research, Innovation, Development and Education, 2014.

[189] A. T. Schulz and R. Stiefelhagen. A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction. In *Proc. of IEEE ITSC*, pages 173–178, 2015.

[190] A. T. Schulz and R. Stiefelhagen. Pedestrian intention recognition using latent-dynamic conditional random fields. In *Proc. of IEEE IV*, pages 622–627, 2015.

[191] David C. Schwebel, Despina Stavrinos, Katherine W. Byington, Tiffany Davis, Elizabeth E. O'Neal, and Desiree de Jong. Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention*, 45:266 – 271, 2012.

[192] Dongmei Shi, Wenyao Zhang, and Binghong Wang. Modeling pedestrian evacuation by means of game theory. *Journal of Statistical Mechanics: Theory and Experiment*, (4), 2017.

[193] Xiaodan Shi, Xiaowei Shao, Zhiling Guo, Guangming Wu, Haoran Zhang, and Ryosuke Shibasaki. Pedestrian trajectory prediction in extremely crowded scenarios. *Sensors*, 19(5):1223, 2019.

[194] J. J. (Jelmer) Slob. State-of-the-art driving simulators, a literature survey. Technical report, 2008.

[195] Dazhi Sun, Satish Ukkusuri, Rahim Benekohal, and S Travis Waller. Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. 51, 2003.

[196] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett. 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. In *Proc. of IEEE ICRA*, 2018.

[197] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. *Proc. of ACM/IEEE HRI*, 2011.

[198] Y. Tamura, P. D. Le, K. Hitomi, N. P. Chandrasiri, T. Bando, A. Yamashita, and H. Asama. Development of pedestrian behavior model taking account of intention. In *Proc. of IEEE/RSJ IROS*, pages 382–387, 2012.

[199] Kardi Teknomo. Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model. *arXiv preprint arXiv:1610.00029*, 2016. PhD Thesis.

[200] Leah L Thompson, Frederick P Rivara, Rajiv C Ayyagari, and Beth E Ebel. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury Prevention*, 19(4):232–237, 2013.

[201] H.J.P. Timmermans. *Pedestrian behavior: models, data collection and applications*. Emerald, 2009.

[202] Isabelle Tournier, Aurélie Dommes, and Viola Cavallo. Review of safety and mobility issues among older pedestrians. *Accident Analysis & Prevention*, 91:24–35, 2016.

[203] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *Proc. of IEEE/RSJ IROS*, pages 797–803, 2010.

[204] Alasdair Turner and Alan Penn. Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. 29:473–490, 2002.

[205] A. Turnwald, W. Olszowy, D. Wollherr, and M. Buss. Interactive navigation of humans from a game theoretic perspective. In *Proc. of IEEE/RSJ IROS*, pages 703–708, 2014.

[206] Annemarie Turnwald, Daniel Althoff, Dirk Wollherr, and Martin Buss. Understanding human avoidance behavior: Interaction-aware decision making based on game theory. *International Journal of Social Robotics*, 8(2):331–351, 2016.

[207] University of Iowa. Driving simulator, 2018. https://www.nads-sc.uiowa.edu/.

[208] Sebastiano Vascon, Eyasu Zemene Mequanint, Marco Cristani, Hayley Hung, Marcello Pelillo, and Vittorio Murino. A game-theoretic probabilistic approach for detecting conversational groups. In *Proc. of the Asian Conference on Computer Vision*, pages 658–675, 2014.

[209] Pavan Vasishta, Dominique Vaufreydaz, and Anne Spalanzani. Natural Vision Based Method for Predicting Pedestrian Behaviour in Urban Environments. In *Proc. of IEEE ITSC*, 2017.

[210] D. Vasquez. Novel planning-based algorithms for human motion prediction. In *Proc. of IEEE ICRA*, pages 3317–3322, 2016.

[211] Hendrik Vermuyten, Jeroen BeliÃ«n, Liesje De Boeck, Genserik Reniers, and Tony Wauters. A review of optimisation models for pedestrian evacuation and design problems. *Safety Science*, 87:167 – 178, 2016.

[212] Giuseppe Vizzari and Stefania Bandini. Studying pedestrian and crowd dynamics through integrated analysis and synthesis. *IEEE Intelligent Systems*, 28(5):56–60, 2013.

[213] Benjamin Völz, Holger Mielenz, Gabriel Agamennoni, and Roland Siegwart. Feature relevance estimation for learning pedestrian behavior at crosswalks. In *Proc. of IEEE ITSC*, pages 854 – 860, 2015.

[214] Benjamin Völz, Holger Mielenz, Igor Gilitschenski, Roland Siegwart, and Juan Nieto. Inferring pedestrian motions at urban crosswalks. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):544–555, 2018.

[215] Esther J. Walker, Sophie N. Lanthier, Evan F. Risko, and Alan Kingstone. The effects of personal music devices on pedestrian behaviour. *Safety Science*, 50(1):123 – 128, 2012.

[216] Peng Wang, Xinyu Huang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[217] O Michael Watson and Theodore D Graves. Quantitative research in proxemic behavior. *American Anthropologist*, 68(4):971–985, 1966.

[218] Annette Werner. New colours for autonomous driving: An evaluation of chromaticities for the external lighting equipment of autonomous vehicles. *Colour Turn*, (1), 2019.

[219] D.G. Wilson, D.G. Wilson, and G.B. Grayson. *Age-related Differences in the Road Crossing Behaviour of Adult Pedestrians*. Road User Characteristics Division, Safety Department, Transport and Road Research Laboratory, 1980.

[220] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In *Proc. of IEEE CVPR*, pages 794–801, 2009.

[221] J. Wu, J. Ruenz, and M. Althoff. Probabilistic map-based pedestrian motion prediction taking traffic participants into consideration. In *Proc. of IEEE IV*, pages 1285–1292, 2018.

[222] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. In *Proc. of IEEE CVPR*, pages 3530–3538, 2017.

[223] Dongfang Yang, Linhui Li, Keith Redmill, and Ümit Özgüner. Top-view trajectories: A pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus. In *Proc. of IEEE IV*, 2019.

[224] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Pedestrian behavior understanding and prediction with deep neural networks. In *Proc. of ECCV*, pages 263–279, 2016.

[225] Yufei Yuan, Bernat Goñi-Ros, Ha H. Bui, Winnie Daamen, Hai L. Vu, and Serge P. Hoogendoorn. Macroscopic pedestrian flow simulation using smoothed particle hydrodynamics (SPH). *Transportation Research Part C: Emerging Technologies*, 111:334 – 351, 2020.

[226] M.Suzanne Zeedyk and Laura Kelly. Behavioural observations of adult-child pairs at pedestrian crossings. *Accident Analysis & Prevention*, 35(5):771 – 776, 2003.

[227] Wei Zhan, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilian Naumann, Julius Kümmerle, Hendrik Königshof, Christoph Stiller, Arnaud de La Fortelle, and Masayoshi Tomizuka. INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. *arXiv:1910.03088 [cs, eess]*, 2019.

[228] S. Zhang, R. Benenson, and B. Schiele. Citypersons: A diverse dataset for pedestrian detection. In *Proc. of IEEE CVPR*, pages 4457–4465, 2017.

[229] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proc. of IEEE ICCV*, 2015.

[230] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In *Proc. of IEEE ICCV*, pages 3774–3782, 2017.

[231] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proc. of IEEE CVPR*, pages 2871–2878, 2012.

[232] Brian D. Ziebart, Andrew Maas, J. Andrew Bagnell, and Anind K. Dey. Maximum entropy inverse reinforcement learning. In *Proc. of AAAI*, pages 1433–1438, 2008.

[233] Brian D. Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J. Andrew Bagnell, Martial Hebert, Anind K. Dey, and Siddhartha Srinivasa. Planning-based prediction for pedestrians. In *Proc. of IEEE/RSJ IROS*, pages 3931–3936, 2009.

Chapter 4

Evaluating Pedestrian Interaction Preferences with a Game Theoretic Autonomous Vehicle in Virtual Reality

Abstract

Localisation and navigation of autonomous vehicles (AVs) in static environments are now solved problems, but how to control their interactions with other road users in mixed traffic environments, especially with pedestrians, remains an open question. Recent work has begun to apply game theory to model and control AV-pedestrian interactions as they compete for space on the road whilst trying to avoid collisions. But this game theory model has been developed only in unrealistic lab environments. To improve their realism, this study empirically examines pedestrian behaviour during road crossing in the presence of approaching autonomous vehicles in more realistic virtual reality (VR) environments. The autonomous vehicles are controlled using game theory, and this study seeks to find the best parameters for these controls to produce comfortable interactions for the pedestrians. In a first experiment, participants' trajectories reveal a more cautious crossing behaviour in VR than in previous laboratory experiments. In two further experiments, a gradient descent approach is used to investigate participants' preference for the AV driving style. The results show that the majority of participants were not expecting the AV to stop in some scenarios, and there was no change in their crossing behaviour in two environments and with different car models suggestive of car and last-mile style vehicles. These results provide some initial estimates for game theoretic parameters needed by future AVs in their pedestrian interactions and more generally show how such parameters can be inferred from virtual reality experiments.

1 Introduction

The widely predicted arrival of autonomous vehicles (AVs) on the roads poses several concerns regarding their future interaction with other road users, in particular with pedestrians. Unlike static objects in the environment which can be mapped and routed around by an AV, pedestrians are active and interactive agents, who move around to actively obtain their own

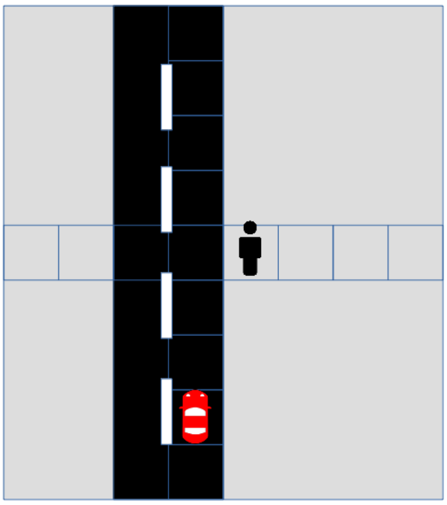


Figure 4.1: Two agents negotiating for priority at an intersection

goals and also interactively in response to the AV's own actions. Pedestrians can now be detected and tracked quite reliably [4] but modelling and controlling interactions with them remains an open question [5].

Recent trials of autonomous minibuses in European cities [28][1] has shown that pedestrians can easily take advantage over AVs: these autonomous minibuses were programmed to stop when any pedestrian stepped in front of them. After a few days observing the AV's behaviour, some pedestrians appeared to learn this safety feature and started stepping intentionally in front of the AV, with instances of this behaviour occurring around once every three hours. Human drivers would not allow this to occur and would instead usually control their vehicles in ways to suggest some threat to such pedestrians, interacting with them to encourage them to get out of their way. This inability of current AVs to similarly control this type of interaction is one of their biggest problems, known as the 'freezing robot problem' or 'the Big Problem with self-driving cars' [2]. To make progress towards creating suitable AV interaction controllers, we thus recently proposed a game theory model, called 'sequential chicken' for such interactions [19], where a pedestrian encounters an autonomous vehicle at an unsignalized intersection, as shown in Fig. 4.1. Game theory offers a framework to model decision-making between rational agents, it has been widely used, for example, in Economics [33] and for coordinating multi-robot systems [31]. We do not use conventional statistical analyses because they rely on a separation of cause and effect, or controlled and observed

[1] https://www.youtube.com/watch?v=PUr8ljfb2Cg

variables. But when studying interactions between agents, we inherently have both agents taking both roles, affecting one another, which is a better fit to game theoretic models than statistical methods.

After finding mathematical solutions to the model in terms of its free parameters, we then showed how the numerical values of its parameters can be fitted from empirical data. Unrealistic laboratory experiments were used to demonstrate this method. We first asked participants to simulate interactions in a board game in [9]. Secondly, participants were asked to play the game in person moving on squares with a set of two speeds (SLOW, FAST) [6]. Finally, participants played the game by moving continuously towards each other at their preferred pace [7]. While providing a proof of concept of the method for finding parameters, these laboratory experiments showed unrealistic results, with participants preferring to save time rather than avoiding collisions in order to win what they perceived as games against the other player rather than protect their safety as they may value more in real life.

The present study aims to extend these experiments by applying the same parameter fitting method to new more realistic interaction scenarios. The new scenarios use virtual reality (VR) to enable a subject to interact with a game theoretic autonomous vehicle in the same road crossing scenario. VR offers the opportunity to experiment on human behaviour in simulated real world environments that can be dangerous or difficult to study, such as pedestrian road crossing, in which experiments need to explore human behaviour leading up to and during actual collisions between vehicles and pedestrians [12, 21]. Virtual reality provides a much greater realism than the previous laboratory experiences, including a real sense of fear from being hit by the vehicle due to its apparent physical presence. These experiments are intended to show how more realistic game theory parameters can be recovered from VR interactions. These parameters could then be built into future AV software to help control their interactions with pedestrians, as well as providing interesting insight into pedestrian behaviour itself.

AVs are on their way not only to roads, but also to pavements in the form of autonomous last-mile robots used for urban delivery tasks [22, 11]. Last-mile delivery vehicles are usually smaller than road vehicles, share the same pavements as pedestrians, and drive at lower speeds. To better understand this new and important use-case for AV interaction control, we also investigate participants' behaviour with these last-mile type vehicles to test if humans prefer to interact with them differently from on-road cars.

2 Related Work

This section gives an overview of related studies on pedestrian crossing behaviour and pedestrian–AV interactions using virtual reality, showing that previous work does not yet provide the game theoretic parameters of interest and thus motivating the new experiments.

2.1 Pedestrian crossing behaviour in virtual reality

In recent years, pedestrian crossing behaviour has been studied using virtual reality environments. In particular, VR has been used for teaching safe crossing behaviour to child pedestrians [44, 29, 46]. For example, [46] studied child and young adults crossing behaviour in VR, and recommended the use of VR for future studies in this domain. Other studies have focused on hazardous crossing situations such as [30] where an investigation was carried on child and adult pedestrians' ability to detect dangerous situations while crossing in a virtual environment. The study showed that the awareness of hazardous situations increases with the age. [52] studied older and younger adults crossing behaviour in a virtual environment. They recorded pedestrian behavioural data, such as their head and eye movement. Their results showed a safer crossing behaviour from younger adults and that older adults tend to look at the ground rather than the other side of the street. [15] investigated pedestrian crossing behaviour and risk acceptance in a virtual environment. Their results suggest that VR creates realistic simulations and allows to test pre-crash events without injuries. [16] studied pedestrian behaviour in critical crossing scenarios using presence questionnaires for gap acceptance analysis. Their results showed no significant difference between the crossing behaviour in their different scenarios. [45] investigated distracted pedestrian behaviour in VR and at a real intersection. Their results showed that pedestrians self-reported a behavioural change but no significant difference has been observed in the real world. [51] studied pedestrian crossing decisions at roundabouts, mainly evaluating pedestrian gap acceptance between moving vehicles in a virtual environment. Their results were consistent with real-world data. [1] studied pedestrian crossing behaviour in virtual and real environments for different tasks. Their results showed no difference in most tasks except for the vehicle speed estimation and pedestrian's presence.

2.2 Pedestrian–AV interactions in virtual reality

Some VR studies have also specifically begun to study autonomous vehicle interactions with pedestrians. [50] developed five different behaviours for an autonomous vehicle. The vehicle behaviour was successfully tested in different simulated traffic scenarios such as at intersections and for lane changing, in a simulated city and highway road networks. [23] studied autonomous vehicles interactions with pedestrians in a virtual environment. In one of their experiments, participants were asked to cross a road in front of them while a vehicle is approaching. Their experiment differs from ours in that the AV stops and shows (or not) a stop intent to pedestrians. This study aimed to show the importance of substituting communications between pedestrians and drivers by some explicit communication forms for self-driving cars. [37] performed an experiment with participants on their crossing behaviour using virtual reality. They used task analysis to divide pedestrian–vehicle interaction as a sequence of actions giving two outcomes, either the vehicle passes first or the pedestrian crosses. [21] proposed a testing procedure for studying safety critical systems, e.g. autonomous vehicles interacting with pedestrians, using VR techniques. This test bed can take into account different factors that could influence pedestrian behaviour such as their understanding of the environment, their body movement and their personality. [43] investigated social cues in pedestrian–AV interactions in a VR environment. Their study showed that VR is a powerful tool for studying pedestrian–AV interactions but also that social cues could be manufactured through the vehicle trajectory. [47] validated the use of virtual reality for pedestrian–AV interactions. Moreover, their study showed that explicit HMI improves the interactions between autonomous vehicles and pedestrians. [13] investigated pedestrian preferences for external features on a fully autonomous vehicle in VR. Their results showed a significant change in pedestrian crossing due to the external displays. [14] showed that facial communication cues such as eye contact do not play a major role in pedestrian crossing behaviour, and that the motion pattern and behaviour of vehicles are more important. The field study in [42] showed similar results with an "unmanned" vehicle, suggesting that the same results could be found with autonomous vehicles. [41] also showed that vehicle movement is sufficient for indicating the intention of drivers and presented some motion patterns of road users such as advancing, slowing early and stopping short. [10] developed an AV prototype with "eyes" in a VR study. Their results showed that pedestrians were quicker at making their crossing decision and they feel safer knowing that the AV has seen them. [3] studied pedestrian reactions (trust, safety) to different AV manoeuvres in a virtual environment. Their results showed that VR is realistic for studying pedestrian behaviour. [36] studied pedestrian–vehicle interactions using recorded 360° videos displayed in VR. Their results showed that pedestrians may change their crossing behaviour based on an AV appearance. In [35], the same authors studied pedestrian crossing behaviour in VR. Pedestrian trust levels were measured and they showed a higher crossing intention. No crossing difference was found between vehicle types. The authors used a mixed-model binomial logistic regression and found that the presence of a zebra crossing and large gaps between vehicles lead to more pedestrian crossing.

2.3 Game theory for pedestrian–AV interactions

Game theory has been widely used for various applications in transportation, such as vehicle to vehicle (V2V) communications [48, 24, 49], freight transportation [17], driver-AV interactions [34, 18]. The few game theory models that focused on pedestrian–AV interactions are very recent. For example, [32] proposed two variants of a game theory model for AV interactions with cyclists and pedestrians. [39] developed a game theory model for pedestrian motion and walking behaviors. [38] then extended this model and built upon it a game theoretic framework for pedestrian–vehicle and pedestrian–pedestrian interactions. [26] proposed a level-k game theory model for autonomous vehicle controller at unsignalised intersections, based on a discrete time, set of actions and a reward function. The game theory model in this work called the sequential chicken model was proposed in [19], it is based on the famous game of chicken. The model is detailed in Sec. 3.2.

Summary of the contributions:

The above related work has shown that virtual reality is a reliable tool for studying human behaviour. Despite these numerous studies, it finds no previous study with a game theoretic vehicle interacting with human pedestrians in a VR environment. The present study fills this gap and uses VR to run the game theoretic model proposed in [19] on a virtual autonomous vehicle and then evaluates the behavioural preferences of human participants. Thus, this paper:

- shows the first attempt to quantitatively evaluate pedestrian behaviour during interaction scenarios with a game theoretic autonomous vehicle in a virtual reality environment;
- proposes a new method sufficient to infer specific numerical values for use in AV interaction control software;

- demonstrates the importance of VR for pedestrian behaviour study and for the development and testing of autonomous vehicle algorithms.

3 Methods

Our method consists in controlling an AV in VR using the game theory model, then measuring human subjects' behaviour during, and their responses after road crossing interactions with the AV under varied parameter settings of the game theory controller. In our previous work, we inferred game theory parameters to describe *human* behaviours, but here in contrast it is the parameters of the AV which are varied and studied. We seek the best parameters for the AV controller, which could for example then be built into real vehicles as part of their control.

3.1 VR Setup

The study was conducted using an HTC Vice Pro head mounted display (HMD). Participants did not use the HTC Vice controllers, as no interactions other than walking were required. The HMD was used with the HTC wireless adapter in order to facilitate easier movement during the simulation. We used an area of approximately 6 m by 3 m to conduct the simulation (as shown in Fig. 4.2), which was mapped using the usual HTC Vive room mapping system. The size of this area slightly exceeds that recommended by the manufacturer; however, we experienced no technical problem with tracking or system performance. The start position on the floor was marked with an "X" using floor tape, so that participants knew where to stand at the start of each simulation, prior to placing the HMD on their head. The simulation was created using the Unity 3D engine[1], and was run under Windows 10 on a PC based on an Intel Core i7-7700K CPU, with 32GB of RAM, and an Nvidia GeForce GTX 1080 GPU.

3.2 Game-theoretic AV behaviour model

The virtual AV was designed to drive using the sequential chicken model [19]. In this model, two agents (e.g., pedestrian and/or human or autonomous driver) called Y and X are moving towards each other at an unmarked intersection. This process occurs over a discrete space (the path is formed of squares) as in Fig. 4.3 and discrete times ('turns') during which the agents can adjust their discrete speeds. Here a turn corresponds to one discrete time step,

[1] https://unity.com/

Figure 4.2: VR Lab

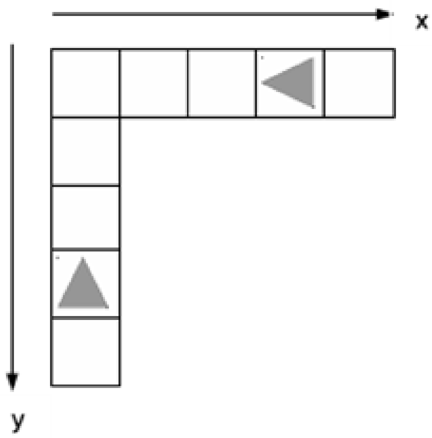


Figure 4.3: Sequential Chicken Model

i.e. the time offered to the agents to make a new decision. They simultaneously select their speed of either 1 square per turn (SLOW) or 2 squares per turn (FAST), at each turn. Space and time are discrete to keep the model simple and computationally tractable. Both agents want to pass the intersection as soon as possible to avoid travel delays, but if they collide, they are both bigger losers as they both receive a negative utility, U_{crash}. Otherwise if the players pass the intersection, each receives a time delay penalty, $-TU_{time}$, where T is the time from the start of the game and U_{time} represents the value of saving one turn of travel time.

The model assumes that the two players choose their actions (speeds) $a_Y, a_X \in \{1, 2\}$ simultaneously then implement them simultaneously, at each of several discrete-time turns. There is no lateral motion (positioning within the lanes of the roads) or communication between the agents other than via their visible positions. The game is symmetric, as both players are assumed to know that they have the same utility functions (U_{crash}, U_{time}), hence they both have the same optimal strategies. These optimal strategies are derivable from game theory together with meta-strategy convergence, via recursion. Sequential chicken can be viewed as a sequence of one-shot sub-games, whose payoffs are the expected values of new games resulting from the actions, and are solvable by standard game theory.

The (discretised) locations of the players can be represented by (y, x, t) at turn t and their actions $a_Y, a_X \in \{1, 2\}$ for speed selection. The new state at turn $t + 1$ is given by $(y + a_Y, x + a_X, t + 1)$. We define $v_{y,x,t} = (v^Y_{y,x,t}, v^X_{y,x,t})$ as the value (expected utility, assuming all players play optimally) of the game for state (y, x, t). As in standard game theory, the value of each 2×2 payoff matrix can then be written as,

$$v_{y,x,t} = v(\begin{bmatrix} v(y-1, x-1, t+1) & v(y-1, x-2, t+1) \\ v(y-2, x-1, t+1) & v(y-2, x-2, t+1) \end{bmatrix}), \tag{4.1}$$

which can be solved using dynamic programming assuming meta-strategy convergence equilibrium selection. Under some approximations based on the temporal gauge invariance described in [19], we may remove the dependencies on the time t in our implementation so that only the locations (y, x) are required in computation of $v_{y,x}$ and optimal strategy selection.

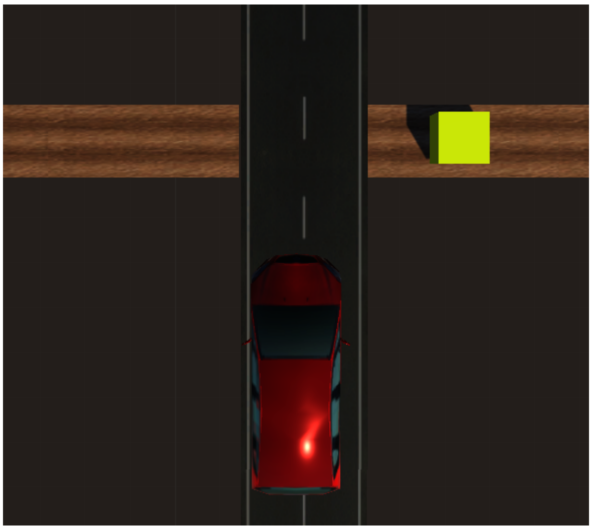
The virtual car model was imported from Unity Asset Store. The AV began driving 40 meters away from the intersection. The vehicle moved and adapted its behaviour to participants' motion. Every time step, the AV observed the current position of the pedestrian and made its decision based on the game theory model. The AV was designed not to stop completely for any pedestrian, rather it was designed only to slow to a lower but nonzero speed if necessary to yield to them. This was because a complete stop could potentially last forever, while ensuring a positive speed at all times guarantees a finite length interaction, which is required by the finite mathematics of the game theory model. In fact, in the sequential chicken model, if the two players play optimally, then there must exist a non-zero probability for a collision to occur. Intuitively, if we consider an AV to be one player that always yields, it will make no progress as the other player will always take advantage over it, hence there must be some threat of collision.

3.3 Human experiment

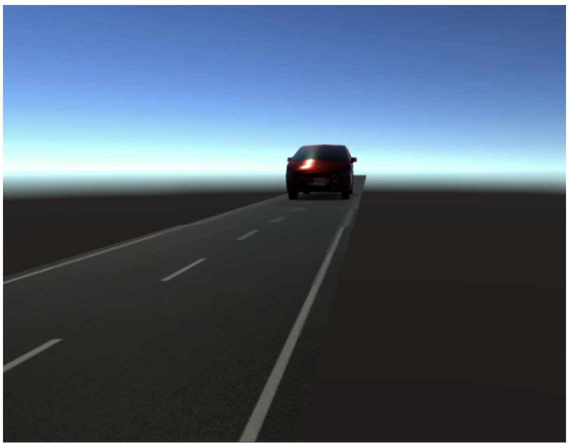
We invited members of staff and students from the University of Lincoln to take part in our study composed of three experiments, under the University of Lincoln Research Ethics. A few participants did the three experiments at different moments, some did two experiments and some others did only one experiment. Participants were not informed about the virtual vehicle behaviour, so they did not know that it was an autonomous vehicle nor that it had a game theoretic behaviour.

3.3.1 Experiment 1

We had 11 participants, 10 males and 1 female aged between 19 and 37 years old, who took part in this first experiment, seven of them had previous experience with VR. Participants were asked to cross a road in front of them as they would do in everyday life. They should stop moving on the other side of the road, when they reach a yellow cube used as a VR obstacle which people would avoid. The cube was located there for safety reasons, so the participants do not walk into a wall in real life, as shown in Fig. 4.2. A vehicle approaches from their right hand side. The AV's full speed was 30km/h, its lowest speed was 15km/h and it updated its decision every $0.02s$. Participants began walking about 4 meters away from the intersection. Prior to the experiment, participants were introduced to the experimental setup and trained on walking within the VR environment with the VR headset. There were 6 trials per participants in the virtual environment with the first trials considered as practice runs in order to get the subjects comfortable with the setup before the actual data collection.



(a) Top view of the scene used for Experiments 1 and 2



(b) Virtual Autonomous Vehicle

(c) Participant taking part in the study

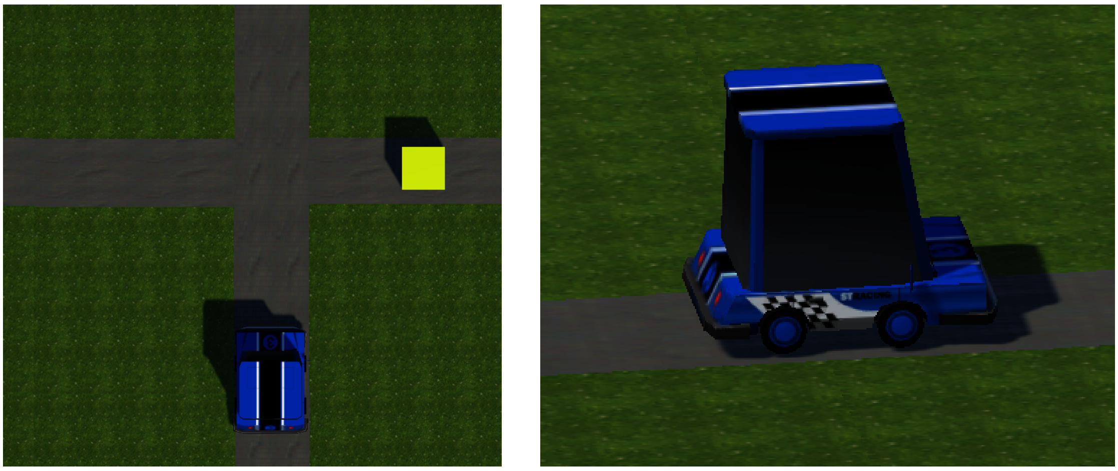
Figure 4.4: VR Experiment

3.3.2 Experiment 2

Nine participants, 7 males and 2 females, aged from 21 to 39 years old took part in the study. Seven participants had previous experience with VR. Participants were given the same instructions as in Experiment 1, the environment and the AV's speed were also the same. The particularity here is that participants were asked, after each interaction, whether they preferred their last interaction with the vehicle or the previous one, in the sense of whether they found the vehicle behaviour more "natural" and more "realistic". Note that this is different from asking for a preference based on their own utility such as whether they managed to cross quickly. At each new interaction, the parameters were adjusted by the experimenter using a manual gradient descent, to seek parameters for the autonomous vehicle that were the most preferred by the participant. Two parameters were changed, the first one being about the spatial motion i.e. the number of discrete cells used in the sequential chicken model and the second parameter was about the time delay i.e. the amount of time that would elapse between two decisions made by the AV. There were 8 proposed parameters in the spatial axis {3 cells, 5 cells, 10 cells, 15 cells, 20 cells, 25 cells, 30 cells, 40 cells} and 3 proposed in the temporal axis {0.02s, 0.5s,1.0s}. The experimenter would ideally move one step along each axis per interaction, but the experimenter's subjective intuition was also allowed to hypothesize other parameter changes to try to speed up the gradient descent. This is an acceptable use of experimenter subjectivity because the aim of gradient descent is only to find the best parameters, so any form of proposal is acceptable if it gives better results than previous ones.

3.3.3 Experiment 3

This experiment investigated pedestrian interactions with a last-mile type delivery vehicle. The protocol was exactly the same as in Experiment 2, except that here, the environment was designed to look more like a park or a garden, by replacing the wide tarmac road with a narrower pathway without markings as shown in Fig. 4.5a. This was to test whether this type of environment alters pedestrian behaviour. The type of vehicle used was also different, it was smaller, with a different colour and looked like a single person podcar, as shown in Fig. 4.5b, and for this reason, the AV's lowest speed was set to 4km/h, to show a significant deceleration. The 3D car model was imported from Unity Asset Store. Six participants, 5 males and 1 female, aged from 21 to 39 years old took part in the study, with 5 participants having had previous experience with VR.



(a) Top view of the scene

(b) Small virtual vehicle

Figure 4.5: Experiment 3

3.4 Gaussian process parameter posterior analysis

We used Gaussian process (GP) regression [40] to fit the posterior belief over the behavioural parameters of interest, $\theta = (U_{crash}, U_{time})$ from the observed data, D. Under the sequential chicken model, M, these are,

$$P(\theta|M, D) = \frac{P(D|\theta, M)P(\theta|M)}{\sum_{\theta'} P(D|\theta', M)P(\theta'|M)}. \tag{4.2}$$

We assume a flat prior over θ so that,

$$P(\theta|M, D) \propto P(D|\theta, M), \tag{4.3}$$

which is the data likelihood, given by,

$$P(D|\theta, M) = \prod_{game} \prod_{turn} P(d_X^{game,turn}|y, x, \theta, M'), \tag{4.4}$$

where $d_{player}^{game,turn}$ are the observed action choices, and y and x are the observed player locations at each *turn* of each *game*. Here M' is a noisy version of the optimal sequential chicken model M, which plays actions from M with probability $(1-s)$ and maximum entropy random actions (0.5 probability of each speed) with probability s. This modification is necessary to allow the model to fit data where human players have made deviations from optimal strategies which would otherwise occur in the data with probability zero. Real humans are unlikely to be perfectly optimal at any time as they may make mistakes of perception and decision making. This is a common method to weaken psychological models to allow non-zero probabilities for such mistakes if present [27, 9, 6].

For a given value of θ, we may compute the optimal strategy for the game by dynamic programming as shown in Algorithm 1. Optimal strategies are in general probabilistic, and prescribe the $P(d_X^{game,turn}|y,x,\theta,M)$ terms to compute the above data likelihood. We then use a Gaussian process with a Radial Basis Function (RBF) kernel to smooth the likelihood function over all values of θ beyond a sample whose values are computed explicitly. In practice, this is performed in the log domain to avoid numerical computation problems with small probabilities. The resulting Gaussian process is then read as the (un-normalized, log) posterior belief over the behavioural parameters $\theta = \{U_{time}, U_{crash}\}$ of interest.

To fit parameters of the discrete sequential chicken model to the continuous pedestrian trajectory data, it was discretised assuming an average speed of 1m/s and sampled every 3 time steps; a similar approach was used in [7].

Algorithm 1 Optimal solution computation

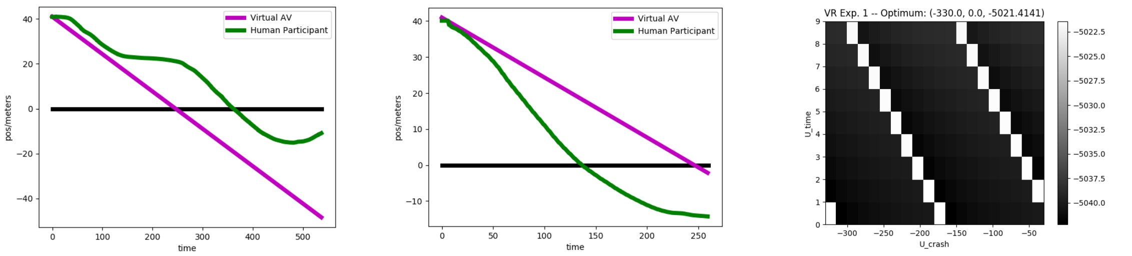
for U_{crash} in range($U_{crash_{min}}$, $U_{crash_{max}}$) **do**
  for U_{time} in range($U_{time_{min}}$, $U_{time_{max}}$) **do**
    S $\leftarrow$ strategy matrix($NY \times NX \times 2$) for P(player X chooses speed $2|y,x$)
    loglik = 0
    for each game in data **do**
      for each turn in game **do**

$$\text{loglik} = \prod_{game}\prod_{turn}(1-s)P(d_X^{game,turn}|y,x,\theta,M) + s(\frac{1}{2})$$

      end for
    end for
    Store loglik(U_{crash}, U_{time})
  end for
end for
maxloglik $\leftarrow$ max of loglik(U_{crash}, U_{time})

Table 4.1: Statistics about pedestrian crossing choices

Pedestrian Action	Experiment 1	Experiment 2	Experiment 3
Crossing	6	12	30
Stopping	49	118	58
Total	55	130	88



(a) Pedestrian–AV trajectories: stopping

(b) Pedestrian–AV trajectories: crossing

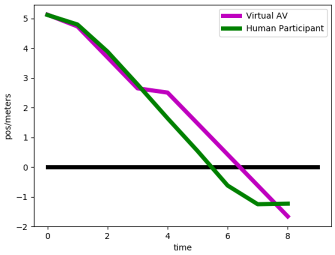
(c) Pedestrian behavioural preference

Figure 4.6: Results Experiment 1

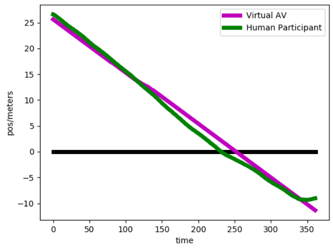
4 Results

4.1 Statistics

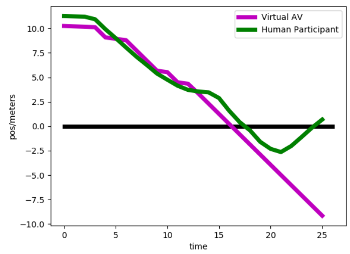
Table 4.1 shows some statistics about pedestrian crossing choices in the three experiments. We observe that very few pedestrians decided to cross in Experiments 1 and 2, about 10% crossings in each, whereas in Experiment 3, participants were more assertive and crossed in 34% of the interactions. This result is not surprising, in fact, this can be easily explained by the difference in the AV's slow speed, 15km/h (Experiments 1and 2) versus 4km/h (Experiment 3), showing that pedestrians adapt their behaviour to the vehicle's behaviour rather than on its appearance, as found in [14, 41].



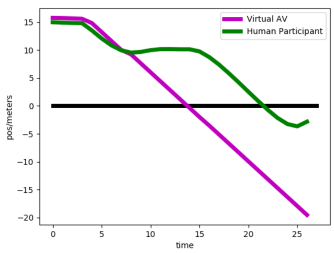
(a) AV parameters =(5 cells, 1s)



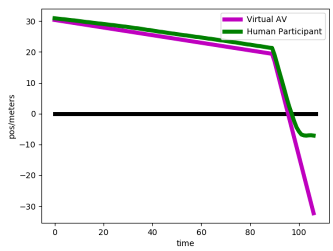
(b) AV parameters = (25 cells, 0.02s)



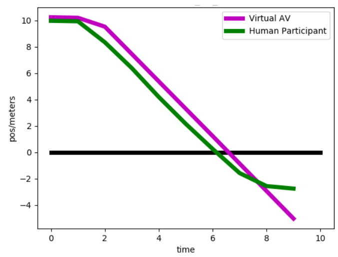
(c) AV parameters = (10 cells, 0.5s)



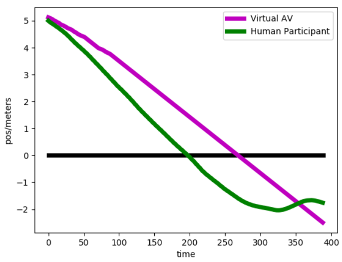
(d) AV parameters = (15 cells, 0.5s)



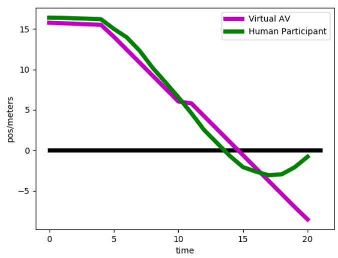
(e) AV parameters = (30 cells, 0.02s)



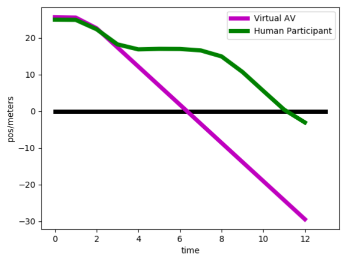
(f) AV parameters = (10 cells, 1s)



(g) AV parameters = (5 cells, 0.02s)



(h) AV parameters = (15 cells, 0.5s)



(i) AV parameters = (25 cells, 1s)

Figure 4.7: Examples of pedestrian–AV trajectories from Experiments 2 and 3

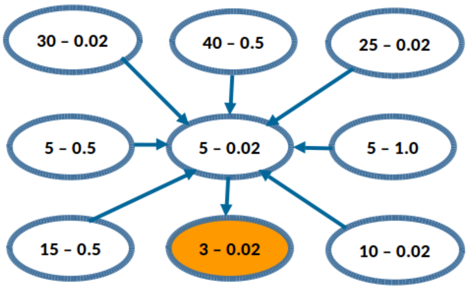
4.2 Pedestrian behaviour in Experiment 1

In total, 55 pedestrian–vehicle interactions were recorded. Among those interactions, pedestrians managed to cross the road before the AV reached the intersection only 6 times. These crossings usually happened after the first trials, by pedestrians who felt more confident after evaluating/gauging the AV driving style. Most interactions looked similar to Figs. 4.6a (pedestrian stopping) and 4.6b (pedestrian crossing), which show the trajectories of the human participant and the virtual autonomous vehicle. In particular, the trajectory profile in Fig. 4.6a shows that pedestrians were slowing down very quickly after seeing the AV, they were not playing optimally the game of chicken, so that the AV could cross most of the time.

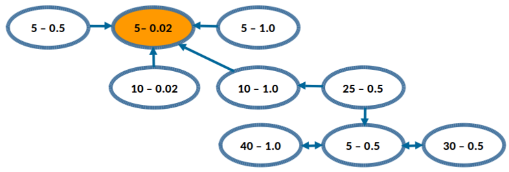
Using Algorithm 1 for Experiment 1 trajectories, we obtain a behavioural parameter $\theta = U_{crash}/U_T = -330/0$, for participants, as shown in Fig. 4.6c. This reveals that pedestrians valued the avoidance of a crash 330 times more than a 0.02s time saving per turn, resulting in pedestrians being less assertive in crossing the road. In comparison, previous laboratory experiments found that participants valued time saving more than collision avoidance [9, 6]. Thus, the use of virtual reality has made the interactions much more realistic.

4.3 Pedestrians' evaluation of the virtual AV behaviour using gradient descent (Experiments 2 and 3)

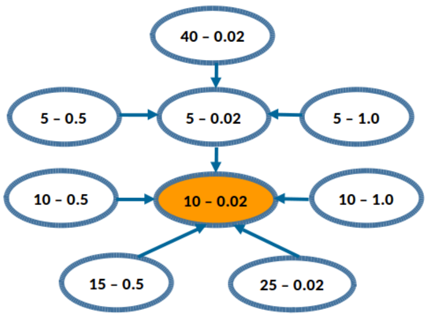
Fig. 4.7 shows examples of pedestrian–AV trajectories from Experiments 2 and 3. Examples are presented in Fig. 4.8 from four participants' interactions with the virtual AV for finding their most preferred AV parameters using the gradient descent approach. The results for pedestrians' most preferred parameters are summarized in Fi.g. 4.9a for Experiment 2 and in Fig. 4.9b for Experiment 3. The mean parameter values for the experiments $\{$mean_exp2 = (16 cells, 0.34s), mean_exp3 = (19 cells, 0.35s)$\}$, are found to be quite similar. That suggests that pedestrians had similar preferences for the AV behaviour in both environments but also that they behaved in the same way.



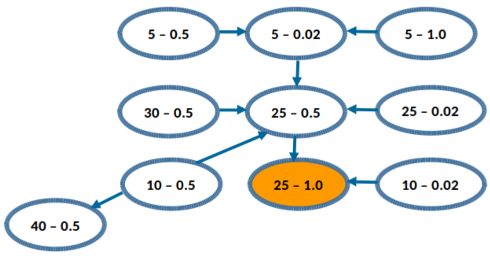
(a) Preferred AV parameters = (3 cells, 0.02s)



(b) Preferred AV parameters = (5 cells, 0.02s)

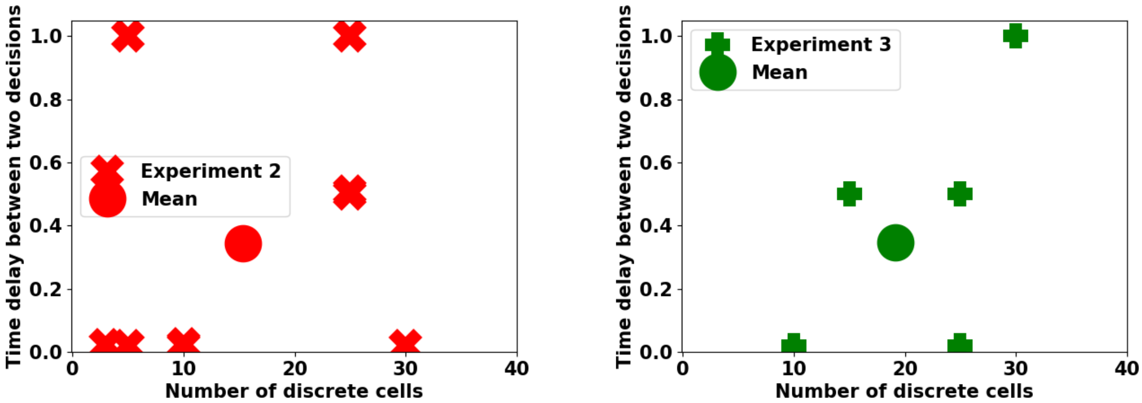


(c) Preferred AV parameters = (10 cells, 0.02s)



(d) Preferred AV parameters = (25 cells, 1s)

Figure 4.8: Examples of gradient descent approach for pedestrian most preferred AV parameters selection



(a) Experiment 2

(b) Experiment 3

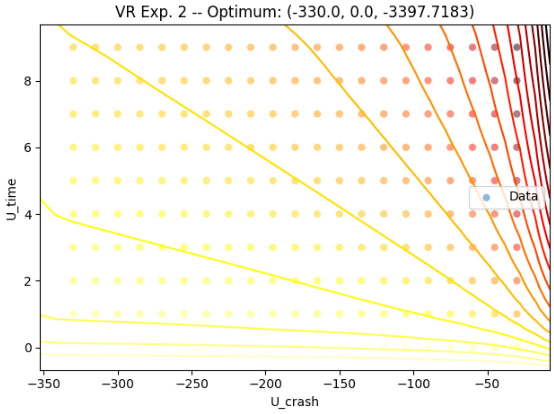
Figure 4.9: Results of pedestrian most preferred AV parameters using gradient descent

4.4 Evaluation of pedestrian crossing behaviour using Gaussian process regression (Experiments 2 and 3)

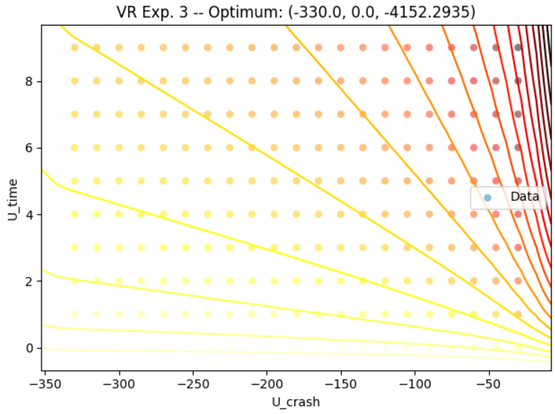
After applying Gaussian process regression and optimising s to maximise the likelihood at the Maximum A Posteriori (MAP) point of θ, the posterior distribution over $\theta = \{U_{crash}, U_{time}\}$ are shown in Figs. 4.10a and 4.10b. The MAP estimate of the parameters is found to be the same for Experiments 2 and 3, it is around $U_{crash} = -330$, $U_{time} = 0$, at $s_2 = 0.32$, and $s_3 = 0.2057$, respectively. Unsurprisingly, the same parameter estimate was found in Experiment 1. The 330 : 0 ratio in the utilities means that assuming the noisy model M' the subjects valued the avoidance of a crash 330 times than saving any time. And the s_i value , $i \in \{2, 3\}$, means that the subjects make mistakes from optimal behaviour in 32% and 20.57% of actions in Experiments 2 and 3, respectively. Significance of the results can be seen by inspection of the thin standard deviation widths of 1D slices through the 2D posterior as shown in Figs. 4.10c and 4.10d. The finding of the same MAP estimate for both experiments shows that participants behaved similarly within the two environments and with the different car models.

5 Discussion

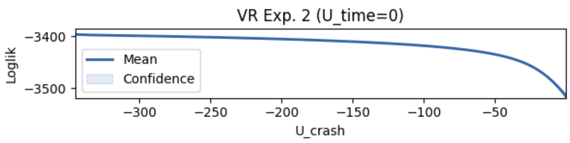
The results provide new estimates for specific numerical parameters which AV controller software could use in the sequential chicken model to control interactions with pedestrians.



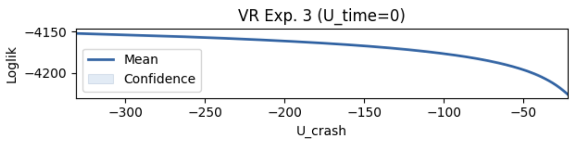
(a) Gaussian process regression for Experiment 2



(b) Gaussian process regression for Experiment 3



(c) GP slice for Experiment 2



(d) GP slice for Experiment 3

Figure 4.10: Pedestrian behavioural preferences for Experiments 2 and 3: Figs. 4.10a and 4.10b show the Gaussian process log-posterior over behavioural parameters. Figs. 4.10c and 4.10d show the slices through the Gaussian process showing standard deviation log-posterior confidence.

The result in study 1 is important as it shows that virtual reality makes pedestrian crossing behaviour more realistic than in the previous laboratory experiments [9, 6]. Pedestrians had a higher preference for avoiding collisions in VR, which gives confidence that the VR environment is more realistic than the previous laboratory experiments and therefore that the numerical parameters found next are good.

The other two experiments then showed that when interacting with an autonomous vehicle, pedestrians care more about the vehicle behaviour than its appearance, i.e. whether it should slow down, keep driving or completely stop for them, as found with the gradient descent method and in [14, 42, 41]. An interesting point to raise here is that the gradient descent method provided numerical results that could be inserted into future experiments or even practical vehicles.

The results from the Gaussian process regression also showed that participants behave similarly in different environments and with different car models, similar to the results in [35]. In particular, in Experiment 3, it appeared that the smaller car and the park environment did not make much difference in pedestrian crossing behaviour. These VR studies also confirm that pedestrian behaviour can be represented by one parameter θ and that there is a linear mapping between U_{crash} and U_{time}, a similar result was found in [9, 6].

There are some limitations with these experiments. The gradient descent takes a long time to run and it was hard for the experimenter to hypothesise which direction to follow, because after several interactions, participants sometimes rejected a set of preferred parameters that they approved several times before. It was also confusing and confounding to infer parameters for both pedestrians' own behaviour and their preferred AV behaviour. Hence, other methods of learning the best behavioural parameters for the autonomous vehicle will be explored in future studies. At first, we plan to simplify the protocol by replacing the virtual game theoretic AV by a human participant driving a virtual vehicle, so that to learn the behavioural parameters of participant drivers and used them for the game theoretic AV. Future work will also further investigate pedestrian crossing behaviour with different car models and within different environments with a larger number of participants.

The GP results from a previous laboratory experiment showed that $U_{time} = 45$ was bigger than $U_{crash} = -30$ [6], which is unrealistic, here instead we obtain an absolute measure of $U_{crash} = -330$ that is much bigger than $U_{time} = 0$, meaning that participants valued collision avoidance much more than saving time. However, the statistics may not be powerful enough to distinguish between values of $U_{crash} \to -\inf$ and $U_{crash} =$ -330, this is shown in Figs. 4.10c and 4.10d with the curve becoming more and more horizontal for $U_{crash} < -300$. We believe that collecting larger data would help make the distinction and measure the parameters more accurately.

The results of these experiments should be consistent if moved from the VR lab to the real world (UK), because subjects were asked to behave as in real life and they are not incentivised to do otherwise. However, moving from one country to another, the numbers may vary because of different cultural norms [25]. The statistically implied values of travel times and human lives, and risk appetites, are well-known to vary between cultures and the sequential chicken analysis might help to better model and understand these relationships from data in the future.

Previous work [20] has shown that pedestrian–driver interactions at semi-crosswalks are different when the road changes from one way to a two-way street. For instance, it was observed that drivers tend to decelerate or stop more on the two-way setting. Thus, future work should investigate the effects of road settings on pedestrian–AV interactions.

Potential policy implications of this work include better understanding and regulating autonomous vehicle interactions with pedestrians. The sequential chicken model shows that unless AVs are able to inflict some kind of negative utility onto pedestrians, then pedestrians can always push in front of them to win interactions and impede the AV's progress. A collision is an obvious but extreme form of negative utility which policy obviously wishes to avoid. By understanding and quantifying how the tradeoff between time saved and risk of collision works, as in the present study, the game theory mathematics could then be used to replace the rare but extreme risk of collision with some other, more common but less extreme negative utility. One possible solution, proposed in [8], is to use humans' sense of psychological discomfort when their personal space is invaded by other agents (proxemics) as such a negative utility. Without this replacement, policy would either have to tolerate occasional actual, deliberate collisions by AVs, or risk them making no progress. With the replacement, AVs can operate safely and efficiently but within existing regulations which prevent collisions.

Author Contributions

Conceptualization, F.C. and C.F.; methodology, F.C. and C.F.; software, F.C. and P.D.; validation, F.C.; formal analysis, F.C.; investigation, F.C.; resources, P.D. and C.F; data curation, F.C; writing–original draft preparation, F.C.; writing–review and editing, F.C, P.D and C.F.; visualization, F.C.; supervision, C.F.

Funding

This project has received funding from EU H2020 interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Rajaram Bhagavathula, Brian Williams, Justin Owens, and Ronald Gibbons. The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 2056–2060. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[2] Rodney Brooks. The big problem with self-driving cars is people and we'll go out of our way to make the problem worse, 2017. IEEE SPECTRUM.

[3] Christopher G Burns, Luis Oliveira, Vivien Hung, Peter Thomas, and Stewart Birrell. Pedestrian attitudes to shared-space interactions with autonomous vehicles–a virtual reality study. In *International Conference on Applied Human Factors and Ergonomics*, pages 307–316. Springer, 2019.

[4] Fanta Camara, Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, and Charles W. Fox. Pedestrian models for autonomous driving Part I: low-level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems*, 2020. https://doi.org/10.1109/TITS.2020.3006768.

[5] Fanta Camara, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Anna Schieben, Gustav

Markkula, Fabio Tango, Natasha Merat, and Charles W. Fox. Pedestrian models for autonomous driving Part II: high-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 2020. https://doi.org/10.1109/TITS.2020.3006767.

[6] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. Towards pedestrian-av interaction: method for elucidating pedestrian preferences. In *IEEE/RSJ Intelligent Robots and Systems (IROS) Workshops*, 2018.

[7] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. *Continuous Game Theory Pedestrian Modelling Method for Autonomous Vehicles.* In Human Factors in Intelligent Vehicles. River Publishers, 2020.

[8] Fanta Camara and Charles Fox. Space invaders: Pedestrian proxemic utility functions and trust zones for autonomous vehicle interactions. *Int J of Soc Robotics*, 2020. https://doi.org/10.1007/s12369-020-00717-x.

[9] Fanta Camara, Richard Romano, Gustav Markkula, Ruth Madigan, Natasha Merat, and Charles W. Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Measuring Behavior 2018: 11th International Conference on Methods and Techniques in Behavioral Research.* Manchester Metropolitan University, March 2018.

[10] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. Eyes on a car: an interface design for communication between an autonomous car and a pedestrian. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 65–73, 2017.

[11] Stef de Groot. Pedestrian acceptance of delivery robots: Appearance, interaction and intelligence design. Master's thesis, 2019.

[12] Shuchisnigdha Deb, Daniel W Carruth, Richard Sween, Lesley Strawderman, and Teena M Garrison. Efficacy of virtual reality in pedestrian safety research. *Applied ergonomics*, 65:449–460, 2017.

[13] Shuchisnigdha Deb, Lesley J Strawderman, and Daniel W Carruth. Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation research part F: traffic psychology and behaviour*, 59:135–149, 2018.

[14] Debargha Dey and Jacques Terken. Pedestrian interaction with vehicles: Roles of explicit and implicit communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '17, pages 109–113, New York, NY, USA, 2017. ACM.

[15] Igor Doric, Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Sebastian Wittmann, Matheus Zimmermann, and Thomas Brandmeier. A novel approach for researching crossing behavior and risk acceptance: The pedestrian simulator. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 39–44, 2016.

[16] Ilja Feldstein, André Dietrich, Sasha Milinkovic, and Klaus Bengler. A pedestrian simulator for urban crossing scenarios. *IFAC-PapersOnLine*, 49(19):239–244, 2016.

[17] Miguel A Figliozzi, Hani S Mahmassani, and Patrick Jaillet. Repeated auction games and learning dynamics in electronic logistics marketplaces: Complexity, bounded rationality, and regulation through information. In *Managing Complexity: Insights, Concepts, Applications*, pages 137–175. Springer, 2008.

[18] M. Flad, L. Fröhlich, and S. Hohmann. Cooperative shared control driver assistance systems based on motion primitives and differential games. *IEEE Transactions on Human-Machine Systems*, 47(5):711–722, Oct 2017.

[19] Charles W. Fox, Fanta Camara, Gustav Markkula, Richard Romano, Ruth Madigan, and Natasha Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *Proceedings of VEHITS 2018: 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, January 2018.

[20] Jon D. Fricker and Yunchang Zhang. Modeling pedestrian and motorist interaction at semi-controlled crosswalks: The effects of a change from one-way to two-way street operation. *Transportation Research Record*, 2673(11):433–446, 2019.

[21] M. Hartmann, M. Viehweger, W. Desmet, M. Stolz, and D. Watzenig. Pedestrian in the loop: An approach using virtual reality. In *International Conference on Information, Communication and Automation Technologies (ICAT)*, 2017.

[22] Thomas Hoffmann and Gunnar Prause. On the regulatory framework for last-mile delivery robots. *Machines*, 6(3):33, 2018.

[23] F. Keferböck and A. Riener. Strategies for negotiation between autonomous vehicles and pedestrians. In De Gruyter Oldenbourg, editor, *Mensch und Computer 2015 – Workshopband*, 2015.

[24] Reza Kim, Changwon & Langari. Game theory based autonomous vehicles operation. *International Journal of Vehicle Design (IJVD), Vol. 65*, 2014.

[25] Yee Mun Lee, Ruth Madigan, Oscar Giles, Laura Garach-Morcillo, Gustav Markkula, Charles Fox, Fanta Camara, Markus Rothmueller, Signe Alexandra Vendelbo-Larsen, Pernille Holm Rasmussen, et al. Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology & Work*, 2020.

[26] Nan Li, Ilya Kolmanovsky, Anouck Girard, and Yildiray Yildiz. Game theoretic modeling of vehicle interactions at unsignalized intersections and application to autonomous vehicle control. In *2018 Annual American Control Conference (ACC)*, pages 3215–3220, 2018.

[27] Yuhao Lu, Tom Stafford, and Charles Fox. Maximum saliency bias in binocular fusion. *Connection Science*, 28(3):258–269, 2016.

[28] Ruth Madigan, Sina Nordhoff, Charles Fox, Roja Ezzati Amini, Tyron Louw, Marc Wilbrink, Anna Schieben, and Natasha Merat. Understanding interactions between automated road transport systems and other road users: A video analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66:196 – 213, 2019.

[29] Joan McComas, Morag MacKay, and Jayne Pivik. Effectiveness of virtual reality for teaching pedestrian safety. *CyberPsychology & Behavior*, 5(3):185–190, 2002.

[30] Anat Meir, Tal Oron-Gilad, and Yisrael Parmet. Are child-pedestrians able to identify hazardous traffic situations? measuring their abilities in a virtual reality environment. *Safety science*, 80:33–40, 2015.

[31] Yan Meng. Multi-robot searching using game-theory based approach. *International Journal of Advanced Robotic Systems*, 5(4):44, 2008.

[32] U. Michieli and L. Badia. Game theoretic analysis of road user safety scenarios involving autonomous vehicles. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1377–1381, Sep. 2018.

[33] Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[34] Xiaoxiang Na and David J Cole. Game-theoretic modeling of the steering interaction between a human driver and a vehicle collision avoidance controller. *IEEE Transactions on Human-Machine Systems*, 45(1):25–38, 2014.

[35] J Pablo Nuñez Velasco, Haneen Farah, Bart van Arem, and Marjan P Hagenzieker. Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. *Transportation research part F: traffic psychology and behaviour*, 66:1–14, 2019.

[36] Pablo Nuñez Velasco, Haneen Farah, Bart Van Arem, and Marjan Hagenzieker. Wepod welly in delft: Pedestrians' crossing behaviour when interacting with automated vehicles using virtual reality. In *Proceedings of the 15th International Conference on Travel Behaviour Research, Santa Barbara, CA, USA*, pages 15–20, 2018.

[37] A. Pillai. Virtual reality based study to analyse pedestrian attitude towards autonomous vehicles. Master's thesis, 10 2017.

[38] Yalda Rahmati and Alireza Talebpour. Learning-based game theoretical framework for modeling pedestrian motion. *Phys. Rev. E*, 98:032312, Sep 2018.

[39] Yalda Rahmati, Alireza Talebpour, Archak Mittal, and James Fishelson. Game theory-based framework for modeling human–vehicle interactions on the road. *Transportation Research Record*, 2674(9):701–713, 2020.

[40] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[41] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. pages 186–192, 11 2017.

[42] Dirk Rothenbücher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 795–802. IEEE, 2016.

[43] Henri Schmidt, Jack Terwilliger, Dina AlAdawy, and Lex Fridman. Hacking non-verbal communication between pedestrians and vehicles in virtual reality. *arXiv preprint arXiv:1904.01931*, 2019.

[44] David C Schwebel and Leslie A McClure. Using virtual reality to train children in safe street-crossing skills. *Injury prevention*, 16(1):e1–e1, 2010.

[45] David C Schwebel, Leslie A McClure, and Bryan E Porter. Experiential exposure to texting and walking in virtual reality: A randomized trial to reduce distracted pedestrian behavior. *Accident Analysis & Prevention*, 102:116–122, 2017.

[46] Gordon Simpson, Lucy Johnston, and Michael Richardson. An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, 35(5):787–796, 2003.

[47] Sebastian Stadler, Henriette Cornet, Tatiana Novaes Theoto, and Fritz Frenkler. A tool, not a toy: using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. In *Augmented Reality and Virtual Reality*, pages 203–216. Springer, 2019.

[48] Alireza Talebpour, Hani S Mahmassani, and Samer H Hamdar. Modeling lane-changing behavior in a connected environment: A game theory approach. *Transportation Research Part C: Emerging Technologies*, 59:216–232, 2015.

[49] Z. Tian, X. Gao, S. Su, J. Qiu, X. Du, and M. Guizani. Evaluating reputation management schemes of internet of vehicles based on evolutionary game theory. *IEEE Transactions on Vehicular Technology*, 68(6):5971–5980, June 2019.

[50] H. Wang, J. K. Kearney, J. Cremer, and P. Willemsen. Steering behaviors for autonomous vehicles in virtual environments. In *IEEE Virtual Reality*, 2005.

[51] Haojie Wu, Daniel H Ashmead, and Bobby Bodenheimer. Using immersive virtual reality to evaluate pedestrian street crossing decisions at a roundabout. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pages 35–40, 2009.

[52] Giuseppe Angelo Zito, Dario Cazzoli, Loreen Scheffler, Michael Jäger, René Martin Müri, Urs Peter Mosimann, Thomas Nyffeler, Fred W Mast, and Tobias Nef. Street crossing behavior in younger and older pedestrians: an eye-and head-tracking study. *BMC geriatrics*, 15(1):176, 2015.

Chapter 5

Space Invaders: Pedestrian Proxemic Utility Functions and Trust Zones for Autonomous Vehicle Interactions

Abstract

Understanding pedestrian proxemic utility and trust will help autonomous vehicles to plan and control interactions with pedestrians more safely and efficiently. When pedestrians cross the road in front of human-driven vehicles, the two agents use knowledge of each other's preferences to negotiate and to determine who will yield to the other. Autonomous vehicles will require similar understandings, but previous work has shown a need for them to be provided in the form of *continuous* proxemic utility functions, which are not available from previous proxemics studies based on Hall's *discrete* zones. To fill this gap, a new Bayesian method to infer continuous pedestrian proxemic utility functions is proposed, and related to a new definition of 'physical trust requirement' (PTR) for road-crossing scenarios. The method is validated on simulation data then its parameters are inferred empirically from two public datasets. Results show that pedestrian proxemic utility is best described by a hyperbolic function, and that trust by the pedestrian is required in a discrete 'trust zone' which emerges naturally from simple physics. The PTR concept is then shown to be capable of generating and explaining the empirically observed zone sizes of Hall's discrete theory of proxemics.

1 Introduction

Autonomous vehicles (AVs) are claimed by many organisations to be close to commercial reality, but their lack of human behaviour understanding is raising concerns. While robotic localisation and navigation in static environments [76] and pedestrian detection [9] are well understood, AVs do not yet have the social abilities of human drivers – who can read the intentions of other road users, predict their future behaviour and then interact with them [10]. Pedestrians, unlike other road users such as cyclists, do not usually follow specific traffic rules, in particular when crossing the road at unsigned crossing points, making them especially difficult to model, predict, and interact with. Pedestrians and human drivers communicate

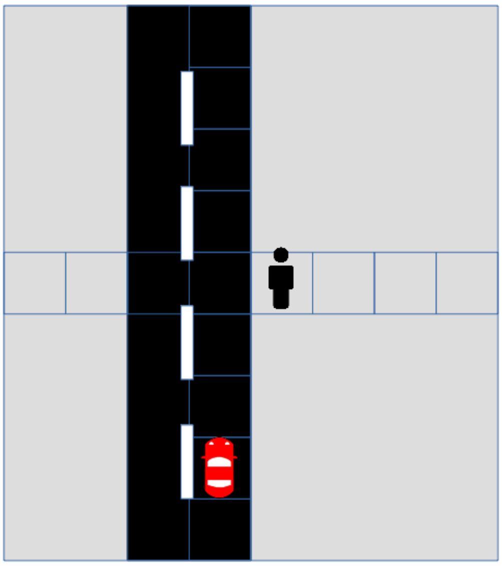


Figure 5.1: Road-crossing scenario

and interact with one another via nonverbal signals including their positions and speeds, which are used to transmit intent information as well as to make progress on the road [66]. For example, a vehicle which drives deliberately close to a pedestrian to scare them is telling them to yield, while a vehicle which maintains a larger distance from them is inviting them to cross.

Recent trials of autonomous minibuses in La Rochelle (France) and Trikala (Greece) [52], highlighted the major drawback of perfectly safe self-driving cars: it was found that pedestrians were intentionally stepping in front of the AV several times in a day, delaying their progress in the knowledge that they would always yield to the pedestrian. This abuse of perfect safety systems is known as the 'big problem with self-driving cars' [8], and in the limiting case of optimal pedestrian behaviour and large crowd size becomes the 'freezing robot problem' of vehicles making no progress at all, as they are constantly forced to yield in every interaction [78].

To make progress towards such understanding, we recently proposed and solved a game-theoretical mathematical model of the road-crossing scenario represented in Fig. 5.1, based on the famous game of 'chicken' and called 'sequential chicken' [27]. In this model, the pedestrian and vehicle compete for space in the road as they move towards one another and threaten to collide with one another, by making a temporal series of game theoretic decisions to advance or yield. The model's utility parameters for collisions and value of time were fit to human behaviours in a series of laboratory experiments [16] [11] [12] [13]. We also analysed real-world pedestrian–vehicle interactions through sequence analysis [15] to learn the most important features and how their ordering could be predictive of the outcome of an interaction [14]. The simplest mathematical solution of this game theoretic model was found to require the AV to *deliberately hit the pedestrian with a small probability*, in order to create a *credible threat* which discourages other pedestrians from taking advantage of it in the rest of the interactions [27]. This is not an ethical or legal arrangement for programming AVs in practice [79]. But the model then also suggested the possibility of an alternative solution: if the rare, large penalty of collisions could be replaced with more frequent but smaller negative utilities inflicted on pedestrians, then the same average penalty could be created and progress made by AVs without having to hit any pedestrians.

This motivates a new search for ways in which an AV could inflict *small* negative utilities onto pedestrians. Humans have evolved a sense of comfort and discomfort around one another as part of their social interaction mechanisms, which could provide a convenient and legal source of small negative utilities. For example, two pedestrians who actually collide with one another while trying to reach their destinations will obviously experience a real, physical negative utility, but it is found empirically that they also experience discomfort – a purely internally generated, psychological negative utility – when they are close but not actually touching. The study of this relationship was named *proxemics* by Hall [30]. Hall classified four discrete distance zones between people – intimate, personal, social and public – corresponding to distances where most people feel distinct levels of comfort or discomfort during interactions. If humans have evolved to feel real psychological negative utilities in the presence of only a *possibility* of collision, without it actually having to take place, then simply invading their personal space could be sufficient to penalise them enough to satisfy the game theory requirements.

It is not necessary for the reader of the present study to understand the game theory model, which provides only the motivation for the present study rather than any methods. The key motivation, taken only from its conclusions, is that it requires a utility function to directly assign numerical utilities to agents as a function of their positions. Positions are in general continuous values so a continuous proxemic utility function is required. Section 2 reviews the proxemic literature and finds that this is not yet available, which motivates the present study to develop new methods to infer it in the required form.

The method in section 3 then forms a first step towards inferring pedestrian behaviour preferences for autonomous vehicle interaction control. It consists in directly inferring the continuous proxemic utility function of pedestrians from offline data from human driver–pedestrian interactions. This is the function that could then be programmed into autonomous vehicles using the sequential chicken game theory model to provide small negative utilities.

To link continuous proxemic utility functions to the more conventional views and models of proxemics from this literature, which are mostly based on Hall's discrete zones, section 4 then introduces a new concept: 'physical trust requirement'. We show that this concept partitions the set of possible states of the world during interactions into three subspaces, for each agent. In the first, a negative utility such as a collision will happen and there is nothing either agent can do to prevent it. In the second, the negative utility may happen but only the *other* agent can choose to act to prevent it – this is the 'trust zone'. In the third, the negative utility may happen but the pedestrian themself can act to prevent it, without needing to trust the other agent. This definition of physical trust requirement may be general to many human–robot interactions in physical or abstract state spaces, but in the case of autonomous vehicle interactions with pedestrians, we provide results showing that it maps cleanly and numerically to Hall's physical proxemic zones, offering an explanation for why they emerge as discrete zones even when the proxemic utility function itself is continuous.

Section 5 finally applies both the proxemic utility function inference and physical trust requirement concept to existing public datasets, to report a real world continuous proxemic utility and physical trust requirements for the first time.

2 Related Work

This section gives a survey of related work, to search for any existing reports of numerical proxemic functions, or for any related results which might be used to infer such functions without the need for a new experiment. In particular, Hall's influential work has encouraged most studies to measure and report results in terms of discrete zones, discarding the continuous distance information which we now require. This motivates the present study to infer continuous proxemic utility functions for the first time.

2.1 Proxemics in Social Sciences

Measuring interpersonal distances during social interactions is a well-studied topic in the social sciences since the introduction of the concept by Hall [30]. For example, it was found for human–human interactions that the intimate space is up to 0.45cm, the personal space is up to 1.2m, the social space is up to 3.6m, and the public space is beyond this [45]. Thompson et al. [75] measured individuals' interaction preferences via the rating of videotapes. This study showed that people have a distance where they feel comfortable during their interactions and when the distance is smaller or greater than that, they feel more discomfort. Hayduk [31] showed via a study with university students that personal space is a two-dimensional noncircular and flexible space that can vary in shape and size. Hecht et al. [32] performed two laboratory experiments (including one in a virtual environment) with subjects and found that personal space has a circular shape with about a 1-meter radius. However, we believe that personal space can be modelled using only one dimension in the present road-crossing scenario. Stamps [72] [73] whose work is based on the theory of permeability, i.e. how people perceive (e.g. their safety) and make preferences within an environment, studied the effects of distance on participants' perception of threat. These results showed that the perceived threat decreases with larger distances.

2.2 Proxemics in Human–Robot Interactions

Proxemics is also an active research area in human–robot interaction (HRI), as shown in the review proposed by Rios-Martinez et al. [65] which focuses on social cues, signals and proxemics for robot navigation. A recent review on nonverbal communication for human–robot interaction was proposed in [69].

Walters et al. [81] proposed a framework that shows how to measure proxemic features in HRI. Their study involved participants interacting with different robots and their preferences were measured. It is explained that factors that may change human proxemics even by 20mm to 150mm can be significant. In [3], a mobile robot was developed with an autonomous proxemic system that could approach and avoid people using the distances from [81]. Koay et al. [42] measured participants' proxemics preferences using comfort level device during an HRI task.

Mead et al. [55] proposed an automatic method for annotating spatial features from 3D data of indoor human–robot interactions. In [56], the same data was used to train a Hidden Markov Model (HMM) to classify the interactions either as initiating or terminating based on the extracted physical Mehrabian's metric [59] or psychophysical Hall's metric [30]. In [57], the same authors studied the interaction between a robot and more participants, one by one. The interactions consisted in moving the robot towards the participants and backwards several times. The results showed that individuals' pre-interaction proxemic preference (mean = 1.14m, std = 0.49m) was consistent with previous studies. With a uniform performance in the robot behaviour, the proxemic preference reached a mean = 1.39m and a std = 0.63m, the participants adapted their proxemic preferences to improve the robot performance. Mead et al. [58] also investigated the influence of proxemics on human speech and gestures and measured how that impacts on the robot speech and gesture production. Their study consisted in recruiting 20 participants interacting by pairs (10 in total) who didn't know each other and each participant had to interact with the robot (PR2). Their results for human–human interactions (HHI), with a mean = 1.44m and a std = 0.34m, was consistent with previous studies but the HRI result (mean = 0.94m, std = 0.61m) was much larger than in previous studies, which could be explained by the presence of robot gestures.

Heenan et al. [33] used proxemics and Kendon's greeting observations [40] for a Nao robot interacting with human encounters. They applied Takayama and Pantofarou's [74] empirical results for proxemics, which are 0.4m to 0.6m (average interpersonal distances) with a 1.35m robot's height. They observed a larger distance between women participants and the robot, while men kept the same distance in HHI and HRI. In these experiments, the researchers found an improvement of the robot's social skills thanks to the proxemic behaviour and its greeting manner. Warta et al. [83] measured levels of social presence in HRI in a hallway. Participants were given a questionnaire to complete after interacting with a robot for a navigation task. In [39], Joosse et al. used a coding system to detect a set of attitudinal (likeability, human-likeness, trust) and behavioural attributes including non-verbal behaviour (eye-gaze, proxemics, emotion etc.) from participants interacting with a robot. The study showed some strong human reactions to a robot invading their personal space.

Kostavelis et al. [44] proposed a dynamic Bayesian network on top of an interaction unit to model human behaviour for a robot. Their method takes proxemic distances into account, allowing the robot to approach people at different distances depending on their current activity. Torta et al. [77] performed two psychometric experiments with subjects interacting with a small humanoid robot and proposed a parametric model of the personal space based on the results of these experiments. The model takes into account the distance and the direction of approach, and was evaluated with a user study where subjects are sitting and approached by the robot.

Henkel et al. [35] evaluated two predefined proxemic scaling functions (linear and logarithmic) for human–robot interactions. Their approach is different from ours in that the robot computes a gain value based on the proxemic distance with the human and then moves accordingly. Their experiments with participants in a search and rescue scenario and followed by a questionnaire showed a preference for a logarithmic proxemic scaling function. Patompak et al. [63] developed an inference method to learn human proxemic preferences. Their method is based on the social force model and reinforcement learning. They argued that proxemic spaces can be limited to two zones, the first being the quality interaction area where a robot could go without creating discomfort, and the private area which is the personal space. In addition, we believe that one more area is needed to model the trust relationship between humans and robots.

2.3 Proxemics in Pedestrian–AV Interactions

A comprehensive review on pedestrian models for autonomous driving is proposed in [9] [10], ranging from low-level sensing, detection and tracking models [9] to high-level interaction and game theoretic models [10]. In the context of autonomous vehicles, more work has been focused on pedestrian crossing behaviour [53], trajectory prediction [84] and for eHMI (external Human–Machine Interface) [20][54][50][29]. Very few studies have investigated interpersonal distances for pedestrian-vehicle interactions.

Risto et al. [66] studied the use of drivers' movement to signal intent and how these signals were understood by other road users. They video recorded pedestrian–vehicle interactions at different intersections and observed that pedestrian discomfort can be created by the vehicle approaching very close to the crosswalk boundary, which leads the pedestrian to slightly change their trajectory towards the other edge of the crosswalk. It was also noted that drivers tend to stop short, i.e. those who intended to stop used to do so much earlier than required by the law (i.e. at the white line for stop or crosswalk). Interview responses and observations showed that pedestrians use to understand 'some forms of movement from the vehicle as communicating a message'. For example, [15] and [47] showed evidence that such implicit signalling through speed and positioning are the main form of signalling used in road-crossing interactions, as explicit forms of signalling such as hand gestures and facial expressions are not often used.

Domeyer et al [24] investigated the quantitative parameters (i.e. time) of pedestrian–vehicle interactions at four pedestrian crossings, using annotated videos. In particular, the authors were interested in the effects of vehicle stopping short time (i.e. their proximity with the pedestrians). Their results showed that the median short stop time was around $1s$. They also found that vehicles, that had higher short stop times, were creating more safety margins, thus were more delayed. However, it was found that the stopping short time did not increase the overall time that the vehicle and the pedestrian would spend at an intersection.

2.4 Trust in Human–Robot Interactions

Various definitions of trust have been used for human–robot interactions. This section introduces some of these definitions and reports findings from several studies.

For instance, Lee and See [46] reviewed the concept of trust in automation. They defined trust as an 'attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability'. In [71], Smithson described trust as 'a psychological state that entails the willingness to take risks by placing oneself in a vulnerable position with respect to the trustee'. He described uncertainty as being prevalent to a trust relationship, there is no trust without any risks. Henschke [36] described trust as a 'key value' in the development of autonomous systems. This paper discussed the ethical issues with autonomous systems but also referred to trust in these systems as a complex concept which could be defined as either reliability, predictability, goodwill, affect or public trust.

Floyd et al. [26] introduced the idea of inverse trust. It proposed a mathematical decision model for an autonomous system to measure the level of trust of a human team-mate and then adapts its own behaviour accordingly. Devitt et al. [23] described that with complex and intelligent autonomous systems, humans could become 'overly trusting or overly skeptical', especially when robots become intelligent enough and could manipulate their trust. Agrigoroaie and Tapus [2] focused their work on human informal behaviour and proxemics. The study showed that autonomous systems that are capable of understanding the processes behind human decision-making can have better interactions with them and are more likely to be trustworthy.

In [80], van den Brule et al. argued that not only the robot performance is important but also its behavioural style can have some influence on people's level of trust. Their experiments in video and in VR showed that task performance is key for trustworthiness but that the robot behavioural style was also significant in the videos. Lewis et al. [48] explained that trust is dynamic (i.e. changing over time), and in [62] trust towards automation is directly related to reliance.

2.5 Trust in Human–AV Interactions

The study of pedestrian trust in AVs is a recent research topic. Previous work has mainly investigated the concept of trust for passengers of autonomous vehicles during shared-driving mode [4] [19]. More often, pedestrian trust in AVs has been investigated via the design and testing of external Human–Machine Interfaces [20] [54]. Rothenbuecher et al. [67] found that pedestrians lacked trust when interacting with a vehicle 'disguised' into an AV because they could not see a human driver inside, but at the same time they expected to trust more the AV because of its algorithmic capabilities. Deb et al. [22] performed a study using questionnaires to evaluate pedestrian receptivity towards autonomous vehicles, showing that males trust AVs more than females. The authors also warn that pedestrians could take advantage of perfectly safe autonomous vehicles.

Saleh et al. [68] proposed a framework that relies on social cues, e.g. intent understanding, to model trust between vulnerable road users and autonomous vehicles. Reig et al. [64] studied pedestrian trust in autonomous vehicles via interviews, showing for example that participants who were favourable to AVs were more likely to trust them and that the lack of knowledge about AV technology leads to mistrust. Using the definition of trust in [46] introduced above, Jayaraman et al. [38] studied pedestrians' trust in autonomous vehicles in a VR experiment followed by a questionnaire using a Likert scale. It is argued that human trust increases with the increase of available information, and found that AV's driving behaviour and the presence of light can influence the trust of pedestrians. This study also showed correlations between pedestrian behaviour (distance to collision, gaze and jaywalking time) and their trust towards the AV.

2.6 Research Aims

Despite the numerous reviewed studies on proxemics and trust from the social science and human–robot interaction research communities, many works rely on qualitative or discretized findings from human experiments using questionnaires, interviews and video analyses. Pedestrian proxemics and trust are very recent topics in the context of autonomous vehicles research. No found studies have inferred continuous valued human proxemic utilities as now required by AV controllers, or linked these to trust concepts. There is little agreement on the definition of trust and new trust concepts are regularly proposed, which are mostly informal rather than directly implementable as mathematics and software for autonomous vehicles. Thus, the rest of the paper will contribute towards filling these gaps.

Summary of contributions This paper proposes:

- A novel Bayesian approach to infer proxemic utility functions;
- A new concept and mathematisation of 'physical trust requirement' for pedestrian–AV interactions, and also applicable to more general human–robot interactions which can numerically generate and explain Hall's proxemic zones;
- Empirical results of our method on two public datasets to infer pedestrian proxemic utility functions and trust zones.

3 Proxemic Utility Modelling

Our method consists in inferring the proxemic utility function of pedestrians from existing public datasets from interactions between human drivers and pedestrians. No new empirical experiments are performed in this study. Bayesian theory is used to fit parameters and compare competing models. The approach is first validated on simulated data whose ground truth correct answer is available, before running on empirical data from two public datasets in section 5.

3.1 Proxemic Utility Definition

It is possible to measure the utilities (i.e. perceived costs and/or benefits) which humans assign to states of the world, by asking for or otherwise observing their preferences between states. Such preference orderings for rational agents can be shown to be mathematically equivalent to the assignment of a single number to each state, which is defined as the utility. This mapping from states to numbers is called the utility function [6].

We consider utility functions U as models M with parameters $\theta = \{a_0, .., a_n\}$,

$$U = M(X, a_0, .., a_n), \tag{5.1}$$

that assigns a real value U to the state X.

We assume that human proxemic utility can be described by such a parametric model with the state X being the physical distance between the two agents. Based on our prior knowledge from Hall's theory, we expect the size of the negative utilities to roughly reduce with distance, so we choose several candidate parametric models, M, with a variable number of parameters, θ, including a hyperbolic function (5.2), a Gaussian function (5.3) and different degrees of polynomials (5.4),

$$M_{hyperbolic}(X, \theta) = a_0 X^{-1}, \tag{5.2}$$

$$M_{Gaussian}(X, \theta) = \mathcal{N}(X, a_1 = \mu, a_0 = \sigma^2), \tag{5.3}$$

$$M_{polynomial(n)}(X, \theta) = a_n X^n + a_{n-1} X^{n-1} + ... + a_1 X + a_0. \tag{5.4}$$

We chose these candidate functions via three considerations. First, if we assume very little about the form of the function – just that it is reasonably smooth – then we need to have at least one highly flexible generic model which is able to fit to any smooth function. This is delivered by the polynomial candidate. Second, we have a prior scientific intuition – a hypothesis to test – that the function will be roughly hyperbolic shaped, starting high and falling off with distance. We include a hyperbolic model for this reason. Finally, the Gaussian is included just because it is a common function which often emerges in solutions of physical processes and easy to include. If additional candidates are proposed in the future, they can also be tested against the ones included here.

Throughout this paper, we assume that all agents are rational and that utility can be measured in units of seconds (roughly equivalent to 'time is money'). Human pedestrians and drivers assign a value of travel time in their journeys [82] [17] [37] [5] [1], and using this as the unit will simplify the analysis. We do not model the negative utility of a crash as an additional explicit term because the proxemic model is already able to include it as the utility of a zero distance contact.

3.2 Proxemic Utility Inference Method

A Bayesian inference method is used to infer the proxemic utility functions from observed data. It consists in fitting different parametric models to the data in order to obtain the best parameters for each model. The observations are the distances between the two agents, X, their speeds, v and v_{ped}, and the outcomes of the interactions (pedestrian crossing or stopping). We used nonlinear least squares optimisation (implemented via the Python Scipy.optimize package) for the model fitting. At each optimisation iteration, we used the candidate model parameters proposed by the optimiser to compute optimal actions for the pedestrian for every possible distance X. These optimal actions are compared against the actual actions seen in the data, for the particular distances in the data, and this comparison is used to compute the probability that given the model, the proposed parameters are the true ones.

This is done using Bayes' theorem as follows: under a given model, M, with parameters θ and data D, we have,

$$P(\theta|M, D) = \frac{P(D|\theta, M)P(\theta|M)}{\sum_{\theta'} P(D|\theta', M)P(\theta'|M)}. \tag{5.5}$$

We assume a flat prior over θ so that,

$$P(\theta|M, D) \propto P(D|\theta, M), \tag{5.6}$$

which is the data likelihood, given by,

$$P(D|\theta, M) = \prod_i P(A_i|x_i, x_{ped_i}, v, v_{ped}, \theta, M'), \tag{5.7}$$

where A_i is the pedestrian observed action choice, e.g. crossing or stopping, x_i and x_{ped_i} are observed car and pedestrian locations at the start of an interaction and v and v_{ped} are observed car and pedestrian speeds. M' is a noisy version of the optimal model M, which plays actions from M with probability $(1 - s)$ and maximum entropy random actions (0.5 probability of each speed) with probability s. This is a standard noise modification, used for example in psychological Bayesian data analysis [49] [16][11], which allows the model to fit data where agents have made deviations from perfectly optimal strategies. Without this noise term, the model would assign probability zero to any deviation from perfect behaviour. But humans – and most other objects modelled using statistics – rarely behave exactly according to any mathematical model, so the noise term enables the models to fit approximate behaviours.

3.3 Model Comparison

To select the best fitting proxemic utility function from the set of candidate models M_i, we would like to compute and take the maximum of $P(M_i|D)$. This is computationally hard due to a required integral over the parameters of the models,

$$P(M_i|D) = P(M_i) \int_{\theta_i} P(D|M_i, \theta_i)P(\theta_i|M_i)\theta_i. \tag{5.8}$$

We instead compute and use the Bayesian Inference Criterion, (BIC) [70] which is a standard approximation to this integral,

$$BIC = \log(n)K - 2\log(L) \approx P(M_i|D). \tag{5.9}$$

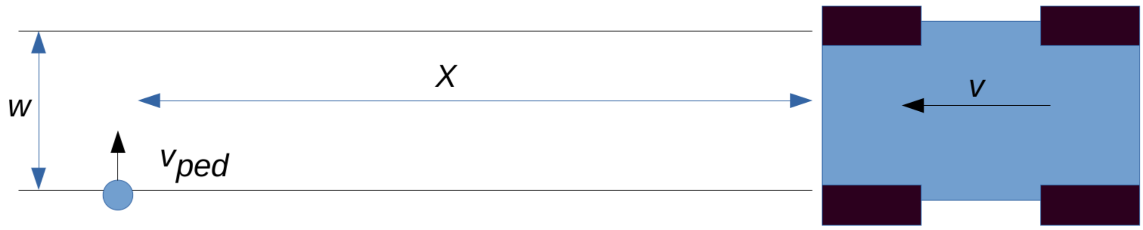


Figure 5.2: Pedestrian–Vehicle Interaction Simulation

The integral, and the BIC approximation to it, are able to correctly compare competing models M_i in cases where the models have differently (K) sized parameter spaces, by combining the likelihood $L = P(D|M_i, \hat{\theta}_i)$ of n observations in data D under the model M_i with the Occam factor arising from the prior over the model's parameter space, $P(\theta_i|M_i)$, assuming a flat prior on the models themselves, $P(M_i) = P(M_j)$. This automatically and correctly penalises models with many parameters for potentially overfitting to data [70].

3.4 Validation via a Simulation Study

To validate our proxemic utility inference method, we developed a simulation with a simple crossing scenario with a pedestrian and a car on a road with a width w, as shown in Fig. 5.2. We simulate the internal reasoning of a pedestrian based on a known (ground-truth) proxemic utility function and the vehicle time utility for a crossing decision. Simulated pedestrian behaviour data is generated, and used to infer back the proxemic function. Validation occurs if the inferred proxemic function matches the input proxemic function used to generate the behaviour.

3.4.1 Assumptions

The purpose of the simulation is only to validate that the system is able to recover the ground truth (i.e. infer the ground truth values used as inputs to the simulation back from the output of the simulation). It does not matter which particular ground truth is used for validation. So to create the simulated data, we choose the following arbitrary settings: the car moves at a constant speed $(2m/s)$ and the pedestrian is standing at the edges of a crosswalk, ready to cross. The pedestrian also moves at constant speed, $1m/s$. The pedestrian is assumed to have an internal reasoning about the utility of crossing and avoid a potential crash with the car. They compare the negative utility (effects) caused by the proximity with the car with the time delay that would occur if they wait for the car. If the proximity cost (measured in seconds, assuming time is a currency) is less than the time delay, i.e. if they are able to cross before the car reaches the intersection, then they are incentivised to do so.

3.4.2 Data Generation and Inference Results

We generated data from a pedestrian–vehicle interaction simulation, using a predefined proxemic utility function. We defined random starts for the vehicle, to create 1000 different pedestrian–vehicle interactions. We then used the data collected to implement and test our inference method to recover the original proxemic utility function. Examples of functions that we tested are shown below.

Hyperbolic Function Firstly, we evaluated our inference method with a ground truth hyperbolic proxemic function,

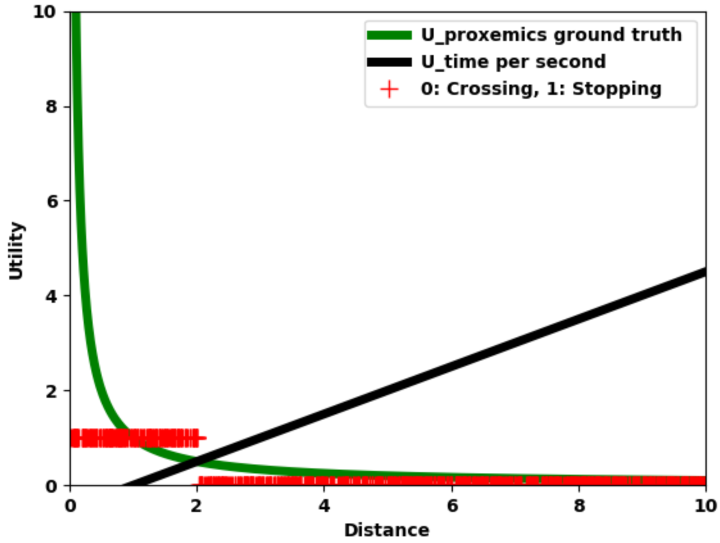
$$M_{hyberbolic}(X, a_0) = a_0 X^{-1}, \tag{5.10}$$

with $a_0 = 1$, as shown in Fig. 5.3a along with the time utility function and the crossing decision for the interactions. As we can see in the results of the model fitting, in Fig. 5.3b, the best model is the hyperbolic function with the maximum likelihood (loglik $= -105.36$) and the lowest BIC value (BIC $= 217.629$). All other models have a lower likelihood and a higher BIC value, for example, the second best model is the quadratic function with a likelihood of -107.55 and a BIC equal to 235.839.

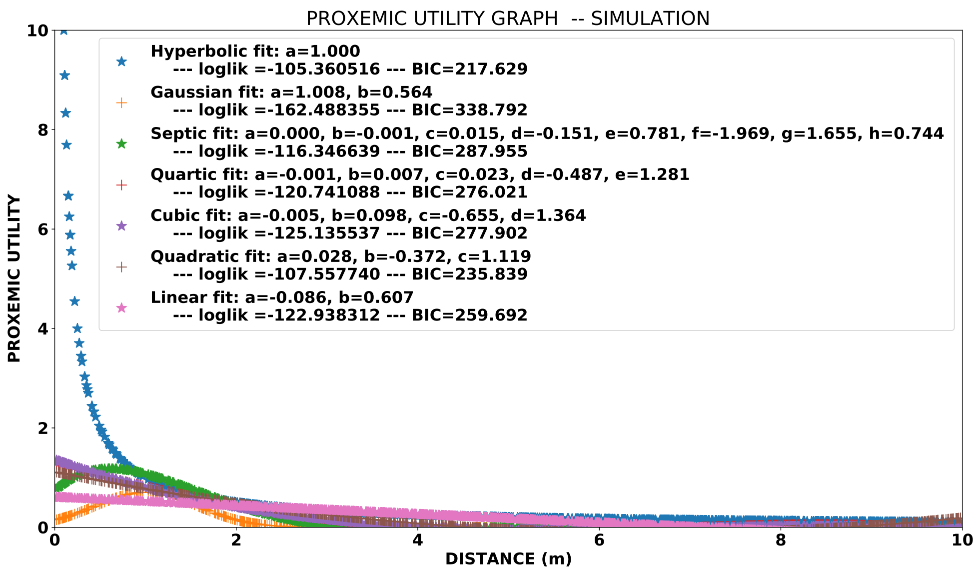
Quadratic Function Secondly, we used an arbitrary quadratic function,

$$M_{quadratic}(X, a_2, a_1, a_0) = -X^2 + 5X + 25, \tag{5.11}$$

as the ground truth. Fig. 5.4a shows the ground truth quadratic proxemic and time utility functions with the pedestrian crossing decisions. As shown in Fig. 5.4b, the best model is the quadratic function with the maximum likelihood (loglik $= -1089.72$) and the lowest BIC value (BIC $= 2200.158$). All other models have a lower likelihood and a higher BIC value, for example, the second best model is the cubic function with a likelihood of -1109.49 and a BIC equal to 2246.615.

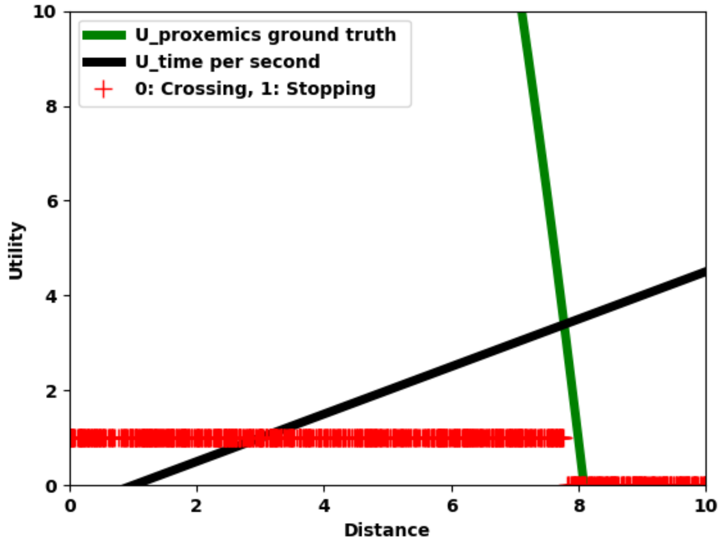


(a) Ground truth utility functions

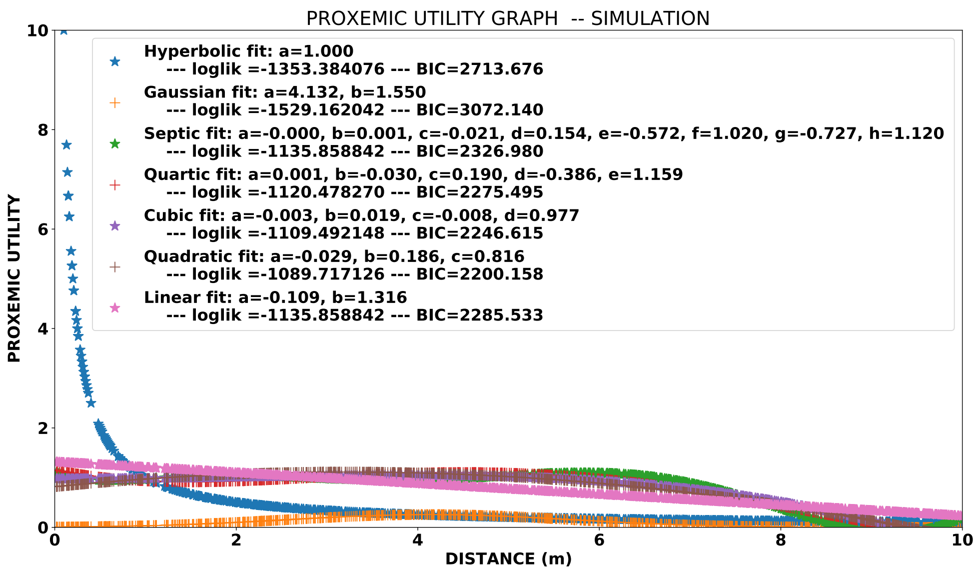


(b) Model fitting results

Figure 5.3: Simulation with a hyperbolic proxemic function



(a) Ground truth utility functions



(b) Model fitting results

Figure 5.4: Simulation with a quadratic proxemic function

Quartic Function Thirdly and lastly, we evaluated our method with an arbitrary quartic function, (i.e. polynomial function of degree 4),

$$M_{polynomial(4)}(X, a_4, a_3, a_2, a_1, a_0) = -0.08X^4 - X^3 + 3X + 0.5, \tag{5.12}$$

as the ground truth as shown in Fig. 5.5a along with the time utility function and the crossing decision for the interactions. The results of the model fitting are shown in Fig. 5.5b. The quartic and septic functions have the maximum likelihood (loglik = −122.93) but the quartic function is ranked as the second best model according to the BIC values with a BIC equal to 280.415. Instead, the Gaussian model (loglik = −129.52, BIC = 272.875) is selected as the best model due to its lower number of parameters. However, we can note here that the shape of the ground truth function shown in Fig. 5.5a looks very similar to a Gaussian, so the selection of the Gaussian model for this case is perfectly understandable.

The above results show that our proposed method for inferring proxemic utility function works on simulated data and with different ground utility functions.

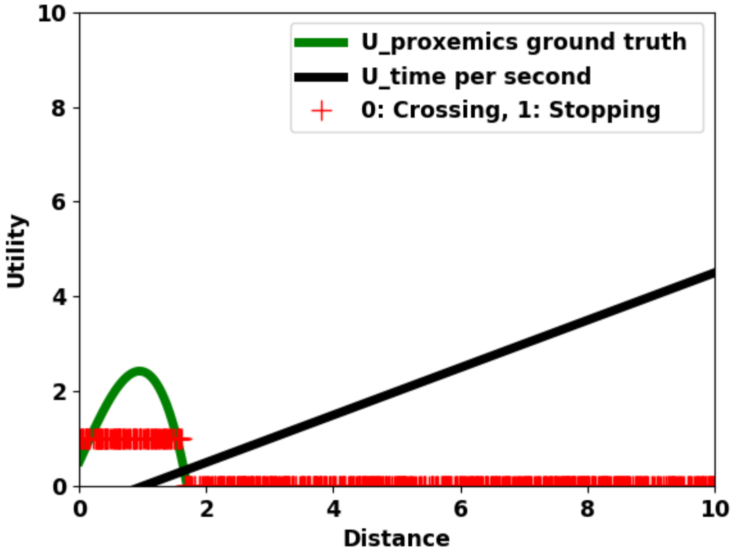
4 Physical Trust Requirement

4.1 Trust Definition

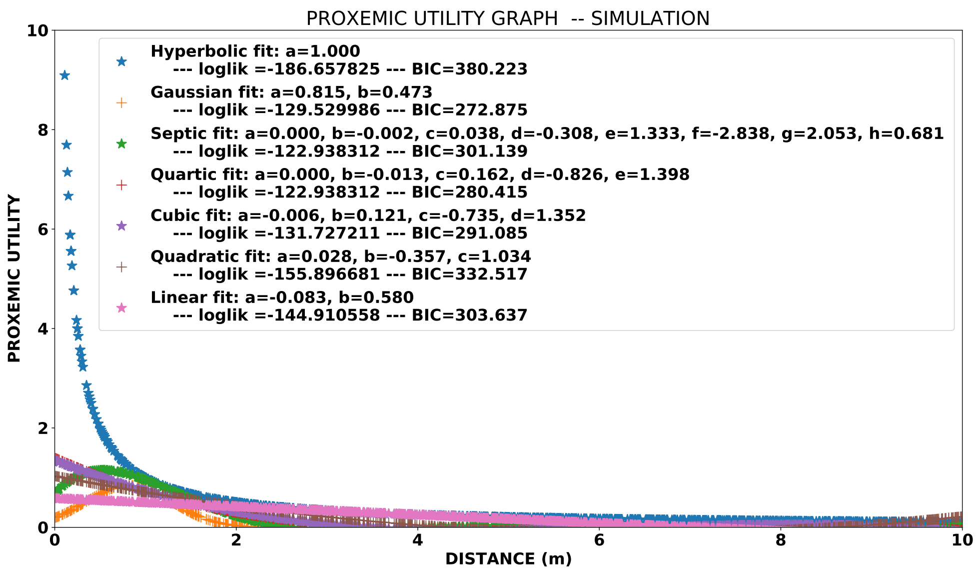
Refining Lee and See's concept of trust [46] reviewed above, where trust is defined as an *attitude* in 'a situation characterised by uncertainty and vulnerability', we define a new related concept: *physical trust requirement* (PTR), a Boolean property of the physical state of the world (not of the psychology of the agents) with respect to one agent during an interaction, true if and only if the agent's future utility is affected by an immediate decision made by another agent.

We thus measure the *need* for trust from pedestrian behaviour in uncertain situations. The PTR divides the proxemic function into three zones as shown in Fig.5.6, as the PTR is true in the trust zone and false in the crash and escape zones. We made some assumptions and used numerical values to obtain specific equations and numbers for the three zones in our road crossing case:

1. ***Crash zone***: This is the region very close to the human agent, where they will be affected by negative consequences and no-one can prevent them from occurring, so no trust is involved. In the road-crossing case, this occurs when the pedestrian is in the



(a) Ground truth utility functions



(b) Model fitting results

Figure 5.5: Simulation with a quartic proxemic function

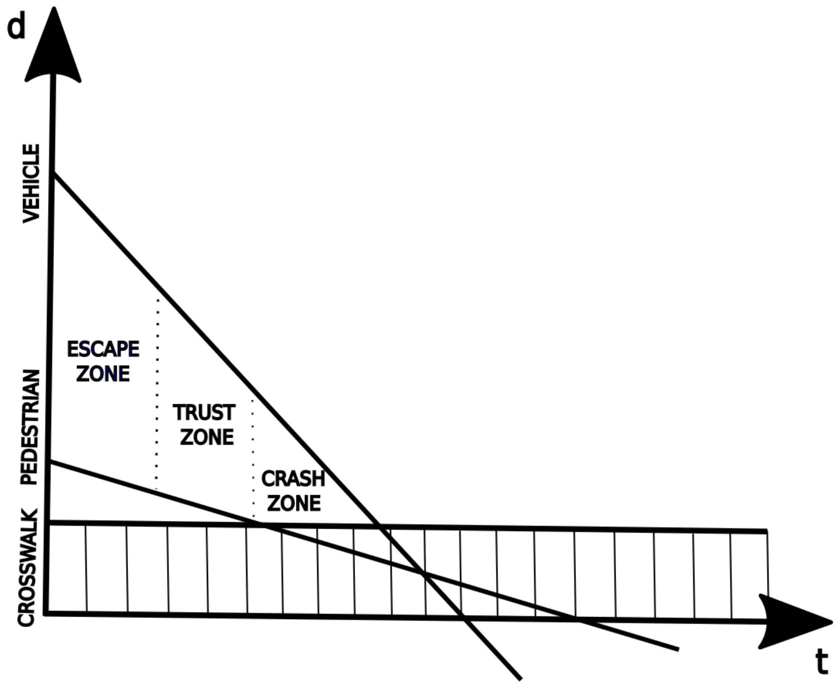


Figure 5.6: Proxemics–Trust relation in pedestrian–vehicle interaction

road and the car is very close, with neither able to run or brake to prevent the collision.

The crash zone, $\{d : 0 < d < d_{crash}\}$, is the region delimited by the reaction and braking distances of the vehicle, given by the standard stopping distance equation [51],

$$d_{crash} = vt_{driver} + \frac{v^2}{2\mu g}, \tag{5.13}$$

where the first term depends on the human driver's psychological thinking reaction time, t_{driver}, and the second term represents the physical braking distance (depending on the physical friction between tyres and tarmac, and equal to the length of any physical skid marks left by the vehicle after the driver begins to apply the brakes), v is the vehicle speed, μ the coefficient of friction and g the gravity of Earth.

2. ***Escape zone***: This defines the area where the human agent is able to choose their own action to avoid the negative utility, rather than relying on the other agent. As such, it does not need to trust the other agent. In our road-crossing case, this occurs when the vehicle is further away from the pedestrian, so that the pedestrian has time to act and

save themself without trusting the vehicle to yield.

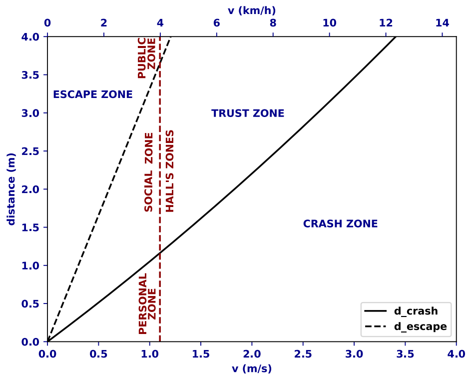
The escape zone, $\{d : d_{escape} < d\}$, is the set of distances beyond which pedestrians do not fear any potential danger from the vehicle. In this zone, pedestrians can complete their crossing before the vehicle arrives. The *escape distance* d_{escape} is the minimum distance at which this is the case. Consider the time $t_{cross} = w/v_{ped}$ it takes for the pedestrian to cross, during this time, the vehicle moves by distance wv/v_{ped}, where v_{ped} is the pedestrian speed and w is the width of the road. When we also add the distance moved by the vehicle during t_{ped}, the human pedestrian's reaction time to make their crossing decision before starting to walk or not walk, then we obtain the escape distance,

$$\begin{aligned} d_{escape} &= vt_{ped} + vt_{cross} \\ &= vt_{ped} + w\frac{v}{v_{ped}}. \end{aligned} \tag{5.14}$$

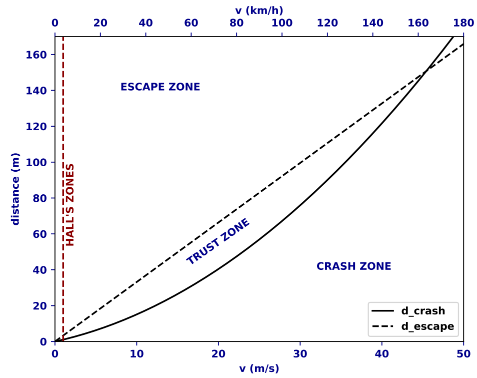
This escape distance then defines the start of the escape zone.

3. ***Trust zone***: We define the trust zone as the region of the proxemic function where the PTR is true. The *other* agent (e.g. the car) can *choose* (e.g. by slowing down) to prevent them from receiving negative effects (e.g. collision), but the human is incapable of making any action to affect the utility outcome themself. In the road crossing case, this occurs when the pedestrian cannot get out of the car's way in time to avoid collision, but the car is able to brake and yield to prevent the collision if it chooses to do so. This excludes the crash zone in which neither agent has any available choice to avert collision, and also excludes the escape zone. So the trust zone is $\{d : d_{crash} < d < d_{escape}\}$, the intermediate space between the crash and escape zones.

When the pedestrian is in the crash zone, the vehicle has no possibility to avoid an accident, whereas in the escape zone the pedestrian can always cross safely. When the pedestrian is in the trust zone, the vehicle has the sole power to decide if a collision will occur. It is thus in the trust zone that it would be important to study whether and how people do or should trust autonomous vehicles or not.



(a) At lower speeds.



(b) At higher speeds.

Figure 5.7: Distances and zones predicted by the PTR model for different car speeds v (5.7a is a close-up of 5.7b)

4.2 Zones Analysis – Comparison with Hall's Zones

We here derive some mathematical results from our zone definitions and link them to previous results on Hall's proxemic zones. Fig. 5.7 shows the distances d_{crash} and d_{escape} and the zones defined by equations 6.1 and 5.14, for variable vehicle speeds v. We here assume: $w = 2m$ for the road width, $t_{driver} = 1s$ as the driver reaction time [28] [21], $v_{ped} = 1.1m/s$ as the average walking speed of the pedestrian [25] [41] [60], $t_{ped} = 1.5s$ as the pedestrian reaction time (chosen to be similar to the driver reaction time but a little larger because drivers may be more focussed on their task than pedestrians) [18], $\mu = 1$ for the coefficient of friction [34] [61] and $g = 9.8m/s^2$ for the gravity of Earth.

By comparison, the related work review found that Hall zones for human–human interactions are usually reported to be around: intimate up to 0.45cm, personal up to 1.2m, social up to 3.6m, and public beyond this [45].

The vertical line in Fig. 5.7 shows the case $v = 1.1m/s$ in which the vehicle has the same speed as the pedestrian, i.e. the vehicle is behaving as if it was a second pedestrian interacting with the first. In this case, the size of the Hall personal zone, 1.2m, closely matches that of our crash zone in Fig. 5.7a, $d_{crash} = 1.16m$ when $v = 1.1m/s$ (as would be the case when the other is another human rather than a vehicle) and retaining other parameters (including, quite unrealistically, retaining the friction model and coefficient walking rather than wheels). The size of the Hall social zone, 3.6m, also closely matches our $d_{escape} = 3.65m$ from the graph.

We also note that Fig. 5.7a predicts that social human–robot interactions in which the robot is *slower* than a human, as is the case for most humanoids, will have smaller crash and trust zones, which matches the related work reviewed in which personal and social zones were found to reduce compared to human–human proxemics. Also, the trust region in Fig 5.7b gets smaller with speed, reaching zero width when linear and quadratic curves meet at around $45m/s = 162km/h$. This is quite close to official and unofficial speed limits on most countries' motorways/freeways.

If we further define and consider R, the zone ratio given by the width of the trust zone relative to the speed of the car,

$$R = \frac{D_{escape}}{D_{crash}} = \frac{vt_{ped}+v(w/v_{ped})}{vt_{driver}+v^2/2\mu g} = \frac{t_{ped}+(w/v_{ped})}{t_{driver}+v/2\mu g}. \tag{5.15}$$

Then we see that as vehicle speed increases, the effect of t_{driver} becomes negligible, and the zone ratio tends to zero, meaning that the crash zone's size comes to dominate the others:

$$v \to \infty \Rightarrow R \to \frac{2\mu g(t_{ped} + (w/v_{ped}))}{v} \to 0, \tag{5.16}$$

and as vehicle speed decreases, the zone size ratio converges to a constant:

$$v \to 0 \Rightarrow R \to \frac{t_{ped} + (w/v_{ped})}{t_{driver}}, \tag{5.17}$$

which shows that if the ratio of zone sizes is considered rather than their absolute size, then all dependency on friction and gravity has vanished in the high and low speed limits. Thus, all road and car specific concepts have vanished to leave a more general proxemic relationship which may be of interest in general human interaction cases rather than only road-crossings. Fig. 5.8 shows the variation of R relative to the speed of the car, and that the value of R in Eq. (5.17) tends to the constant 3.5.

5 Empirical Data Study

To demonstrate the inference of empirical pedestrian proxemic utility functions, we then apply the method to data from real-world pedestrian interactions with manual driven vehicles. We used two public datasets containing tracking data from multiple road users. We only considered the interactions where the pedestrian crosses or stops for utility, i.e. when the gap is greater than the safety distance so that we can learn how the pedestrian adjusts their comfort zone. We then compute the PTR zones for these datasets.

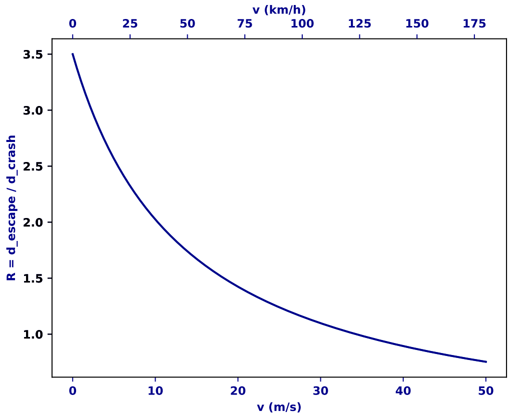


Figure 5.8: The ratio of escape zone size to crash zone size, R, decreases as the car speed v increases, showing that the crash zone dominates at high speeds.

5.1 Datasets

5.1.1 Daimler Pedestrian Benchmark

The Daimler dataset [43] contains 58 pedestrian–vehicle trajectory data and annotations, such as pedestrian crossing decisions. The dataset was not collected from real-world interactions, the pedestrians and drivers were actors. The authors created these interaction scenarios for their work, 44 of these were pedestrian crossing scenarios and the other 14 interactions were stopping scenarios. Fig. 5.9 shows a dash cam image of one interaction scenario. The distribution of vehicle and pedestrian speeds in the dataset is shown in Fig. 5.10.

5.1.2 inD (Intersection Drone Dataset)

The inD dataset [7] is a newly released dataset which provides road users (cars, trucks, cyclists, pedestrians) tracking data. There are 32 videos recording data from 4 different intersections in the dataset, which contains thousands of real-world interactions. But as the videos were not released with the trajectory data, we decided to focus on one intersection, where there is clearly a pedestrian crosswalk, thus pedestrians crossing the road would necessarily interact with the upcoming vehicles. Twelve recordings (n°18 to n°29) contain data from the crosswalk shown in Fig. 5.11. The distribution of vehicle and pedestrian speeds in the dataset is shown in Fig. 5.12.

Figure 5.9: Pedestrian intention with a vehicle, from its dashcam, in the Daimler dataset [43].

Figure 5.10: Histograms of vehicle and pedestrian speeds in Daimler dataset, showing that average speeds $v \approx 5.25m/s$ and $v_{ped} \approx 1.60m/s$ are good approximations.

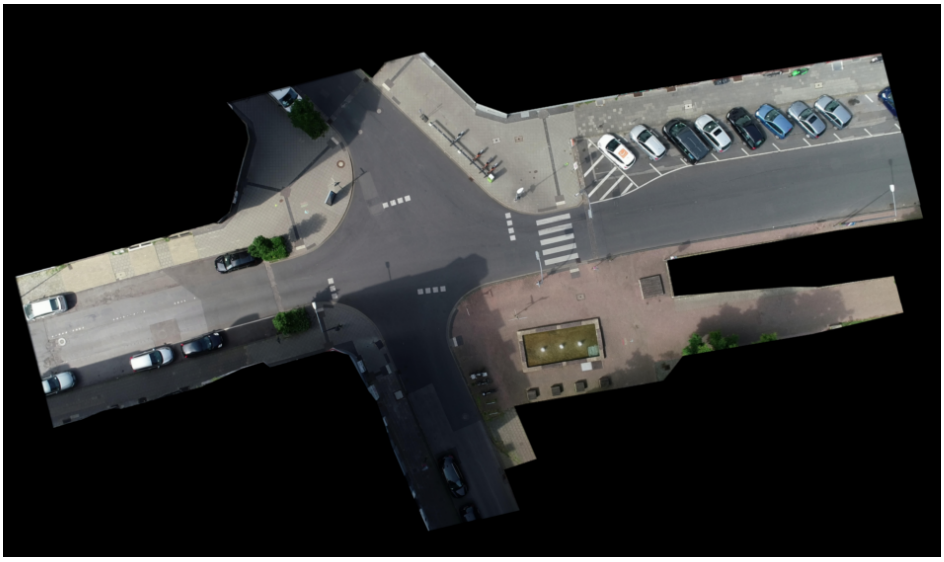


Figure 5.11: Crosswalk in inD dataset [7]

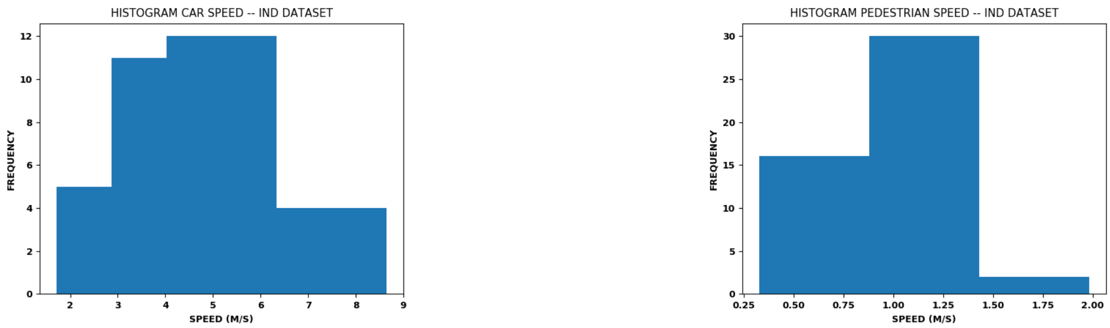


Figure 5.12: Histograms of vehicle and pedestrian speeds in inD dataset, showing that average speeds $v \approx 4.79 m/s$ and $v_{ped} \approx 0.99 m/s$ are good approximations.

5.1.3 Criteria for interactions' selection

As inD dataset contains multiple classes of road users but we were interested in pedestrian–vehicle interactions only, we extracted them from the rest of the data in a semi-automatic manner and annotated them. For each given pedestrian, we find the car that appeared a few frames earlier and then we select the frames where they both appear together. We only kept interactions where the vehicle and the pedestrian were encountering somewhere near the coordinates (x = 62, y = -27), to make sure the pedestrians cross at the crosswalk, not any other locations, where they would jaywalk and we would have no possibility to know the hidden factors behind that decision. We selected trajectories where cars and pedestrians followed a straight path until their encounter, in order to match with our simulation model. We kept pedestrians walking from the bottom right, we didn't consider pedestrians coming from the top right because most of them were not crossing, as there was a car park.

In total, we used the 58 interactions from the Daimler dataset and we collected 48 more interactions from inD dataset, with 24 where the car came from the top right of the image, and the other 24 where cars came from the bottom left of the image. Fig. 5.13 shows some examples of pedestrian–vehicles trajectories from both datasets.

5.2 Proxemic Utility Model Selection

5.2.1 Proxemic Utility Implementation

First, we applied our proxemic utility inference method on the two datasets, similar to the simulation study in section 3.4, except that here we would not know the ground truth function for final comparison. The goal here is thus to infer the unknown proxemic utility function from the data and select the best model with the lowest BIC value.

5.2.2 Proxemic Utility Results

Results of the proxemic utility inference method on the Daimler and inD datasets are shown in Fig. 5.14 and 5.15, respectively. They show that a hyperbolic function best describes pedestrian proxemic behaviour in both cases, with the lowest BIC values (Daimler BIC = 174.482, inD BIC = 62.325). The proxemic utility costs increases with shorter proxemic distances, and with a steep growth near the collision point. These results are consistent with the human experiments in [72] [73], where participants' perception of threat (negative utilities) increases at shorter distances and decreases at longer distances.

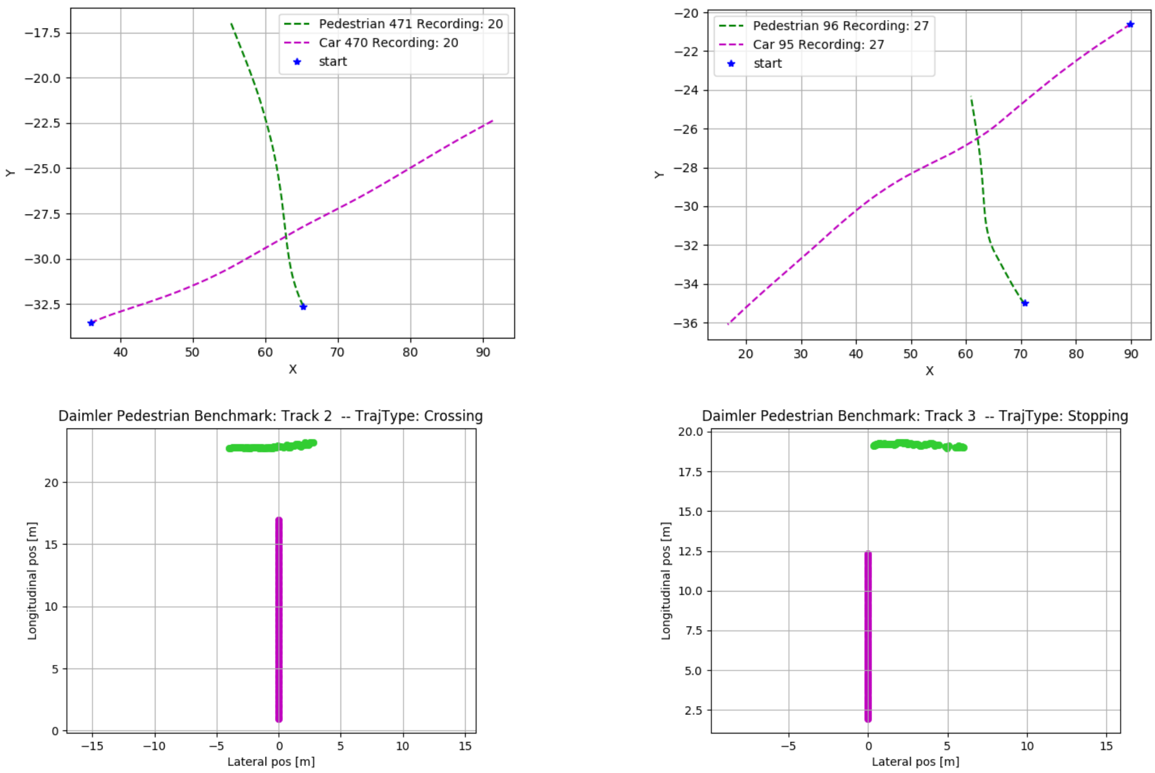


Figure 5.13: Examples of interactions from the datasets

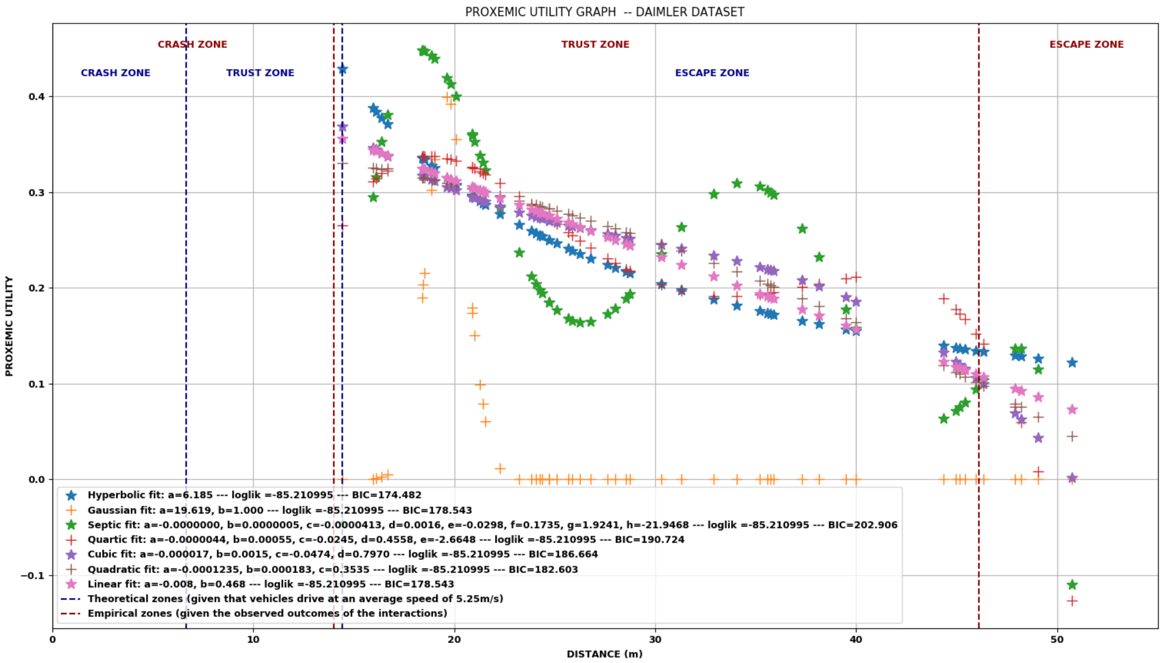


Figure 5.14: Model fitting results for Daimler dataset

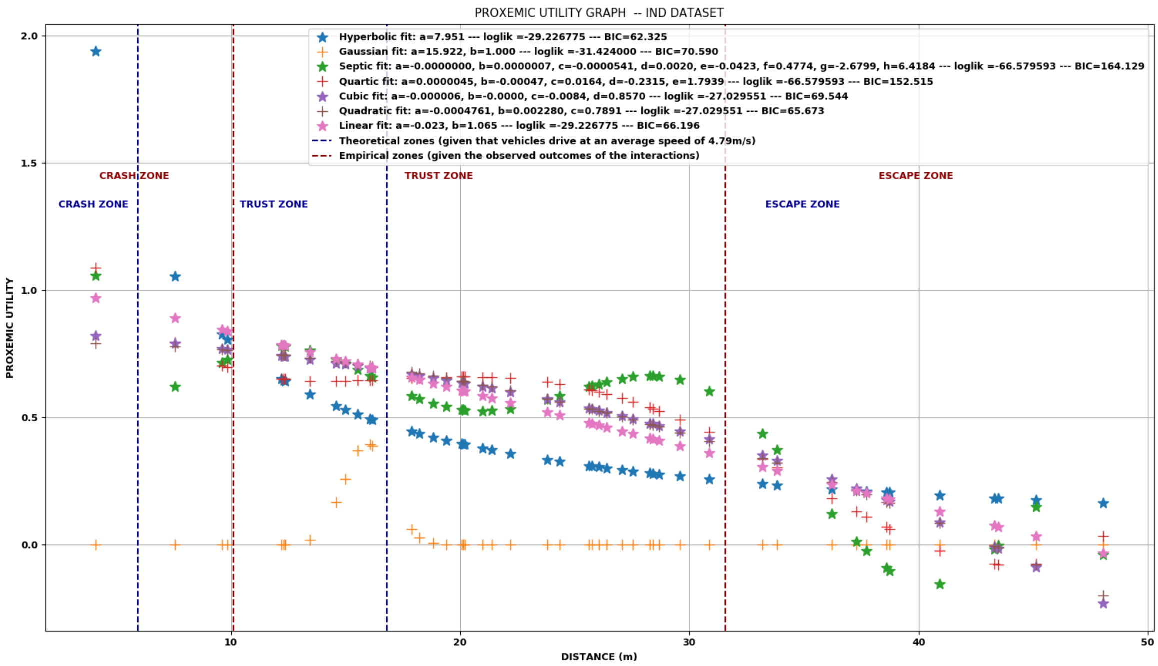


Figure 5.15: Model fitting results for inD dataset

5.3 Zones Computation

5.3.1 Zones Implementation

Second, we computed two different estimates of the zone distances, called 'theoretical' and 'empirical' zones. Both estimates make use of the data. The theoretical estimate makes use only of average speeds from the data, and the empirical estimate makes use of extreme individual behaviours from the data.

We define *theoretical zones* as the solutions of the equations in section 4.1 given by assuming that all vehicles move at the average speed of the vehicles in the dataset, and all pedestrians move at the average speed of pedestrians in the dataset. This assumption is justified approximately by the histograms of these speeds in the datasets, as shown in Figs. 5.10 and 5.12, which show that vehicles are all moving at similar urban speeds of 0–30km/h and pedestrians are all moving at similar walking speeds. The average speed of vehicles in Daimler was $v \simeq 5.25m/s$; and in inD: $v \simeq 4.79m/s$. The average speed of pedestrians was in Daimler: $v_{ped} \simeq 1.60m/s$; and in inD: $v_{ped} \simeq 0.99m/s$. We here use the same constants as in section 4.2, with $w = 2m$ for the road width, $t_{driver} = 1s$ as the driver reaction time, $t_{ped} = 1.5s$, $\mu = 1$ for the coefficient of friction and $g = 9.8m/s^2$ for the gravity of Earth.

We define *empirical zones* by finding in the datasets the maximum distance below which pedestrians always stop and the minimum distances above which they always cross. This is intended to provide only an exploratory measure. It is not a true statistical estimator, because its error increases rather than decreases with sample size due to its dependency on only the most extreme individuals.

5.3.2 Zones Results

Results of the theoretical zone experiments are shown in dark blue in Fig. 5.14 for Daimler dataset and in Fig. 5.15 for inD dataset. The empirical zones are shown in dark red in Fig.5.14 for Daimler dataset and in Fig. 5.15 for inD dataset.

For the Daimler dataset, the theoretical trust zone is between 7–15m and the empirical trust zone is between 14–45m. For the inD dataset, the theoretical trust zone is between 6–17m and the empirical trust zone is between 10–31m.

The theoretical and empirical zones for the two data sets are roughly in agreement which suggests the effect of the actors in Daimler is not important. The boundaries of these zones, both theoretical and empirical, would change if the vehicle drives at a higher or lower speed.

The width of all of our theoretical (crash, trust and escape) zones are smaller than the empirical zones. We found that our theoretical zones were underestimated relative to the empirical zones, by about three times in Daimler dataset and by two times in inD dataset. We compute these coefficients by iteratively updating by increments the theoretical crash and trust zone boundaries. This underestimation of the theoretical zones is expected because we computed them under many simplifying assumptions, including using average speeds across the datasets and guessed other parameters such as the driver reaction time (t_{driver}), the pedestrian reaction time (t_{ped}) and the coefficient of friction (μ). If all the interactions were performed with these average speeds and parameters (t_{driver}, t_{ped} and μ), then the theoretical zones might match the empirical zones. In fact, Fig. 5.16 and 5.17 show the time utilities and outcomes (pedestrian crossing decisions) for each interaction in the Daimler and inD datasets, respectively. In particular, the time utility graphs show the variations of vehicle speeds across the interactions. This may explain why our theoretical trust zones do not match the empirical trust zones. Moreover, if we had computed the theoretical zones for each interaction (with their corresponding speeds), it would not be possible to analyse and to make a general discussion on these zones with respect to the proxemic utility function, which was drawn from all the interactions in each dataset.

For this reason, we will base the rest of our analysis of trust on the empirical zones. We can see that the trust zone is the area of the proxemic utility function where the gradient changes more. This reflects the high uncertainty that lies in the trust zone. The decision of a pedestrian to cross is uncertain here because the pedestrian has to rely on the vehicle to make a decision. In contrast, in the crash and escape zones, we see that the gradient of the proxemic utility function changes less, this is due to the more deterministic outcome in these areas. In the crash zone, the distance and the speed of the vehicle give enough information to the pedestrian for not crossing and in the escape zone, the vehicle behaviour does not interfere into their crossing decision because the danger cannot be perceived by the pedestrian as found in [72] [73], therefore they will cross.

Finally, for the actual average car speeds in the two datasets, equation 5.15, computed by the ratios of the theoretical zones D_{escape}/D_{crash} from Figs. 5.14 and 5.15, gives for Daimler $R = 15/7 = 2.1$, and for inD $R = 17/6 = 2.8$. Using the empirical zone boundaries from the same figures, we obtain Daimler empirical $R = 45/14 = 3.2$; and inD empirical $R = 31/10 = 3.1$. These results closely match the ratio found for Hall's zones in section 4.2.

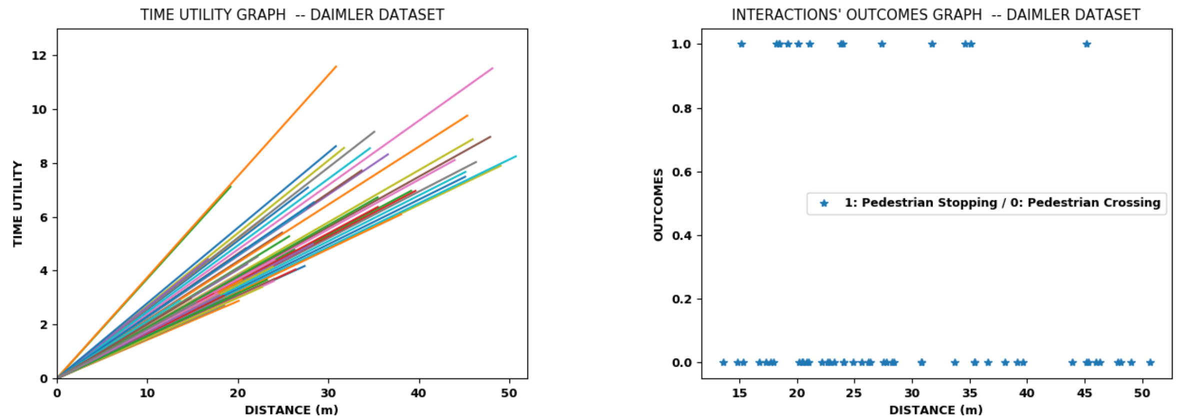


Figure 5.16: Time utility and ground truth interaction outcomes for Daimler dataset

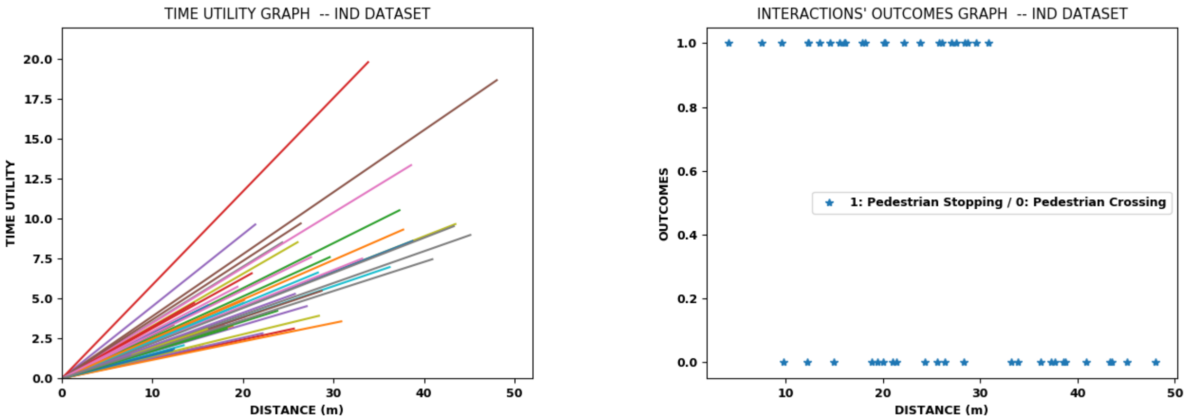


Figure 5.17: Time utility and ground truth interaction outcomes for inD dataset

6 Discussion

Although the proxemic utility inference method has proven successful on simulation and real-world interactions, several simplifying assumptions were made in order to present and test the basic principles of the method, from which future work should try to move away in order to obtain more reliable *results*. In particular, we assumed that all vehicles move at an average vehicle speed rather than their individual speeds, which is a likely cause of the observed discrepancy between the theoretical and empirical trust zones. This discrepancy is a useful self-test of the model's assumptions, so if future work brings them closed that would give some confidence in the proxemic utility results.

The basic premise of this study, as taken from the game theory model conclusions, was that a *proxemic* function captures the feeling of discomfort from space invasion by vehicles. However, speed considerations might be extended into the utility function itself: a pedestrian might feel comfortable standing $10m$ from a car if it is moving towards them at $1m/s$, but not at $10m/s$. Including the speed of the vehicle as an additional parameter in the pedestrian's utility function would formally move future models from being proxemic functions to include a *kinesic* component (i.e. involving speed as well as proximity) as suggested in [24] and this may further improve interaction control.

We assumed that the pedestrian and the vehicle were solely interacting with each other, ignoring simultaneous interactions with other individuals. We also assumed that the agents always moved along straight, orthogonal paths as in the game theory model, thus we did not include the interactions where pedestrians were not crossing straight away. We used only parametric models to infer the proxemic utility function, future work could explore the use of non-parametrics such as Gaussian Processes and compare their performance against the present models, which is possible via the BIC. Reviewed previous work on proxemics has shown that demographics, social, cultural and environmental factors can have an influence on the proxemic distances [55], therefore it would be important to incorporate some of this additional information and to build a more precise inference model on them. Reviewed previous work on human–robot interactions has shown that the physical size such as height of the other agent also affects proxemics zone sizes, which suggests a similar role for physical car sizes in modifying proxemic utilities. In particular, it provides a further explanation for how buying expensive sports utility vehicles (SUVs) can be rational via their infliction of stronger proxemic penalties onto other road users, thus allowing the driver to win more interactions and reduce their own journey times [27].

Additional future work could look into testing our method on human–AV interactions in virtual reality experiments, and demonstrates its effectiveness on a real autonomous system for better interactions with people. In these settings, it would be possible to collect *causal* data rather than the passive data used in the present study, as the vehicle can be actively controlled as an independent variable in order to measure the dependent behaviour of the pedestrian, more clearly separating the causal logic between the two agents during their interaction.

We have mainly focused on pedestrian–vehicle interactions, but the concepts and methods here could be applied to other human–robot interaction tasks. For example, human factory workers collaborating with a robot arm could be modelled by a trust zone in which the arm is able to hit them without time or space to escape.

We have merged Hall's intimate and personal zones to map jointly into our collision zone, and did not attempt to explain any theory of intimacy within this zone. In general proxemics, our collision zone would be the distance at which a physical attack such as a punch or grab (analogous to the vehicle collision) may (a) have already happened or (b) be in unstoppable progress. Possibly this would subdivide with (a) as Hall's intimate zone and (b) as Hall's personal zone, with the width of the intimate zone being the collision area width w.

Using space invasion to inflict small negative utilities via discomfort on members of the public may still be considered unethical or illegal in some cases. In many jurisdictions, such as in the UK, this is an ongoing dilemma under active debate by authorities [79]. We hope the present study will contribute to this debate, by showing how this option trades off against other possible negative utilities, including those inflicted on passengers of such vehicles whose journeys would be delayed by overly assertive pedestrians pushing in front of them. Human drivers already use many such credible threats to encourage pedestrians to get out of the way. In many cases, these threats result in actual collisions. Replacing these threats by automated systems which only invade space rather than potentially collide would improve safety.

7 Conclusions

A previous game theoretic model has suggested that autonomous vehicles must either risk making no progress at all by yielding to all road-crossing pedestrians to stay safe, or maintain a credible threat of actually colliding with them to encourage them to yield. Neither of these are desirable outcomes. The new method developed in the present study now enables the inference of continuous pedestrian proxemic utility functions from pedestrian–driver interaction data. The game theory model shows that this can be used to make their interactions both safe and efficient. This can be done by de-escalating the severe threat of collision to much milder and legally permissible threat of merely invading their personal space to create discomfort as a weaker but still effective penalty for non-collaboration in interactions.

We also defined and mathematically formalised a new concept of trust based on the proxemic function for human–autonomous vehicle interactions. These new, quantitatively defined, zones for the physical trust requirement may assist autonomous vehicle designers in understanding what is meant and required by the concept of trust. The mathematical and empirical results of section 4.2 are evidence that our concept can explain the existence of the classic Hall intimate-personal, social and public zones, quite precisely generating their sizes and ratios, which emerge as a special case for two low speed agents interacting.

Our concept generalises these Hall zones beyond their usual use in human–human interactions to allow for larger zones as the speed of the other agent increases from human to vehicle speed, and shows how trust zones become relatively smaller at higher (e.g. freeway/motorway) speeds. It also generalises to interactions with agents moving *slower* than humans and predicts smaller zones in these cases, which is consistent with the human–robot proxemics studies previously reviewed.

Acknowledgements

This project has received funding from EU H2020 project interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395. The authors are grateful to the Associate Editor and the anonymous reviewers for their valuable time and very useful feedback.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Maya Abou-Zeid, Moshe Ben-Akiva, Michel Bierlaire, Charisma Choudhury, and Stephane Hess. Attitudes and value of time heterogeneity. *90th Annual Meeting of the Transportation Research Board*, 2011.

[2] Roxana Agrigoroaie and Adriana Tapus. Cognitive performance and physiological response analysis. *International Journal of Social Robotics*, pages 1–18, 2019.

[3] Mohammedreza Asghari Oskoei, Michael Walters, and Kerstin Dautenhahn. An autonomous proxemic system for a mobile companion robot. In *Proceedings of the AISB 2010 Symposium on New Frontiers for Human Robot Interaction*. AISB, 2010.

[4] Chandrayee Basu and Mukesh Singhal. Trust dynamics in human autonomous vehicle interaction: A review of trust models. In *2016 AAAI Spring Symposium Series*, 2016.

[5] Richard Batley, John Bates, Michiel Bliemer, Maria Börjesson, Jeremy Bourdon, Manuel Ojeda Cabral, Phani Kumar Chintakayala, Charisma Choudhury, Andrew Daly, Thijs Dekker, et al. New appraisal values of travel time saving and reliability in great britain. *Transportation*, 46(3):583–621, 2019.

[6] José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

[7] Julian Bock, Robert Krajewski, Tobias Moers, Lennart Vater, Steffen Runde, and Lutz Eckstein. The ind dataset: A drone dataset of naturalistic vehicle trajectories at german intersections. *arXiv preprint arXiv:1911.07602*, 2019.

[8] Rodney Brooks. The big problem with self-driving cars is people and we'll go out of our way to make the problem worse, 2017.

[9] Fanta Camara, Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, and Charles W. Fox. Pedestrian models for autonomous driving Part I: low-level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[10] Fanta Camara, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Anna Schieben, Gustav Markkula, Fabio Tango, Natasha Merat, and Charles W. Fox. Pedestrian models for autonomous driving Part II: high-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[11] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. Towards pedestrian-av interaction: method for elucidating pedestrian preferences. In *IEEE/RSJ Intelligent Robots and Systems (IROS) Workshops*, 2018.

[12] Fanta Camara, Patrick Dickinson, Natasha Merat, and Charles W. Fox. Towards game theoretic av controllers: measuring pedestrian behaviour in virtual reality. In *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshops*, 2019.

[13] Fanta Camara, Patrick Dickinson, Natasha Merat, and Charles W. Fox. Examining pedestrian behaviour in virtual reality. In *Transport Research Arena (TRA) (Conference canceled)*, 2020.

[14] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proceedings of the 15th International Conference on Intelligent Autonomous Systems workshops*, 2018.

[15] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. In *Proceedings of IEEE 21st International Conference on Intelligent Transportation Systems*, 2018.

[16] Fanta Camara, Richard Romano, Gustav Markkula, Ruth Madigan, Natasha Merat, and Charles Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Measuring Behavior 2018: 11th International Conference on Methods and Techniques in Behavioral Research*. Manchester Metropolitan University, March 2018.

[17] Carlos Carrion and David Levinson. Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, 46(4):720–741, 2012.

[18] Qianwen Chao, Zhigang Deng, and Xiaogang Jin. Vehicle–pedestrian interaction for mixed traffic simulation. *Computer Animation and Virtual Worlds*, 26(3-4):405–412, 2015.

[19] Jong Kyu Choi and Yong Gu Ji. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, 31(10):692–702, 2015.

[20] Michael Clamann, Miles Aubert, and Mary L Cummings. Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles. Technical report, 2017.

[21] L.C. Davis. Modifications of the optimal velocity traffic model to include delay due to driver reaction time. *Physica A: Statistical Mechanics and its Applications*, 319:557 – 567, 2003.

[22] Shuchisnigdha Deb, Lesley Strawderman, Daniel W Carruth, Janice DuBien, Brian Smith, and Teena M Garrison. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. *Transportation research part C: emerging technologies*, 84:178–195, 2017.

[23] S Kate Devitt. Trustworthiness of autonomous systems. In *Foundations of trusted autonomy*, pages 161–184. Springer, Cham, 2018.

[24] Joshua Domeyer, Azadeh Dinparastdjadid, John D Lee, Grace Douglas, Areen Alsaid, and Morgan Price. Proxemics and kinesics in automated vehicle–pedestrian communication: Representing ethnographic observations. *Transportation Research Record*, 2673(10):70–81, 2019.

[25] Kay Fitzpatrick, Marcus A Brewer, and Shawn Turner. Another look at pedestrian walking speed. *Transportation research record*, 1982(1):21–29, 2006.

[26] Michael W Floyd, Michael Drinkwater, and David W Aha. Adapting autonomous behavior using an inverse trust estimation. In *International Conference on Computational Science and Its Applications*, pages 728–742. Springer, 2014.

[27] Charles W. Fox, Fanta Camara, Gustav Markkula, Richard Romano, Ruth Madigan, and Natasha Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *VEHITS 2018: 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, January 2018.

[28] Marc Green. ”how long does it take to stop?” methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3):195–216, 2000.

[29] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, Dennis Saluäär, and Pontus Larsson. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology*, 9:1336, 2018.

[30] Edward Twitchell Hall. *The hidden dimension*, volume 609. Garden City, NY: Doubleday, 1966.

[31] Leslie A Hayduk. The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 13(1):87, 1981.

[32] Heiko Hecht, Robin Welsch, Jana Viehoff, and Matthew R Longo. The shape of personal space. *Acta Psychologica*, 193:113–122, 2019.

[33] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. Designing social greetings in human robot interaction. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 855–864, 2014.

[34] Gert Heinrich and Manfred Klüppel. Rubber friction, tread deformation and tire traction. *Wear*, 265(7-8):1052–1060, 2008.

[35] Zachary Henkel, Cindy L Bethel, Robin Roberson Murphy, and Vasant Srinivasan. Evaluation of proxemic scaling functions for social robotics. *IEEE Transactions on Human-Machine Systems*, 44(3):374–385, 2014.

[36] Adam Henschke. Trust and resilient autonomous driving systems. *Ethics and Information Technology*, pages 1–12, 2019.

[37] Stephane Hess, Michel Bierlaire, and John W Polak. Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A: Policy and Practice*, 39(2-3):221–236, 2005.

[38] Suresh Kumaar Jayaraman, Chandler Creech, Dawn M. Tilbury, X. Jessie Yang, Anuj K. Pradhan, Katherine M. Tsui, and Lionel P. Robert. Pedestrian trust in automated vehicles: Role of traffic signal and av driving behavior. *Frontiers in Robotics and AI*, 6:117, 2019.

[39] Michiel Joosse, Aziez Sardar, Manja Lohse, and Vanessa Evers. Behave-ii: The revised set of measures to assess users' attitudinal and behavioral responses to a social robot. *International journal of social robotics*, 5(3):379–388, 2013.

[40] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.

[41] Richard L Knoblauch, Martin T Pietrucha, and Marsha Nitzburg. Field studies of pedestrian walking speed and start-up time. *Transportation research record*, 1538(1):27–38, 1996.

[42] Kheng Lee Koay, Kerstin Dautenhahn, SN Woods, and Michael L Walters. Empirical results from using a comfort level device in human-robot interaction studies. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 194–201, 2006.

[43] Julian Francisco Pieter Kooij, Nicolas Schneider, Fabian Flohr, and Dariu M. Gavrila. Context-based pedestrian path prediction. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 618–633, Cham, 2014. Springer International Publishing.

[44] Ioannis Kostavelis, Manolis Vasileiadis, Evangelos Skartados, Andreas Kargakos, Dimitrios Giakoumis, Christos-Savvas Bouganis, and Dimitrios Tzovaras. Understanding of human behavior with a robotic agent through daily activity analysis. *International Journal of Social Robotics*, 11(3):437–462, 2019.

[45] David Lambert. *Body language*. HarperCollins, 2004.

[46] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

[47] Yee Mun Lee, Ruth Madigan, Oscar Giles, Laura Garach-Morcillo, Gustav Markkula, Charles Fox, Fanta Camara, Markus Rothmueller, Signe Alexandra Vendelbo-Larsen, Pernille Holm Rasmussen, et al. Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology & Work*, 2020.

[48] Michael Lewis, Katia Sycara, and Phillip Walker. The role of trust in human-robot interaction. In *Foundations of trusted autonomy*, pages 135–159. Springer, Cham, 2018.

[49] Yuhao Lu, Tom Stafford, and Charles Fox. Maximum saliency bias in binocular fusion. *Connection Science*, 28(3):258–269, 2016.

[50] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlönn, Rikard Fredriksson, Claes Edgren, Stas

Krupenia, and Dennis Saluäär. Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context? In Neville A Stanton, Steven Landry, Giuseppe Di Bucchianico, and Andrea Vallicelli, editors, *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation, July 27-31, 2016, Walt Disney World®, Florida, USA*, pages 485–497. Springer International Publishing, Cham, 2017.

[51] Daniel Lyubenov. Research of the stopping distance for different road conditions. *Transport Problems*, 6:119–126, 2011.

[52] Ruth Madigan, Sina Nordhoff, Charles Fox, Roja Ezzati Amini, Tyron Louw, Marc Wilbrink, Anna Schieben, and Natasha Merat. Understanding interactions between automated road transport systems and other road users: A video analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66:196 – 213, 2019.

[53] Gustav Markkula, Richard Romano, Ruth Madigan, Charles W Fox, Oscar T Giles, and Natasha Merat. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. *Transportation Research Record*, 2018.

[54] Milecia Matthews, Girish Chowdhary, and Emily Kieson. Intent communication between autonomous vehicles and pedestrians. *CoRR*, abs/1708.07123, 2017.

[55] Ross Mead, Amin Atrash, and Maja J Matarić. Proxemic feature recognition for interactive robots: automating metrics from the social sciences. In *International conference on social robotics*, pages 52–61. Springer, 2011.

[56] Ross Mead, Amin Atrash, and Maja J Matarić. Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction. *International Journal of Social Robotics*, 5(3):367–378, 2013.

[57] Ross Mead and Maja J Mataric. Robots have needs too: People adapt their proxemic preferences to improve autonomous robot recognition of human social signals. *New Frontiers in Human-Robot Interaction*, 100:100–107, 2015.

[58] Ross Mead and Maja J Matarić. Perceptual models of human-robot proxemics. In *Experimental Robotics*, pages 261–276. Springer, 2016.

[59] Albert Mehrabian. *Nonverbal communication.* Transaction Publishers, 1972.

[60] Jeannette Montufar, Jorge Arango, Michelle Porter, and Satoru Nakagawa. Pedestrians' normal walking speed and speed when crossing a street. *Transportation Research Record*, 2002(1):90–97, 2007.

[61] Desmond F Moore. Friction and wear in rubbers and tyres. *Wear*, 61(2):273–282, 1980.

[62] Bonnie M Muir and Neville Moray. Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3):429–460, 1996.

[63] Pakpoom Patompak, Sungmoon Jeong, Itthisek Nilkhamhang, and Nak Young Chong. Learning proxemics for personalized human–robot social interaction. *International Journal of Social Robotics*, pages 1–14, 2019.

[64] Samantha Reig, Selena Norman, Cecilia G Morales, Samadrita Das, Aaron Steinfeld, and Jodi Forlizzi. A field study of pedestrians and autonomous vehicles. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pages 198–209, 2018.

[65] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

[66] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. In *Proc. of Driving Assessment Conference*, pages 186–192, 2017.

[67] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *Proc. of IEEE RO-MAN*, pages 795–802, 2016.

[68] Khaled Saleh, Mohammed Hossny, and Saeid Nahavandi. Towards trusted autonomous vehicles from vulnerable road users perspective. In *Proc. of Annual IEEE International Systems Conference (SysCon)*, pages 1–7, 2017.

[69] Shane Saunderson and Goldie Nejat. How robots influence humans: A survey of non-verbal communication in social human–robot interaction. *International Journal of Social Robotics*, 11(4):575–608, 2019.

[70] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[71] Michael Smithson. Trusted autonomy under uncertainty. In *Foundations of trusted autonomy*, pages 185–201. Springer, Cham, 2018.

[72] Arthur E Stamps III. Distance mitigates perceived threat. *Perceptual and motor skills*, 113(3):751–763, 2011.

[73] Arthur E Stamps III. Mitigation of threat by posture, distance, and proximity. *Comprehensive Psychology*, 2:27–50, 2013.

[74] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[75] Donna E Thompson, John R Aiello, and Yakov M Epstein. Interpersonal distance preferences. *Journal of Nonverbal Behavior*, 4(2):113–118, 1979.

[76] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[77] Elena Torta, Raymond H Cuijpers, and James F Juola. Design of a parametric model of personal space for robotic social navigation. *International Journal of Social Robotics*, 5(3):357–365, 2013.

[78] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803, Oct 2010.

[79] UK Law Commission. Automated Vehicles: a joint preliminary consultation paper, 2019. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/11/6.5066_LC_AV-Consultation-Paper-5-November_061118_WEB-1.pdf.

[80] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, and Pim Haselager. Do robot performance and behavioral style affect human trust? *International journal of social robotics*, 6(4):519–531, 2014.

[81] Michael L Walters, Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Dag Sverre Syrdal, and Chrystopher L Nehaniv. An empirical framework for human-robot proxemics. *Procs of new frontiers in human-robot interaction*, 2009.

[82] Mark Wardman. The value of travel time: a review of british evidence. *Journal of Transport Economics and Policy*, pages 285–316, 1998.

[83] Samantha F Warta, Olivia B Newton, Jihye Song, Andrew Best, and Stephen M Fiore. Effects of social cues on social signals in human-robot interaction during a hallway navigation task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1128–1132. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[84] Jingyuan Wu, Johannes Ruenz, and Matthias Althoff. Probabilistic map-based pedestrian motion prediction taking traffic participants into consideration. In *IEEE Intelligent Vehicles Symposium (IV) June 26-30, 2018, Changshu, Suzhou, China*, 2018.

Chapter 6

Extending Quantitative Proxemics and Trust to HRI

Abstract

Human-robot interaction (HRI) requires quantitative models of proxemics and trust for robots to use in negotiating with people for space. Hall's theory of proxemics has been used for decades to describe social interaction distances but has lacked detailed quantitative models and generative explanations to apply to these cases. In the limited case of autonomous vehicle interactions with pedestrians crossing a road, a recent model has explained the quantitative sizes of Hall's distances to 4% error and their links to the concept of trust in human interactions. The present study extends this model by generalising several of its assumptions to cover further cases including human-human and human-robot interactions. It tightens the explanations of Hall zones from 4% to 1% error and fits several more recent empirical HRI results. This may help to further unify these disparate fields and quantify them to a level which enables real-world operational HRI applications.

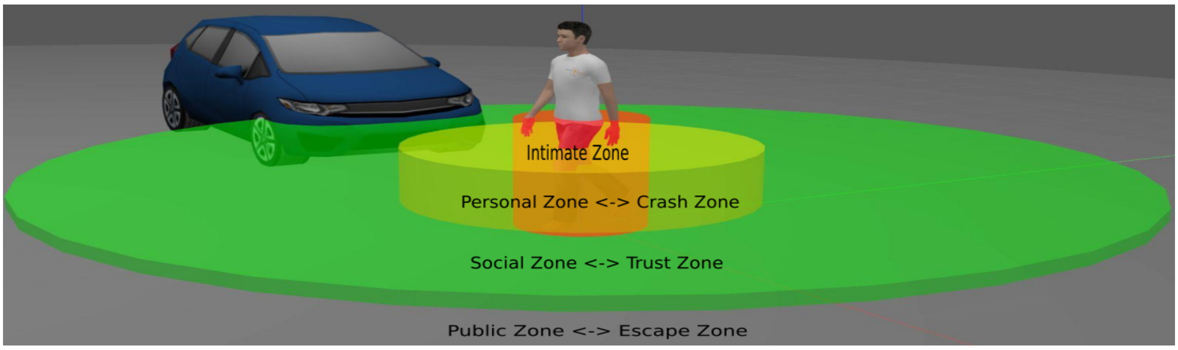


Figure 6.1: Autonomous vehicle entering pedestrian's social zone, which can also be viewed and quantified as a trust region.

1 Introduction

Autonomous robotics including autonomous vehicles (AVs) and service robots are now a reality, spreading from research to real-world social environments around humans [40]. Such environments raise new questions about how humans can trust robots, and how they should share their physical social spaces during human-robot interactions (HRI).

Social interaction is an important factor in making humans and robots acceptable and trustworthy to the humans they assist [4], and has been identified as one of ten major robotics challenges [43]. Two major challenges within Social Robotics were defined as modelling social dynamics, and learning social and moral norms [43]. Robots may be more accepted by people if they are socially aware, i.e. able to understand and reproduce these social norms and conventions. Within these norms, trust is essential for building relationships [26, 35]. Two important factors which influence the acceptance of humans and robots and are used to assess their social abilities are proxemics (i.e. interpersonal distances) and trust.

Robots need a better understanding and models of human social behaviour, especially nonverbal communication which plays an important role in human interactions. For instance, it was shown that people have strong 'social expectations' towards robots' nonverbal cues [5]. This raises concerns: are robots' social abilities good enough to interact with humans? Are they safe? Can we trust them?

Most current models of human social behaviour are based on qualitative studies and descriptive statistics. These are appropriate for reporting scientific findings, but they cannot be easily operationalised into engineered, robotic decision-making algorithms. More quantitative and computational models are thus needed to better understand and prescribe human-robot interactions, because numerical probabilities and utilities are needed by most robotics control systems.

The present paper briefly reviews proxemics, trust, and the recent PTR model [7] which links them quantitatively in the limited case of pedestrian-autonomous vehicle interactions. It then extends the PTR model to new, generalised cases of human-human and human-robot interactions and presents new results comparing the extended model's predictions to empirical data. These links could enable research to be shared and operationalised between models of proxemics, trust, and robotic interactions for the first time.

2 Review of Previous Work

2.1 Review of Proxemics and Trust

This section presents a review of previous work on proxemics and trust for HRI, an extensive review of these topics was introduced in [7].

Proxemics was proposed in the 1960's by Hall [14], defining four distinct zones for human interactions: the intimate, personal, social and public zones. Psychology studies then measured these zones for human-human interactions, finding that the intimate zone goes up to 0.45m, the personal ranges from 0.45m to 1.2m, the social between 1.2m to 3.6m, and the public beyond 3.6m [21]. These numbers are sometimes inserted into costmaps for robotic interaction planning algorithms. But we have lacked a theory to generate and explain these empirical values. Social roboticists have found these proxemic zones change in size when humans interact with robots of different heights, appearances, speeds, voices, and also for different HRI activities [30]. For example, for a short, 1.35m height, humanoid robot approaching or being approached by a human, the personal zone shrinks to the range 0.4m to 0.6m [36].

Trust is commonly defined as 'trusting a person means believing that when offered the chance, he or she is not likely to behave in a way that is damaging to us' [11, 3]. A question is whether humans can build trust with robots as they do with other people and through which means. For instance, a set of questionnaire metrics was designed in [42] to assess users' acceptance and use of robots via five HRI attributes, such as team configuration, team process, context, task, and system, where trust in automation is defined as depending on the level of autonomy of a system and also on its level of intelligence. Thus most HRI trust experiments have studied only humans' qualitative acceptance of robots [12, 25, 34]. But these qualitative models do not provide enough information to directly implement them as quantitative control systems for robotics.

2.2 The PTR Model: Linking Proxemics and Trust

Links between proxemics and trust have been proposed via a quantitative model, intended for use in the limited case of an autonomous vehicle, *$Agent_2$, interacting with a pedestrian,* Agent_1[1], crossing its path [7] as in Fig. 1. In this model, *Physical Trust Requirement (PTR)*

[1] *Terminology*: In the original model, Agent_1 was called 'the pedestrian' and Agent_2 called 'the vehicle'. The terms Agent_1 and Agent_2 are used throughout the present study to emphasise new generalities.

is defined as a Boolean property of the physical state of the world (not of the psychology of the agents) with respect to Agent$_1$ during an interaction, true if and only if Agent$_1$'s future utility is affected by an immediate decision made by Agent$_2$.

The model assumes that the two agents are approaching each other at a right angle, as is the case where one crosses the other's path, as in Fig. 6.1. It then defines the following three zones based on the PTR:

Crash zone is the region close to Agent$_1$, $\{d : 0 < d < d_{crash}\}$,

$$d_{crash} = v_2 t_2 + \frac{v_2^2}{2\mu_2 g}, \tag{6.1}$$

in which a crash is guaranteed and neither party can prevent it. v_2 is Agent$_2$'s speed. The first term depends on Agent$_2$'s thinking reaction time, t_2, and the second term represents the physical braking distance, μ_2 is the coefficient of friction between Agent$_2$'s tyres and tarmac, and g is gravity [23].

Escape zone is the area where Agent$_1$ is able to choose their own action to avoid the collision, without needing to trust Agent$_2$ to behave in any particular way. If w_2 is the width of Agent$_2$, which Agent$_1$ must cross at speed v_1 if they wish to pass first, the escape zone is then $\{d : d_{escape} < d\}$ with

$$d_{escape} = v_2 t_1 + w_2 \frac{v_2}{v_1}. \tag{6.2}$$

Trust zone is the region $\{d : d_{crash} < d < d_{escape}\}$ where the PTR is true. Agent$_2$ can here *choose* to slow down to prevent collision, but Agent$_1$ is incapable of making any action to affect this outcome themselves. This occurs when Agent$_1$ cannot get out of Agent$_2$'s way in time to avoid collision, but Agent$_2$ is able to slow and yield to prevent the collision if it chooses to do so.

The *zone ratio* $R = d_{escape}/d_{crash}$ is a measure of how much trust (in the PTR sense) is involved in an interaction.

Zones are not symmetric between Agent$_1$ and Agent$_2$. They describe when Agent$_1$ must trust Agent$_2$. Their roles must be swapped and the zones recomputed to see when Agent$_2$ must trust Agent$_1$. The crash, escape, and trust zones were mapped to Hall's personal, public, and social zones respectively, for Agent$_1$ [7], cf. Fig 6.1. The trust/social zone is the region in which physical trust is required. This may be a prerequisite for some types of interactions, with physical trust being useful to enable the content of the interaction. The evidence for this mapping came from the observation that if an autonomous vehicle Agent$_2$ is set to drive at the same speed as a pedestrian Agent_1, the model generates Hall's proxemic social zone to within 4% quantitative accuracy. This unexpected result, found only by studying how an AV should interact with road-crossing pedestrians, is suggestive that this scenario may be a special case of a more general HRI theory of proxemics and trust.

2.3 Limitations of the PTR Model

The PTR model made three key assumptions which limit its application to general HRI:

Assumption 1: Agent_2 is a wheeled vehicle, having momentum and a braking time. These dynamics are not appropriate for other types of Agent_2 such as walking humans and humanoids.

Assumption 2: The width of Agent_2 is much larger than that of Agent_1, so it treated Agent_1 as a point and Agent_2 as a rectangle, because a vehicle is bigger than a pedestrian and most vehicles are rectangular. These geometric assumptions are not appropriate for two human-like agents of similar size.

Assumption 3: The pedestrian has a *goal*: to cross the road. The road crossing is orthogonal to the road. Thus the pedestrian's velocity is orthogonal to the vehicle's. This is a strong constraint which is not appropriate to general HRI scenarios. Agent_2 might in general approach Agent_1 from any direction, not just at right angles to Agent_1's initial heading. We are now only interested in explaining the size of the trust zone which we assume is independent of any goal for Agent_1 other than avoiding a collision.

3 New Extensions to the PTR Model

The original PTR model was intended only for pedestrian road-crossing interactions with vehicles. We here expand its relevance to explain and predict new types of agents and scenarios, including human-human and human-humanoid robot interactions with approaches from arbitrary rather than orthogonal directions. We extend and generalize the model to address each of the above assumptions as follows.

Assumption 1: The second term on the right of Eq. (6.1) is only applicable to wheeled vehicles as it models their braking time. If Agent_2 is a walking agent, we will now assume this second term is omitted, as walkers are always in static equilibrium so can stop instantly once a decision is made. Models for running agents [20] or finer detailed models of walkers [28] could insert different braking terms here.

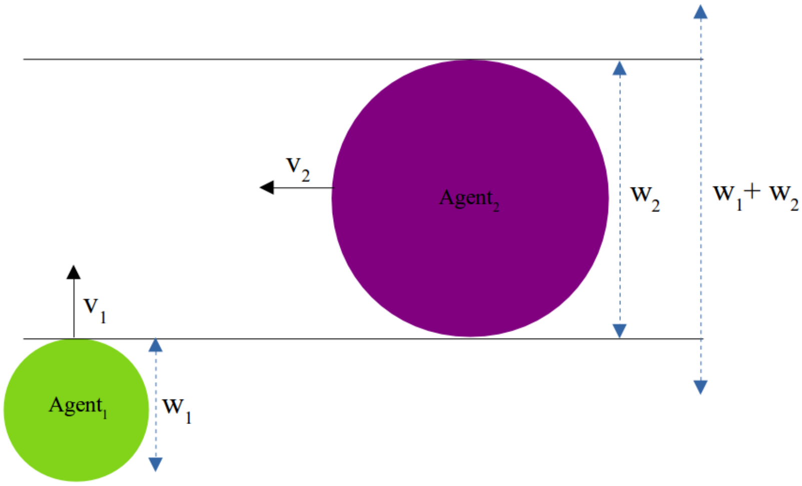


Figure 6.2: New assumed geometry for the two agents.

Assumption 2: To allow for interactions between similarly sized agents, we now modify Eq. (6.2) to:

$$d_{escape} = v_2 t_1 + (w_1 + w_2)\,\frac{v_2}{v_1}, \tag{6.3}$$

where $w_1 + w_2$ is the total distance that Agent$_1$ must travel in front of Agent$_2$ in order to avoid contact with Agent$_2$. These widths may now be mapped to Hall's intimate/personal zones of the agents, i.e. the ranges at which *actual* contact may occur between the agents. Justification for this modification can be seen in Fig. 6.2, which shows how Agent$_1$ must move its center point by half its own width at the start and end of the path as well as passing by the width of Agent$_2$, to avoid the *minimal* possible collision.

Assumption 3: As our focus is now purely on understanding proxemic and trust zones, we now drop the assumption that Agent$_1$ has a goal location, and consider that they simply want to avoid being hit by Agent$_2$. We thus want to allow Agent$_2$ to approach Agent$_1$ from any heading θ, measured relative to Agent$_1$'s own initial heading as in Fig. 6.3. The previous change from rectangular to circular agents is a first step towards enabling this. We then need to consider the direction in which Agent$_1$ moves to escape from Agent$_2$. The best way to escape is *always* by moving orthogonal to the heading of Agent$_2$[1]. There are at least four different modelling options for whether and how this is possible:

[1]Moving towards Agent$_2$ is obviously useless. Moving away from Agent$_2$ is useless if $v_1 < v_2$, but if $v_1 > v_2$ then there are no zones at all as it is trivial to escape. Any other direction is a linear combination of an optimal orthogonal escape plus one of these useless directions.

- Option 1: Assume that Agent_1 can turn on the spot *instantly* to face any direction. In this case, the optimal strategy is to first turn to a heading orthogonal to that of Agent_2, then walk at speed v_1 to escape. This makes Agent_1's initial heading irrelevant and reduces the model back to the original assumption of orthogonal velocities.

- Option 2: Assume that Agent_1 can *only* walk in the direction of their initial heading. They cannot rotate at all. By substituting v_1 in Eq. (6.3) for its component orthogonal to Agent_2's heading, $v_1|\sin(\theta)|$,

$$d_{escape} = v_2 t_1 + (w_1 + w_2)\frac{v_2}{v_1|\sin(\theta)|} \tag{6.4}$$

- Option 3: Assume Agent_1 can turn on the spot or twist during forward travel, where turning takes place at up to maximum angular velocity $\dot{\theta}$. If $\dot{\theta}$ is very fast then it will behave like Option 1. If $\dot{\theta}$ is very slow then it will behave like Option 2. Options 1 and 2 are thus special, limiting cases of Option 3.

- Option 4. Extending Option 3, further available motions such as sidesteps and stepping backwards could be added and optimised.

4 Results

The present section shows some validations of the extended model by comparing its predictions to data from a selection of previously published empirical studies of interest. Option 2 is chosen to model the direction of Agent_1. This is because it includes some consideration of the initial heading, unlike option 1, but without requiring a full solution of option 3 or 4 which may form extensive future work.

4.1 Two Walking Humans

We first show that the extended model can numerically reproduce and explain Hall's original observations of proxemic zone sizes for interactions between two walking humans (unlike the previous study's [7] with a walking human and autonomous vehicle). By choosing the realistic parameters: $t_1 = t_2 = 1.1$s, $v_1 = 1.1$m/s, $w_1 = w_2 = 1.19$m, found by optimisation, the extended model then generates values $d_{crash} = 1.21$m and $d_{escape} = 3.59$m, matching

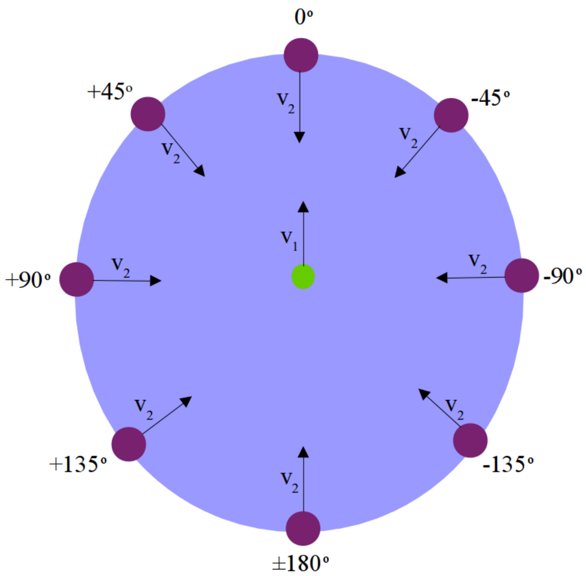
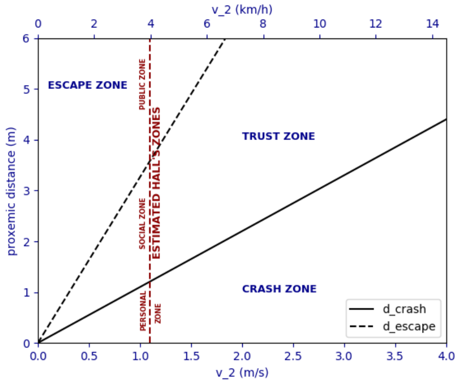


Figure 6.3: Possible interaction geometries. Green=Agent$_1$; Purple= different possible positions and headings for Agent$_2$. (θ is the angle of Agent$_2$'s approach from Agent$_1$'s perspective.)

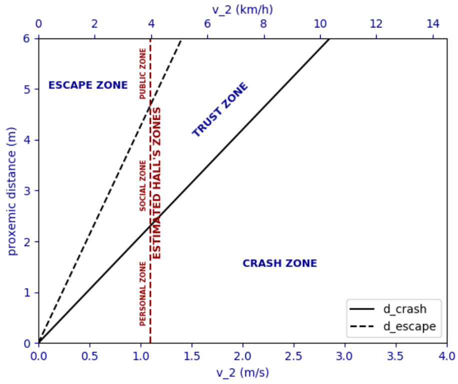
Hall's data, as shown by the vertical line in Fig. 6.4a, where $v_1 = v_2 = 1.1$m/s [21]. This result from the extended model shows a better fit to Hall zones, with an error of less than 1% compared to the previous model's 4% error [7]. The zone ratio is found to be $R_{H-H} = 3$ [7]. This will serve as a comparator for the following experiments.

4.2 Distracted Walking Human Interactions

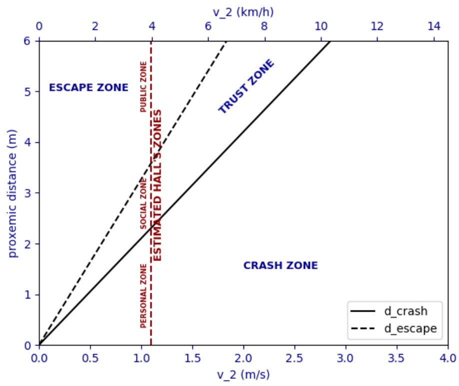
We next model the effect of distraction on the walking humans – such as attending to headphones, phones, or billboards – by increasing their reaction times in the model by 1s [10]. With both distracted, Fig. 6.4b shows that their crash zone size then increases from 1.21m to 2.31m and the escape zone is also increased from 3.59m to 4.69m, therefore the zone ratio R_{H-H} reduces to 2.03. With only one distracted, Fig. 6.4c shows that the crash zone size increases from 1.21m to 2.31m but the escape zone starts at 3.59m as in Sect. 4.1, leading to a smaller trust zone size in this case. The zone ratio, $R_{H-H} \approx 1.55$, is much smaller than the comparator. These findings are consistent with and explain empirical data that there is more distance, less trust, and hence less social interactions between distracted people [37].



(a) No distraction: $t_1 = t_2 = 1.1$s.

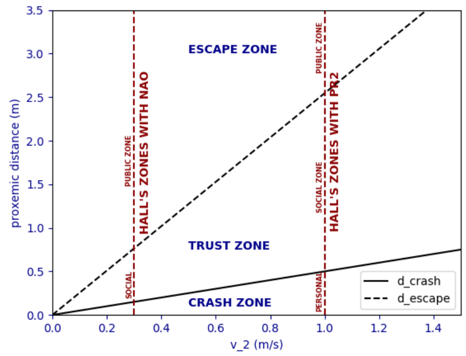


(b) Both distracted: $t_1 = t_2 = 2.1$s.

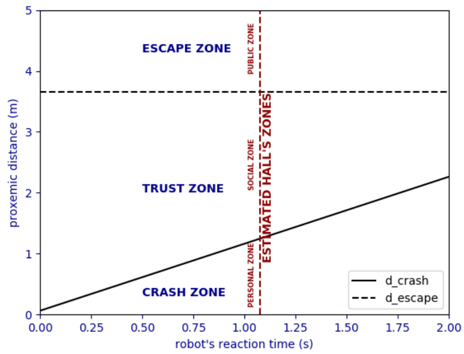


(c) One distracted: $t_1 = 1.1$s and $t_2 = 2.1$s.

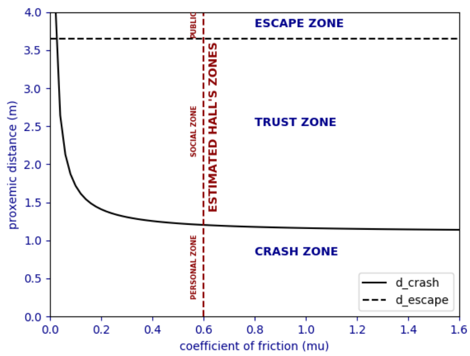
Figure 6.4: PTR distance and zone predictions for two walking humans at normal speed with different reaction times.



(a) Human - Humanoid interaction.



(b) Reaction time.



(c) Coefficient of friction.

Figure 6.5: Human-robot interactions. (6.5a) shows the PTR distance and zone predictions for a walking human interacting with humanoid robots at different speeds. (6.5b) and (6.5c) show the implied parameters for an interacting robot.

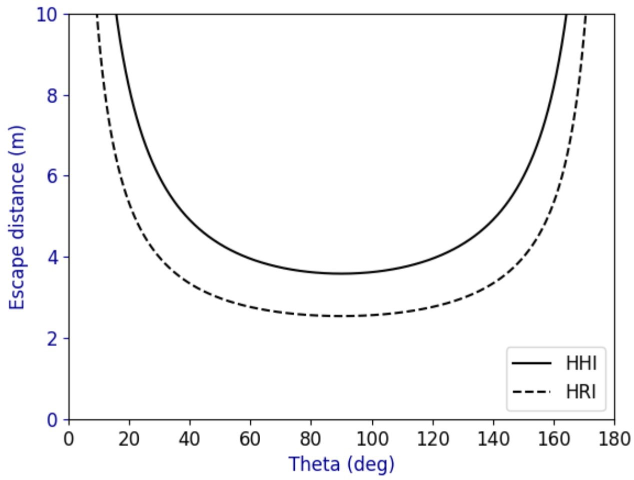


Figure 6.6: Example of predicted escape distance for different interaction angles between Agent_1 and Agent_2.

4.3 Walking Human vs Humanoid Robot

We now consider human-robot interactions. Fig. 6.5a shows predicted zone sizes for a human walker interacting with two different humanoids, NAO (~0.6m tall) and PR2 (~1.4m tall). The parameters used are: $t_1 = 1.1$s, $t_2 = 0.5$s, $v_1 = 1.1$m/s, $w_1 = 1.19$m, $w_2 = 0.4$m. With NAO at speed $v_2 = 0.3$m/s, the model predicts zone sizes: $d_{crash} = 0.15$m and $d_{escape} = 0.76$m. For PR2, having speed $v_2 = 1.0$m/s, zone sizes are: $d_{crash} = 0.5$m and $d_{escape} = 2.54$m. The sizes found for these human-robot interactions are much smaller than for human-human interactions, which is consistent with and matches closely results from previous empirical experiments with humanoid robots [18, 36, 39]. The zone ratios $R_{H-NAO} = 5.06$ and $R_{H-PR2} = 4.62$ are much bigger than the comparator from above. This explains existing empirical results that humans may be more sociable and friendly with humanoids than human strangers [16], and that people might not perceive robots as 'social entities' having an intimate zone [38].

4.4 Effects of Different Approach Headings

Fig. 6.6 shows the predicted escape distance for different approach headings between Agent_1 and Agent_2. In the HRI case, the prediction matches the previous result for a PR2 robot at 90° with $d_{escape} = 2.54$m, assuming the following parameters: $t_1 = 1.1$s, $t_2 = 0.5$s, $v_1 = 1.1$m/s,

$v_2 = 1\text{m/s}$, $w_1 = 1.19\text{m}$ and $w_2 = 0.4\text{m}$. In the HHI scenario, the parameters are as follows: $t_1 = t_2 = 1.1\text{s}$, $v_1 = v_2 = 1.1\text{m/s}$, $= 1\text{m/s}$ and $w_1 = w_2 = 1.19\text{m}$, and the prediction at $90°$ closely matches Hall's zone, with $d_{escape} = 3.59\text{m}$. The results of this extended model match and explain recent empirical data that d_{escape} i.e. public zone may be noncircular [15] while d_{crash} i.e. personal zone is always circular [29]. This is because d_{escape} is a function of v_1 (Eq. 6.3) and v_2 while d_{crash} depends only on v_2 (Eq. 6.1). The escape distance goes to infinity as $\theta \to 0°$ and $\theta \to 180°$ because it is impossible for Agent_1 to escape if their heading is constrained to be the same as Agent_2's.

4.5 Measuring Human Beliefs About Robots

It is possible to measure human's beliefs about robots' proxemic behaviour via implied parameters from the model and experimental data. For example by optimising the reaction time of the robot (Fig. 6.5b) or its coefficient of friction (Fig. 6.5c) to best fit results from human interaction. Assuming $v_1 = v_2 = 1.1\text{m/s}$, $t_1 = 1.5\text{s}$, $w_1 + w_2 = 2\text{m}$, $\mu = 1.0$ and $g = 9.8\text{m/s}^2$, as in Fig. 6.5b, the best reaction time for this case would be $t_2 \approx 1.075\text{s}$ if the robot wants to behave like a human and reproduce Hall's empirical zones. Similarly the coefficient of friction is found by keeping the previous parameters except μ which becomes unknown and by now setting $t_1 = t_2 = 1.1\text{s}$. Fig. 6.5c shows that the best coefficient of friction for the robot would then be $\mu = 0.6$. This should enable roboticists to learn and program their robots with the best parameters, with the possibility to vary the parameters for different people and in different environments. Current HRI proxemics results may suggest that humans have this natural ability to measure a robot's parameters and thus adapt their behaviour accordingly.

5 Discussion

The new extensions generalise the unification of proxemics and trust previously presented in the special case of AV-pedestrian interactions, to more general HRI interactions. This was achieved by modifying the assumptions to allow interactions between agents of similar sizes, approaching at arbitrary angles, and by removing the need for a goal location. The new model was validated by successfully fitting and explaining varied classical recent empirical proxemics and trust results.

We have here simulated two identical walking agents, but in the real world it is unlikely that two humans will share the same exact behavioural parameters. This new model could help to better understand proxemics and trust dynamics by simulating agents with differing parameters, without the costs or hazards associated with human experiments. The model for two walkers at normal walking speed is also valid for walkers at higher speeds e.g. 2.2m/s because the form of the equations scale, though for runners new dynamic equilibrium terms may be needed to model their stopping distance. In some cases, such as interactions with large cars, the old rectangular vehicle geometry may have to be restored and more complex equations used to compute shape overlaps and collisions. Future work should replace the use of Option 2 with a full solutions to Option 3 then 4.

Some possible applications for this work include:

Social Robotics: People are 'the big problem with self-driving cars' [6]. AVs are one case of social robots, which must understand social dynamics and norms, especially in crowded and mixed pedestrian-vehicle areas, in order to negotiate for space safely [33]. These negotiations are typically competitive rather than collaborative, with the aim of each agent being to get to their own destination quickly rather than to specifically interact with the other. Other forms of Social Robotics such as interactions with service and assistive robots may also benefit from quantitative understanding of proxemics and trust [17, 22, 24]. Unlike AVs, interaction with these robots is often cooperative.

Gaming & Extended Reality (XR) seeks to understand human proxemics in simulations of crowds, both for improving realism of video games and movie special effects, and for serious games such as simulations of evacuations, human locomotion in obstructed environments or group interactions in immersive virtual environments [1, 9, 31, 32].

Behavioural & Social Sciences: As trustors, humans are known to be more trusting (and gullible) depending on personality and environmental factors, and neuroscientific factors such as oxytocin hormones which may be physically transmitted through physical proximity [19]. As trustees, humans also maintain different reputations for trustworthiness, as studied by social network theorists [2, 41]. Hall zones are known to change in size across different human cultures [13]. Future work may need to take account of and replicate these factors for different human cultures. The Covid-19 pandemic has put a focus on human-human physical social interactions via the concept of *social distancing*. This is the encouragement or enforcement of a minimum proxemic distance between people when meeting. This requires hard numerical distance limits to be decided but there is a debate about what this distance should be. If the distance is too small, infections may be transmitted. A meta-review [8] found that 1m distance reduces transmission risk by 86%; 2m by 93%; and 3m by 96%. Others argue that if social distance is too large, trust will be harder to build [27]. An analogous debate to human-robot trust exists here, with arguments that physical proximity is sometimes needed to build human-human trust which may be jeopardized through social distancing and remote working. For example many workers are happy to hold technical meetings online but want to meet physically and closely to make contacts and deals which require trust.

Acknowledgements

This project has received funding from EU H2020 project interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395. The authors would like to thank Prof. Samantha Jamson, Prof. Simon Parsons, Dr. Paul Baxter, and Dr. Patrick Dickinson for their useful feedback on a first draft of this paper.

References

[1] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[2] Jay B Barney and Mark H Hansen. Trustworthiness as a source of competitive advantage. *Strategic management journal*, 15(S1):175–190, 1994.

[3] Chandrayee Basu and Mukesh Singhal. Trust dynamics in human autonomous vehicle interaction: a review of trust models. In *2016 AAAI spring symposium series*, 2016.

[4] Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. *Springer handbook of robotics*, pages 1935–1972, 2016.

[5] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*, pages 708–713. IEEE, 2005.

[6] Rodney Brooks. The big problem with self-driving cars is people. *IEEE Spectrum: Technology, Engineering, and Science News*, 2017.

[7] Fanta Camara and Charles Fox. Space invaders: Pedestrian proxemic utility functions and trust zones for autonomous vehicle interactions. *Int J of Soc Robotics*, 2020. https://doi.org/10.1007/s12369-020-00717-x.

[8] Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, Amena El-harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The Lancet*, 395(10242):1973–1987, 2020.

[9] Patrick Dickinson, Kathrin Gerling, Kieran Hicks, John Murray, John Shearer, and Jacob Greenwood. Virtual reality crowd simulation: effects of agent density on user experience and behaviour. *Virtual Reality*, 23(1):19–32, 2019.

[10] Thomas F Fugger, Bryan C Randles Jr, Anthony C Stein, William C Whiting, and Brian Gallagher. Analysis of pedestrian gait and perception-reaction at signal-controlled crosswalk intersections. *Transportation Research Record*, 1705(1):20–25, 2000.

[11] Diego Gambetta et al. Can we trust trust. *Trust: Making and breaking cooperative relations*, 13:213–237, 2000.

[12] Ilaria Gaudiello, Elisabetta Zibetti, Sébastien Lefort, Mohamed Chetouani, and Serena Ivaldi. Trust as indicator of robot functional and social acceptance. an experimental study on user conformation to icub answers. *Computers in Human Behavior*, 61:633–655, 2016.

[13] Edward T Hall. A system for the notation of proxemic behavior 1. *American anthropologist*, 65(5):1003–1026, 1963.

[14] Edward Twitchell Hall. *The hidden dimension*, volume 609. Garden City, NY: Doubleday, 1966.

[15] Leslie A Hayduk. The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 13(1):87, 1981.

[16] Takayuki Kanda, Takahiro Miyashita, Taku Osada, Yuji Haikawa, and Hiroshi Ishiguro. Analysis of humanoid appearances in human–robot interaction. *IEEE Transactions on Robotics*, 24(3):725–735, 2008.

[17] Philipp Kellmeyer, Oliver Mueller, Ronit Feingold-Polak, and Shelly Levy-Tzedek. Social robots in rehabilitation: A question of trust. *Sci. Robot*, 3(21), 2018.

[18] Kheng Lee Koay, Dag Sverre Syrdal, Michael L Walters, and Kerstin Dautenhahn. Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 564–569, 2007.

[19] Michael Kosfeld, Markus Heinrichs, Paul J Zak, Urs Fischbacher, and Ernst Fehr. Oxytocin increases trust in humans. *Nature*, 435(7042):673–676, 2005. https://doi.org/10.1038/nature03701.

[20] Taesoo Kwon and Jessica K Hodgins. Control systems for human running using an inverted pendulum model and a reference motion capture sequence. In *Symposium on Computer Animation*, pages 129–138, 2010.

[21] David Lambert. *Body language*. HarperCollins, 2004.

[22] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience & Biobehavioral Reviews*, 104:231–239, 2019.

[23] Daniel Lyubenov. Research of the stopping distance for different road conditions. *Transport Problems*, 6:119–126, 2011.

[24] Ross Mead and Maja J Matarić. Autonomous human–robot proxemics: socially aware navigation based on interaction potential. *Autonomous Robots*, 41(5):1189–1201, 2017.

[25] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, pages 1–23, 2020.

[26] Cristina Olaverri-Monreal. Promoting trust in self-driving vehicles. *Nature Electronics*, 3(6):292–294, 2020. https://doi.org/10.1038/s41928-020-0434-8.

[27] Gary M Olson and Judith S Olson. Distance matters. *Human–computer interaction*, 15(2-3):139–178, 2000.

[28] Lalit Patnaik and Loganathan Umanand. Physical constraints, fundamental limits, and optimal locus of operating points for an inverted pendulum based actuated dynamic walker. *Bioinspiration & Biomimetics*, 10(6):064001, 2015.

[29] Pakpoom Patompak, Sungmoon Jeong, Itthisek Nilkhamhang, and Nak Young Chong. Learning proxemics for personalized human–robot social interaction. *International Journal of Social Robotics*, pages 1–14, 2019.

[30] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

[31] Daniel Roth, Constantin Klelnbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 215–222, 2018.

[32] Ferran Argelaguet Sanz, Anne-Hélène Olivier, Gerd Bruder, Julien Pettré, and Anatole Lécuyer. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *IEEE Virtual Reality (VR)*, pages 75–80, 2015.

[33] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50):24972–24978, 2019.

[34] Shervin Shahrdar, Luiza Menezes, and Mehrdad Nojoumian. A survey on trust in autonomous systems. In *Science and Information Conference*, pages 368–386. Springer, 2018.

[35] Keng Siau and Weiyu Wang. Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2):47–53, 2018.

[36] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[37] Leah L Thompson, Frederick P Rivara, Rajiv C Ayyagari, and Beth E Ebel. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury prevention*, 19(4):232–237, 2013.

[38] Michael L Walters, Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Christina Kaouri, Sarah Woods, Chrystopher Nehaniv, David Lee, and Iain Werry. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication.*, pages 347–352, 2005.

[39] Michael L Walters, Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Dag Sverre Syrdal, and Chrystopher L Nehaniv. An empirical framework for human-robot proxemics. *Procs of New Frontiers in Human-Robot Interaction*, 2009.

[40] Jochen Wirtz, Paul G Patterson, Werner H Kunz, Thorsten Gruber, Vinh Nhat Lu, Stefanie Paluch, and Antje Martins. Brave new world: service robots in the frontline. *Journal of Service Management*, 2018. https://doi.org/10.1108/JOSM-04-2018-0119.

[41] Sze-Sze Wong and Wai Fong Boh. Leveraging the ties of others to build a reputation for trustworthiness among peers. *Academy of Management Journal*, 53(1):129–148, 2010.

[42] Rosemarie E Yagoda and Douglas J Gillan. You want me to trust a robot? the development of a human–robot interaction trust scale. *International Journal of Social Robotics*, 4(3):235–248, 2012.

[43] Guang-Zhong Yang, Jim Bellingham, Pierre E Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, et al. The grand challenges of science robotics. *Science Robotics*, 3(14):eaar7650, 2018.

Chapter 7

OpenPodcar: an Open Source Vehicle for Self-Driving Car Research

Abstract

OpenPodcar is a low-cost, open source hardware and software, autonomous vehicle research platform based on an off-the-shelf, hard-canopy, mobility scooter donor vehicle. Hardware and software build instructions are provided to convert the donor vehicle into a low-cost and fully autonomous platform. The open platform consists of (a) hardware components: CAD designs, bill of materials, and build instructions; (b) Arduino, ROS and Gazebo control and simulation software files which provide standard ROS interfaces and simulation of the vehicle; and (c) higher-level ROS software implementations and configurations of standard robot autonomous planning and control, including the move_base interface with Timed-Elastic-Band planner which enacts commands to drive the vehicle from a current to a desired pose around obstacles. The vehicle is large enough to transport a human passenger or similar load at speeds up to 15km/h, for example for use as a last-mile autonomous taxi service or to transport delivery containers similarly around a city center. It is small and safe enough to be parked in a standard research lab and be used for realistic human-vehicle interaction studies. System build cost from new components is around USD7,000 in total in 2022. OpenPodcar thus provides a good balance between real world utility, safety, cost and research convenience.

Metadata Overview

Main design files: https://github.com/OpenPodcar/OpenPodcar
Target group: researchers and hobbyists interested in autonomous vehicle research and robotics.
Skills required: Mechanical assembly – intermediate (drilling steel); electrical assembly – intermediate (PCB soldering); Software – easy (Linux command line).
Replication: The current OpenPodcar is being used by some of the authors for human-robot interaction experiments and a second copy will be built from the documentation to improve its accuracy. The design is currently being forked for a courier-type manually-driven platform by a commercial UK vehicle manufacturer.

Keywords

Autonomous vehicle, automation, self-driving car, mobility scooter, open source platform.

---

1 Overview

1.1 Introduction

Autonomous Vehicles (AVs, also known as 'self-driving cars'), is a fast-moving research field in both academia and the industry. Open source software (OSS) for localisation, mapping and control of AVs is available [25] but hardware vehicle platforms remain expensive and proprietary, making it difficult for researchers with low resources to develop algorithms or reproduce complete research systems. There is thus a need for a standard, low-cost, reproducible hardware platform, compatible with the standard open source software stack.

Open source hardware (OSH) allows for more effective and accessible sharing and collaboration among researchers [16]. By combining OSH and OSS, a standard platform can be produced for use by all members of a research community, who may then reproduce each others work in full, and contribute their new research as functional system components rather than only as reports. Such platforms may evolve from research into development and real-world applications.

To create an OSH platform for the autonomous vehicle research community, several requirements must be met: *low cost* and *easy to build* to enable the community to reproduce and use it; consumer levels of safety and reliability are not required, though *research standards of safety and reliability* are required; the system should be designed to enable *easy modification* so that it can be forked to operate with similar but different vehicles; the system should be physically *light-weight* to ease experimentation and reduce risks of damage, though large enough for *human transport* so that it can be used in real-world applications and in research requiring realistic interactions with other human road users [10, 17].

1.2 Related Systems

SMART [36] is a design to modify an existing donor golf cart vehicle for automation research, this is of a similar size and power to OpenPodcar. Similarly, iCab (Intelligent Campus Automobile) [22] is a research golf car with a ROS (Robot Operating System)-based architecture and that has been tested with Timed-Elastic Band planner [28]. However, these vehicle designs are not open source hardware.

Complete and built mechanical OSH designs for on-road, person-carrying cars exist, including PixBot [37] and Tabby EVO [33]. Building these full size cars is a large task for experts and may require dangerous processes such as welding, purchase of expensive components, and considerable storage space. OpenPodcar is based on a proprietary but commodity mobility scooter which is cheaper and easier to convert than performing these builds.

Several OSH RC-scale cars have been completed and built such as F1Tenth [1], AutoRally [21], BARC [23], MIT Racecar [2], MuSHR [3], [31], and [46]. These platforms are not large enough to drive on public roads or to transport people or goods like OpenPodcar. Open Source Ecology (OSE) [24] is an ambitious programme of projects which ultimately aims to develop fully OSH vehicles including a car and tractor. OSE is optimised for reliability and for users in developing countries so it uses hydraulic power rather than electric as used in OpenPodcar. But its vehicle designs are not yet complete.

Autoware [25] is a heavyweight open source software project to construct a full ROS-based automation stack for on-road cars. Apollo [4] is an open source self-driving software stack and an open hardware interface which may be implemented on vehicles, as done in [26]. These systems could be software interfaced to run with OpenPodcar.

Some AV research can be done in simulation without the need for hardware, hence open source simulation platforms are widely available such as SUMMIT [5], Gym-Duckietown [15], CARLA [18], DEEPDRIVE [39], LGSVL Simulator [41], AirSim[43], and FLOW [47]. The USA state of Georgia provides a level 3 open-source autonomous vehicle based on a Ford-Edge[35], which can be used *gratis* in their Peachtree Corners' Curiosity Lab smart city environment by researchers needing a vehicle but not wanting to build or buy one.

2 Overall Implementation and Design

2.1 Donor vehicle

A Pihsiang TE-889XLSN hard-canopy scooter (branded in UK as Shoprider Traverso, [44]) is used as a donor vehicle. It is an Ackermann-steered [30], hard-canopy, electric mobility scooter. It is powered by two 12V batteries connected in series to provide 24V operating voltage and containing 75Ah. In its standard configuration, its steering is controlled by a human-operated loop handle bar. The speed and braking systems are both powered by an electric motor and an electric brake via the trans-axle assembly, controlled by an AC2 digital controller receiving different voltage signals to drive forward or brake. The manual speeding and braking systems are controlled by three buttons connected in series on the handle bar. A toggle switch in parallel with a resistor (10kΩ) selects speed mode from high (max 8mph) or low (max 4mph); a speed dial knob via a variable resistor (20kΩ) sets a maximum limit speed within the mode. A throttle lever connected with a 5kΩ potentiometer is used to select the speed within the mode and limit.

2.2 Mechanical Modification for Steering

To automate steering, a linear actuator (Gimson GLA750-P 12V DC) with position feedback is mounted between an anchor on the underside of the chassis and the car's front axle via bearings. This actuator has a 8mm/s full load (750N) speed and 250mm stroke length (installation length is 390mm). To access the underside of the vehicle, two axle stands are used as shown in Fig. 7.1a. There is an existing hole in the right front wheel axle. The linear actuator is mounted via a rear hole to the left side of the front chassis and connected through the front hole of the actuator with the hole in the car's right front wheel axle via bearings as shown in Figs. 7.1b and 7.2.

(a) Tilting the vehicle using two axle stands, to enable access to the underside. (Note also lidar mounted to roof.)

(b) Underside with linear actuator added for steering.

Figure 7.1: Vehicle mechanical modification

2.3 Electronics

The new vehicle electronics, which various different voltage power supplies (cf. Fig. 7.3), are packaged on a single new PCB (Printed Circuit Board), as shown in Figs. 7.5a and 7.5b. This is convenient as it reduces the number of small wires between the components by having them directly drawn on the board, and packages them together.

As an OSH design, the PCB hosts several daughter PCBs, mounted using headers. The physical structure of the large PCB comprised of these smaller PCBs reflects the OSH design itself. There are two DC-DC Buck converters with an XL4016 regulator, an Arduino Uno, an MCP4725 DAC (Digital-Analog Converter), a Pololu Jrk 21v3 motor controller with position feedback for the linear actuator, two resistors (10kΩ and 100kΩ) for the potential divider and two terminal blocks. 3D-printed parts support the mounting of the LCD and the 3D lidar to the board. A 3D printed enclosure mounts and protects the PCB board, as shown in Fig. 7.4.

2.4 Steering Automation System

The front wheels are steered by a Pololu Jrk 21v3 PID controller-driver, which takes serial port desired positions as input. It also takes feedback position information as an analog voltage from the linear actuator as an input. It outputs analog high-power voltages to the linear actuator. A gratis, closed-source, Windows program from Pololu is required once, at build time, to set the PID parameters for the linear actuator.

The relationship between the required central turning angle θ of the pair of front wheels

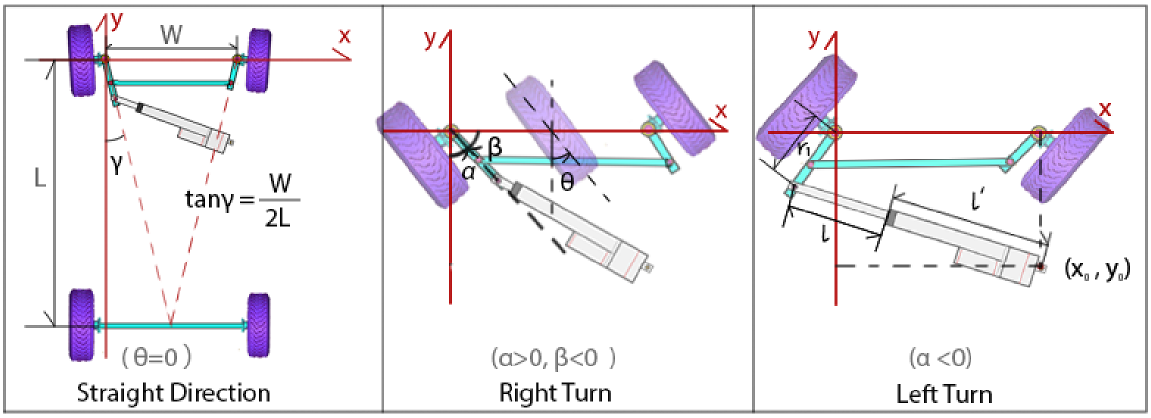


Figure 7.2: Underside view of front wheels' steering relationship including geometric coefficients

and extending length l of linear actuator as in Fig. 7.2 is given by,

$$\theta = \alpha - \arctan(\frac{W}{2H}) \tag{7.1}$$

$$\beta = \alpha - \frac{\pi}{2} \tag{7.2}$$

$$x = r_1 \cos(\beta) \tag{7.3}$$

$$y = r_1 \sin(\beta) \tag{7.4}$$

$$l = \sqrt{(x_0 - x)^2 + (y_0 - y)^2} - L + l_0 \tag{7.5}$$

where r_1, x_0, y_0, W, H and L are the geometric coefficients shown in Fig.1. Among them, the value of y_0 is negative. l_0 is the initial value of the linear actuator position feedback. Table 7.2 specifies the acceptable serial port commands for the linear actuator. Sending commands outside this range may mechanically destroy the system.

2.5 Speed Controller Automation System

An Arduino UNO [34] is used to send electric signals to the vehicle's motor controller in place of the donor vehicleâ€™s paddle controllerâ€™s potentiometer. An Adafruit MCP4725 DAC is connected to the Arduino as in Fig. 7.3, and is used to send clean analog speed command voltages to the donor vehicleâ€™s internal controller.

Arduino firmware source, and upload instructions, are supplied in the repository. When uploaded to the Arduino (using the standard Arduino IDE running on the laptop), the

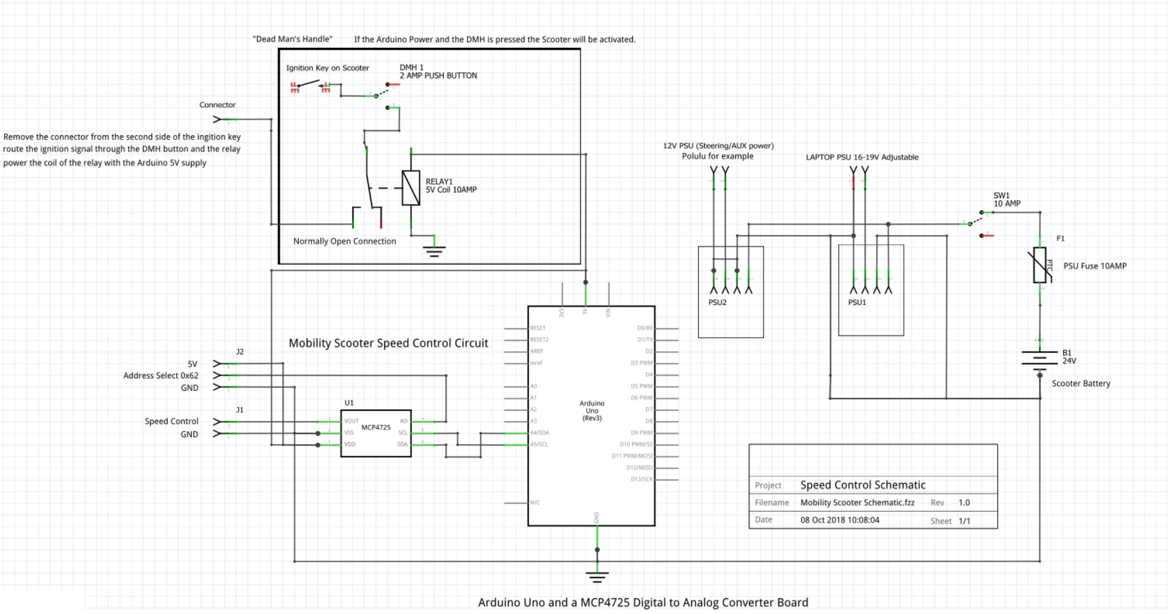


Figure 7.3: Circuit diagram for electronic modifications.

firmware provides a simple serial port API running at 112,000 baud. It receives ASCII commands of the form 'FA:210' as speed commands. Table 7.3 summarises the range of speed commands and their corresponding output voltages.

To start the ignition, the car safety system requires the control voltage to be in the dead range. A problem is that this doesnâ€™t correspond precisely to any fixed speed bytes, due to floating USB power level issues. But if we pick a number solidly in the center of the dead zone, such as 164, this will work for most USB supplies. (i.e. when the vehicleâ€™s battery is flatter, the voltages provided to USB power by it are lower. For example, we might send 164 and get 1.9V instead of the usual 2.26V.) This may result in the vehicle not starting and producing an audible warning beep instead.

Also due to floating voltages from the battery, the Arduino typically receives a lower power e.g. 4.9V instead of its ideal 5V, which gets divided by the DAC value in some calculations.

To deal with these instabilities, a potential divider is added to the battery to monitor its voltage and compensate the podcar control accordingly, as in Fig. 7.6. A "BV" command is provided in the Arduino serial protocol which allows callers to request this current battery voltage. This can then be used by higher-level (Python) systems to decide what speed bytes to sent, including compensating for the floating dead zone.

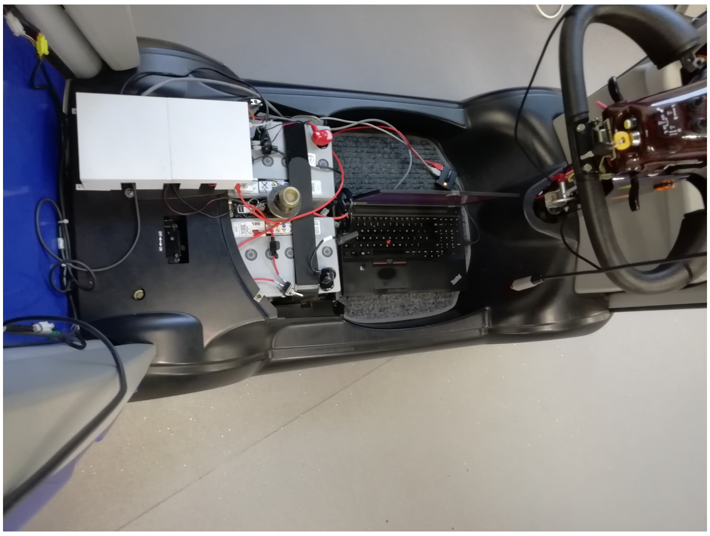
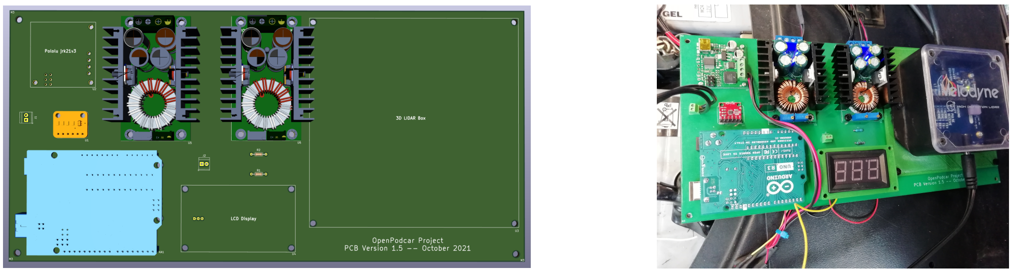


Figure 7.4: PCB enclosure mounted on the vehicle (top left, white box).



(a) PCB Design.

(b) PCB assembly currently used.

Figure 7.5: Electronics with PCB board design and assembly

Table 7.2: Linear actuator acceptable command ranges.

FA:cmd	Effect
2500	turn max right i.e $\sim$ -45 deg
1900	center wheels i.e $\sim$ 0 deg
1000	turn max left i.e $\sim$ +45 deg

Table 7.3: Speed commands and their corresponding output voltages.

Command	Voltage	Effect
FA:0	0	very fast reverse
FA:80	~ 0.9	fast reverse (ROS limit)
FA:132	~ 1.5	slowest reverse motion
		dead zone - allows ignition
FA:170	1.9	stop - zero/home position
FA:201	~ 2.3	slowest forward motion
FA:240	~ 2.7	fast forward (ROS limit)
FA:255	~ 3.0	very fast forward

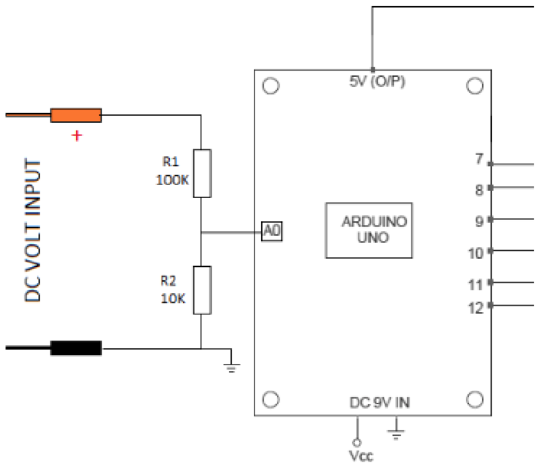


Figure 7.6: Potential divider linked to the battery

2.6 Software Interface (ROS)

A ROS interface to and from the physical vehicle is provided as described below. ROS is an open source operating system for robots based on a publish-subscribe pattern [38], which is the robotics communityâ€™s standard interface. The ROS core and software all run on a consumer laptop computer mounted on-board the vehicle, and that could be powered from a DCDC converter from the vehicle battery, running Xubuntu 16.04 (Xenial) and ROS Kinetic.

The system expects to hear two incoming ROS control messages: /speedcmd_meterssec and /wheelAngleCmd, which contain single floats representing the desired speed in meters per second, and the desired front wheel orientation in radians respectively. These two messages are received by ROS nodes speedm2arduino and wheelAngle2Pololu, which are ROS drivers for the Arduino speed controller and the Polulo steering controller respectively. Converters from a standard ROS USB joystick driver node to the speed and angle command interface messages are provided, by joystick2speedms and joystick2wheelAngle. These use the y axis of a joystick for speed and x for steering.

2.7 3D Lidar Sensor

A Velodyne VLP-16 lidar sensor is mounted on the vehicle roof using a Manfrotto Black Digi Table Tripod 709B. It is mounted at a 10 degree tilt downwards (to allow pedestrians to be most clearly seen in the 16 scan lines). The lidar has a ROS driver.

2.8 High-Level Automation Software

Fig. 7.7 shows an overview of the ROS components used in high level automation, including localisation and mapping, path planning and control, and pedestrian tracking as discussed in the following sections.

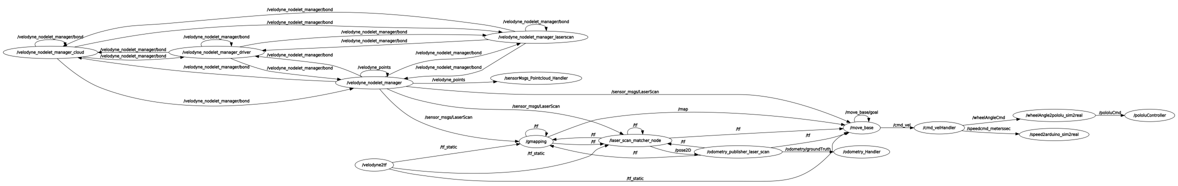
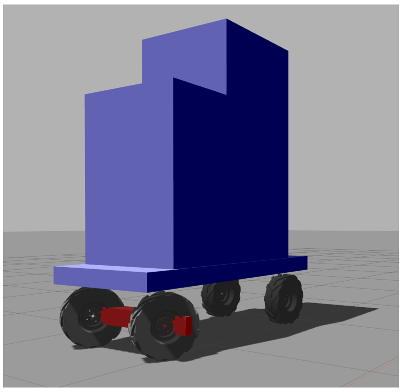


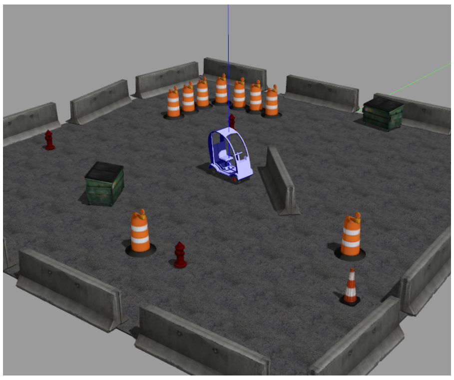
Figure 7.7: ROS nodes used in the autonomous driving mode.

2.8.1 Localisation and Mapping System

Simultaneous Localisation and Mapping (SLAM) [45] is the robotic task of inferring the robotâ€™s location at the same time as building a map of its environment, which is a classic ‘chicken and eggâ€™ problem as the two subtasks depend on one another. Solving SLAM is an NP-hard problem but many standard approximations exist. GMapping [49] is a ROS implementation of a Rao-Blackwellized Particle Filter (RBPF) in which “each particle carries an individual map of the environment”. The information carried by each particle overlaps, and an estimation of a map can be built based on these relationships. As the robot moves around the environment, these estimations are stored, and when a â€˜feedback loopâ€™ is closed, the estimations cascade into a portion of the completed map. These maps take the form of 2D occupancy grids, and can be used later by the navigation stack to plan paths around the environment. To provide reliable odometry data for GMapping, ROS laser_scan_matcher package is used as a stand-alone odometry estimator that matches consecutive laserscans.



(a) Physical simulation of vehicle.



(b) Default Gazebo simulation world

Figure 7.8: OpenPodcar 3D simulation

2.8.2 Path Planning and Control

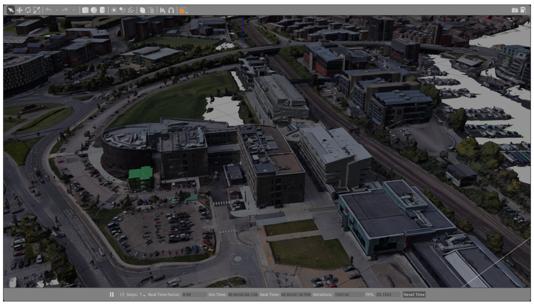
Path planning is the autonomous selection of an entire desired trajectory for a robot to get from a current pose to a desired pose. Path control (or path following) is then the real-time process of executing a path plan by interactively monitoring the robotâ€™s state and sending commands to motors, to make the actualized path close to the desired path. The OpenPodcar software includes path planning and control with the standard ROS tool move_base and Timed Elastic Band (TEB) [42] plugin. These tools implement the requirement geometry of Dubins paths [19] and Ackermann steering. The values for parameters such as minimum turning radius have been calculated from the technical specifications of the base vehicle [44].

2.8.3 Pedestrian Detection and Tracking

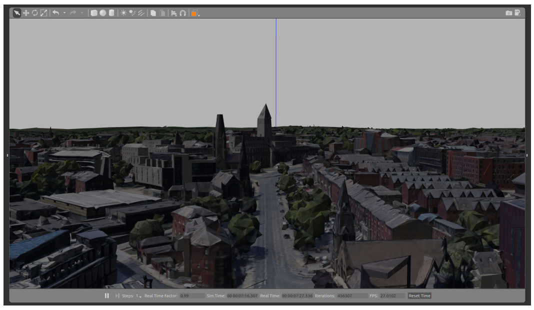
A pedestrian detector and tracker ROS package are included in the system. The lidar-based detections are classified by a SVM (Support Vector Machine) classifier, then a Bayesian multi-target tracker is used to track pedestrians over time. These modules re-use OSS from the EU FLOBOT project [48], merged into the repository.

2.9 Simulation

A physical simulation of the vehicle is provided for use in Gazebo 7 [27] under ROS Kinetic and Ubuntu 16.04 (Xenial). The simulation implements the same ROS interface as the physical vehicle system to enable plug and play inter-operability between them. The physics model



(a) University of Lincoln 3D world



(b) University of Leeds 3D world

Figure 7.9: OpenPodcar additional Gazebo 3D simulation worlds

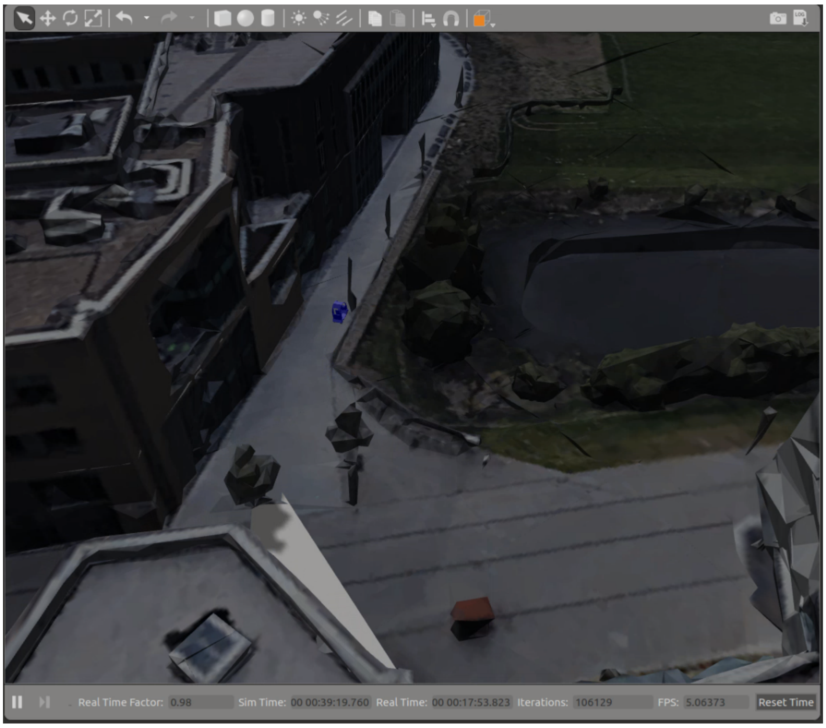


Figure 7.10: OpenPodcar in Lincoln world

is based on a simplified vehicle geometry with two large cuboids containing the vehiclesâ€™ mass, as shown in Fig. 7.8a. Wheel geometry, friction, and motor driver parameters were measured from the physical vehicle. A detailed graphical mesh model of the vehicle is provided for display, rather than physical simulation, purposes. The main difference with the real vehicle is that the effects of the linear actuator are represented by a tracking rod, where is mounted the Kinect sensor used in place of the lidar, as found in Fig. 7.8b.

A basic 3D world containing the podcar and various test objects from Gazebo libraries is provided by default as shown in Fig. 7.8b. Fig. 7.11 shows the complete ROS node configuration used during simulation, under manual joystick control. Moreover, the open source Blender 3D add-on, called MapsModelsImporter [29], was used to create further 3D worlds representative of the University of Lincoln, the testing area for the OpenPodcar, and the University of Leeds campuses, shown in Fig.7.9. Fig. 7.10 shows the OpenPodcar in Lincoln campus environment.

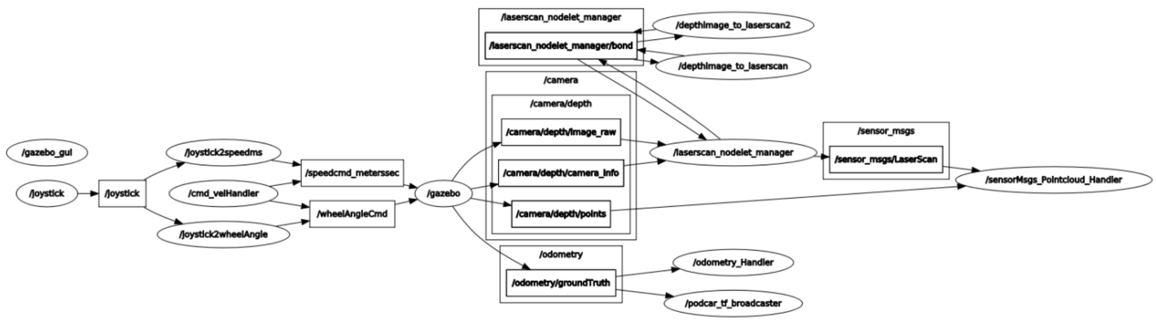


Figure 7.11: ROS nodes used in simulated, manual joystick control mode.

3 Quality Control

3.1 Safety

Autonomous vehicles can present a significant hazard to humans and to the environment in which they operate. Damage to surroundings and possible injury to operators and bystanders could result from inappropriate use or malfunction. A particular risk arises from the speed controller on the donor vehicle being of 'wigwag' style, as is common in mobility scooters. This means it is an analog signal in the range 0-5V, including a dead zone around 2.5V corresponding to no motion. Above the dead zone and up to 5V are forward speed control

Figure 7.12: Steering console showing the newly added relay (with lit LED)

commands of increasing speeds, below the dead zone to 0V are reverse control commands of increasing speeds. Wigwag control is potentially dangerous because a 0V signal might appear due to component failure rather than as a desired max-speed reverse command. Also, if the vehicle batteries run low, the scaling of this signal may be altered resulting in the dead zone position floating and leading to further undesired motion. The following layered safey systems are included to fully mitigate these risks:

Fusing As shown in Fig. 7.3, a 10A fuse is inserted between the vehicle's original 24V battery and the switch to the new electronics. This is in addition to original fusing and other safety features provided by the donor vehicle, which all remain in tact.

Dead Man's Handle It is essential that a suitable emergency stop system is implemented in all autonomous vehicles. Given the research nature of the OpenPodcar, a safety mechanism which stops the vehicle under fault conditions is an especially important part of the design. A wired dead-man's handle (DMH) is included which is required to be pressed by a human experimenter at all times, in order for a hardware relay to actively continue to supply power from the vehicleâ€™s batteries to all other systems. The relay connects to the donor vehicle's keyed ignition switch and will naturally cut out if these signals are absent for any reason, including failures in the safety systems themselves. A photograph of the installed system is shown in Fig. 7.12.

Heartbeat Signal The serial protocol linking the Arduino to ROS includes a heartbeat signal, in which the Arduino code will shut down the motors unless a correctly formatted and timestamped serial command is received within 0.1 seconds. This requires ROS code to actively check and confirm its own status and to send positive confirmation, for example if ROS or Linux go down then this heartbeat will cut off.

Steering Control System Limiter Limitations are placed on the steering controller for the linear actuator commands, to only allow the vehicle to accept and execute input values within the range that will keep the mechanical mounting safe.

3.2 General Testing

A series of sub-component (e.g. Pololu, DAC) acceptance tests, component (e.g. PCB, lidar) hardware unit tests, and system integration tests are defined and included as formal, non-optional steps in the build instructions. The structure of the tests is designed to enable build problems to be immediately localised, so that passing one test means that a failure of the next one must be due to build steps that have occurred between them. Below is a summary of these tests.

At component level, an external power supply, a multi-meter, a clamp meter and a bread-board with some wires are frequently used to recreate smaller electronic circuits in order to check the voltages, currents and the correct functioning of each component during the build. For instance, a circuit with an external power supplying 5V to the Arduino connected to the DAC is temporally created to test the Arduino code and its communication with the DAC. Similarly, another test circuit is created with an external power supplying 12V to the Pololu connected to the linear actuator to send direct commands via the Windows program used to fix the PID parameters. These hardware unit tests are essential to the success of components' integration to the vehicle and make things easier later.

At system integration level, udev rules are used to facilitate testing with the creation of simlinks, i.e. dynamic assignments for the laptop USB ports connected to the Arduino and the Pololu, using their respective product and vendor IDs. This helps in being able to physically interchange the USB ports without having any impact at the software level. For the speed control, the vehicle wheels are lifted from the ground using jacks to stop them from driving off. This technique helps to test and fix the Arduino and ROS speed control code whilst staying in the same place. Vehicle steering is first tested using the Windows app that allows direct commands to be directly sent to the linear actuator. This helps verify and fix the linear actuator mounting as desired. Similarly, using Pololu's C++ API, direct commands are sent from a terminal to the linear actuator, but this time for testing at the software level.

Driving tests are initially performed in the manual joystick control mode in order to ensure that both hardware and software stack work well together. In particular, the LCD on the PCB board helps with checking in real-time the voltage received for each speed command and the LEDs colors displayed on the Pololu also give useful indications about the steering control.

The autonomous driving tests with move_base and TEB are performed with the vehicle speed controls dial know set to '5', corresponding to about 0.2m/s. This relatively low speed is chosen because these tests may be performed in a shared and cluttered research lab around people. Also, a large inflation distance is set in the planner to prevent the vehicle from close contacts with both static and dynamic obstacles. At first, simple and short goals are sent to move_base such as "drive one meter forward and keep your current orientation" or "drive three meters forward and keep your current orientation". Once the vehicle is able to execute and reach these simple goals, more complex goal commands are sent. Once a goal is reached, it is possible to resend immediately another goal without having to turn off the system, which is very convenient for example when one wants to ask the vehicle to return to its starting position or go somewhere else.

Setting a very high accuracy for goals such as 1mm and 0.01rad is achievable on the vehicle and is tested for short drives in the lab. However, in these cases, the short drives may end up taking a lot of time, for example it can take up to three minutes to simply reach a one meter forward goal. This is due to the planner's oscillating behaviour around the goal. To fix this, more tolerance should be given for the goal accuracy, for example, 150mm and 0.15rad give an acceptable vehicle behaviour. During these driving tests, ROS topics and RViz (ROS visualization tool) are particularly monitored to get informed about the vehicle behaviour in real-time.

OpenPodcar was developed, and our own build was heavily tested, between March 2018 and March 2022. With its first automated test drives taking place since summer 2018 and an estimated 100km or more driven to date, the vehicle design has thus proven robust enough for autonomous vehicle research.

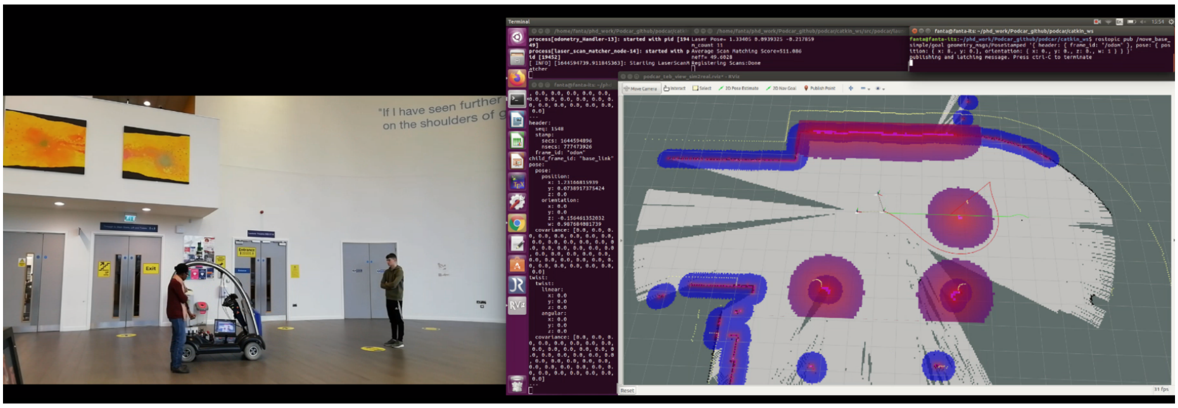


Figure 7.13: OpenPodcar test drive with GMapping SLAM, ROS move_base with TEB planner and obstacle avoidance.

4 Application

4.1 Use Cases

4.1.1 Self-Driving Research

Many AV researchers cannot currently afford the acquisition of a self-driving hardware platform for their work. The OpenPodcar is primarily designed for this purpose, as a low-cost and an all-in-one, software and hardware platform for researchers and hobbyists. Thus, giving them not only the opportunity to reproduce, develop and test algorithms on a physical hardware platform but also to extend its capabilities with new features.

The Related Systems section found that there are many open source software stacks without related open hardware platforms. OpenPodcar thus fills this gap, offering the opportunity not only to deploy Autoware or other types of AV software but also to extend the hardware capabilities to the point where OpenPodcar could become a standard test bed for the AV research community. For example, this platform could be useful to test different SLAM and planning algorithms, parallel and valet parking methods. The objective being that both hardware and software can be tested regularly in real-world conditions and contribute towards the deployment of AVs. The OpenPodcar can avoid both static and dynamic obstacle using the integrated feature in move_base and TEB planner. Fig. 7.13 shows the OpenPodcar test drive with GMapping, move_base and TEB planner in action when it encounters an obstacle on its path.

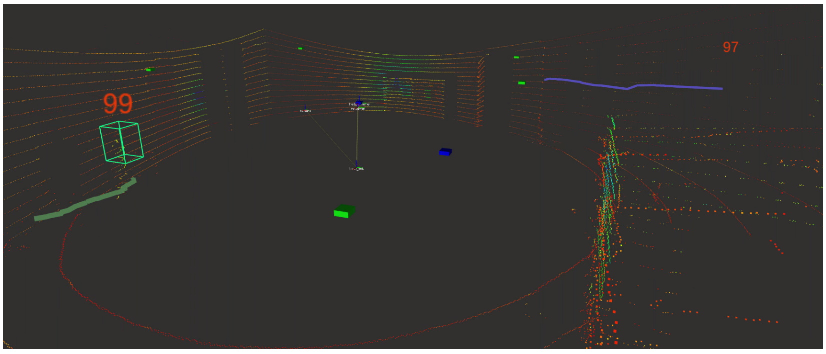


Figure 7.14: Pedestrian detection and tracking output from RViz.

4.1.2 Human-Robot Interaction Research

Understanding human behaviour and interaction strategies are of upmost importance nowadays for autonomous systems. There is a general growing interest from the robotics and autonomous vehicle research communities to tackle the numerous challenges posed by human interactions. Social robots as well as autonomous vehicles need better models of human behaviour [6, 7]. Some of the authors (FC and CF) are particularly interested in improving autonomous vehicles' decision-making using a game theoretic approach for road-crossing scenarios [20]. Several empirical studies e.g. [8, 9, 14], were performed in highly safe lab environments and found that human participants were not interacting realistically with the other agent. A similar experiment performed in a VR environment showed a more realistic behaviour from the participants [10, 11]. An additional model of human proxemics (i.e., interpersonal distances) has been developed and is being combined with the game theory model [13, 12]. In future work, the OpenPodcar will be used to extend these human experiments using a real physical platform and demonstrate the operation of game theoretic behaviour on a autonomous vehicle for the first time. The pedestrian detection and tracking feature will be particularly useful for this task, since the AV needs to track the pedestrian in order to make a decision. Fig. 7.14 shows an example output of the pedestrian detection and tracking integrated in the OpenPodcar.

4.1.3 Practical Transportation

OpenPodcar can carry at least 76kg of payload, such as a person or parcels, making it potentially useful for real-world as well as research applications.

Last mile delivery of parcels could replace human workers for e-commerce deliveries.

Figure 7.15: OpenPodcar test drive in remote control mode.

Urban center retail environments may also be improved by replacing the last mile of supply *to* retail outlets. Instead of driving to a shop to deliver goods, heavy goods vehicles could instead park a mile outside the urban center and transfer the goods to OpenPodcar or similar electric autonomous vehicles to take to the shop, reducing local pollution. The Covid-19 pandemic emphasised a specific need for *autonomous* last-mile delivery: to reduce the need for human contact and potential disease transmission at the point of delivery.

OpenPodcar is able to transport a human passenger, as shown in Fig. 7.15, as it is based on an COTS mobility scooter. For instance fleets of OpenPodcars might one day transport people over the last mile from the train station to their office, as a low cost electric taxi service. This will require more automation software to operate in busy urban environments.

4.2 Reuse Potential and Adaptability

The OpenPodcar design is intended so that the mechanical, electronics and software components can be easily ported to other vehicles/platforms and only require small changes on the software side to adapt it and fix some parameters specific to the new vehicle requirements. This could include future deeper OSH vehicles as well as additional commercial donor vehicles. Cheaper sensors such as depth cameras or stereo cameras could be used instead of the 3D lidar. Such modifications would typically require an advanced rather than intermediate designer/builder.

5 Build Details

5.1 Availability of Materials and Methods

The design is made under the CERN-OSH-W licence which allows for the use of commercially available proprietary components such as the off-the-shelf donor vehicle. However the design is intended to be easily modifiable for transfer to other base vehicles, including those which are OSH at lower levels. The PCB can be manufactured by many online PCB manufacturers. The additional mechanical and electronics used are common parts available from standard online vendors.

5.2 Ease of Build

The vehicle modification requires the use of common hand tools for assembly: spanners, screwdrivers, and pliers. Additionally, a 3D printer is needed to fabricate some components. Basic soldering skills are needed for assembling the PCB.

5.3 Operating Software and Peripherals

The system requires open source software: Arduino IDE, Ubuntu 16.04, ROS Kinetic, Gazebo, KiCad (PCB Design), ROS GMapping, ROS move_base, and Velodyne lidar driver. It also requires the Pololu Configuration Utility Manager software which is available gratis from the manufacturer website. The on-board laptop should have minimal specifications of amd64 3GHz quad-core, 8GB RAM, 250Gb hard-disc, USB and Ethernet ports. The system might also work on lower specifications. Step-by-step instructions for installation of these software dependencies, and the new system software components, are provided in the repository.

5.4 Hardware Documentation and Files Location:

Archive for hardware documentation, build files and software

Name: GitHub

Project repository: https://github.com/OpenPodcar/OpenPodcar

Licence: CERN-OHL-W for hardware design and build instructions; GPL for software source code.

Date published: 09/05/2022

The hardware is structured as two separate formal OSH designs, each licenced as CERN-OSH-W. The first covers all components which are easily transferable to other vehicles without modification. The second contains all components which are specific to the mobility scooter donor vehicle. This structure enables the first design to be used as sub-component of closed products while also preventing closed modifications of it.

6 Discussion

6.1 Conclusions

OpenPodcar is a multi-purpose hardware and software platform for autonomous vehicle research. It provides the required hardware and software tools to carry out research in this field. The platform has a lower-level stack, a higher-level stack and a simulator for initial testing. It has several safety features to prevent hazards. The general testing carried on the vehicle shows a robust and safe design. Several use cases have been identified and successfully tested. OpenPodcar is open source to allow further improvements and extensions of its capabilities from the community. The replication of this work on a second and later vehicles will help identify build issues and continually improve the documentation.

6.2 Future Work

OpenPodcar is designed to be extensible and modular, both at the hardware and software levels. As well as improving the current design, the community is warmly invited to create forks such as replacing the mobility scooter with other donor vehicles – including deeper OSH vehicles – or extending the ROS stack to more complex on-road self-driving systems such as Autoware.

In the current setup, the lidar has limited perception of obstacles that are too close and not as high as the lidar. This is generally fine, because people or objects would be seen before, but this can be problematic with objects such as desks and chairs that are not detected by the laserscans and can create unexpected collisions. For example, a low-cost alternative to lidar is to use a stereo camera for point cloud sensing. In this option, a StereoLabs ZedCam is mounted similarly on the vehicle roof.

The design currently uses ROS1 but the robotics community is slowly shifting to ROS2 for its security, real-time control and increased distributed processing features. OpenPodcar could join this shift when all of its ROS dependencies have themselves completed it.

The donor vehicle currently used it not itself OSH, and it would be interesting and useful to replace it with a more deep OSH vehicle. Such vehicles would be based on OSH motor drivers and controllers such as the brushed OSMC [40] or brushless ODrive v3.5 [32].

Paper author contributions

Fanta Camara performed the physical podcar automation work including the mechanical design, printed circuit board (PCB), ROS localisation and mapping, path planning, people detector and tracker integration, Gazebo 3D simulation worlds of Lincoln and Leeds university campuses, testing, and wrote the documentation for the physical podcar and the manuscript. Chris Waltham designed the initial electronics circuit and safety systems, wrote the Arduino code for the speed control and participated in the initial remotely-controlled driving test. Grey Churchill developed the podcar simulator with path planning in ROS/Gazebo and wrote the documentation for the simulator. Charles Fox supervised the work, wrote the manuscript, and wrote some ROS code. The manuscript was improved by comments from all the co-authors.

Acknowledgements

The authors would like to thank Jacob Lord for creating the vehicle graphical mesh model, Yao Chen for scoping Dubins path methods and the mechanical design, Gabriel Walton for scoping simulation tools, Yicheng Zhang for helping with the 3D printer and many other useful tools for the PCB board, Zak Burrows for the PCB enclosure design and assistance during some autonomous driving tests.

Funding statement

This project has received funding from EU H2020 grant 723395 interACT, and from InnovateUK grant 5949683 C19-ADVs.

Competing interests

The authors declare that they have no competing interests.

References

[1] F1tenth. https://f1tenth.org/.

[2] MIT RaceCar. https://mit-racecar.github.io/.

[3] MuSHR. https://mushr.io/.

[4] ApolloAuto. Apollo. https://github.com/ApolloAuto/apollo.

[5] Panpan Cai, Yiyuan Lee, Yuanfu Luo, and David Hsu. Summit: A simulator for urban driving in massive mixed traffic. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4023–4029. IEEE, 2020.

[6] Fanta Camara, Nicola Bellotto, Serhan Cosar, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, and Charles W. Fox. Pedestrian models for autonomous driving Part I: low-level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[7] Fanta Camara, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Anna Schieben, Gustav Markkula, Fabio Tango, Natasha Merat, and Charles W. Fox. Pedestrian models for autonomous driving Part II: high-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[8] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. Towards pedestrian-av interaction: method for elucidating pedestrian preferences. In *IEEE/RSJ Intelligent Robots and Systems (IROS) Workshops*, 2018.

[9] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. *Continuous Game Theory Pedestrian Modelling Method for Autonomous Vehicles*. River Publishers, 2020.

[10] Fanta Camara, Patrick Dickinson, and Charles Fox. Evaluating pedestrian interaction preferences with a game theoretic autonomous vehicle in virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, 78:410–423, 2021.

[11] Fanta Camara, Patrick Dickinson, Natasha Merat, and Charles W. Fox. Towards game theoretic av controllers: measuring pedestrian behaviour in virtual reality. In *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshops*, 2019.

[12] Fanta Camara and Charles Fox. Extending quantitative proxemics and trust to HRI. *(under review)*.

[13] Fanta Camara and Charles Fox. Space invaders: Pedestrian proxemic utility functions and trust zones for autonomous vehicle interactions. *International Journal of Social Robotics*, 2020. https://doi.org/10.1007/s12369-020-00717-x.

[14] Fanta Camara, Richard Romano, Gustav Markkula, Ruth Madigan, Natasha Merat, and Charles Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Measuring Behavior 2018: 11th International Conference on Methods and Techniques in Behavioral Research*. Manchester Metropolitan University, March 2018.

[15] Maxime Chevalier-Boisvert, Florian Golemo, Yanjun Cao, Bhairav Mehta, and Liam Paull. Duckietown environments for OpenAI Gym. https://github.com/duckietown/gym-duckietown, 2018.

[16] Fisher Daniel K and Gould Peter J. Open-source hardware is a low-cost alternative for scientific instrumentation and research. *Modern instrumentation*, 2012, 2012.

[17] Patricia R DeLucia. Effects of size on collision perception and implications for perceptual theory and transportation safety. *Current directions in psychological science*, 22(3):199–204, 2013.

[18] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.

[19] Lester E Dubins. On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. *American Journal of mathematics*, 79(3):497–516, 1957.

[20] Charles W. Fox, Fanta Camara, Gustav Markkula, Richard Romano, Ruth Madigan, and Natasha Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *VEHITS 2018: 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, January 2018.

[21] Brian Goldfain, Paul Drews, Changxi You, Matthew Barulic, Orlin Velev, Panagiotis Tsiotras, and James M Rehg. Autorally: An open platform for aggressive autonomous driving. *IEEE Control Systems Magazine*, 39(1):26–55, 2019.

[22] D Gomez, P Marin-Plaza, Ahmed Hussein, A Escalera, and J Armingol. Ros-based architecture for autonomous intelligent campus automobile (icab). *UNED Plasencia Revista de Investigacion Universitaria*, 12:257–272, 2016.

[23] Jon Gonzales. *Planning and Control of Drift Maneuvers with the Berkeley Autonomous Race Car*. PhD thesis, University of California at Berkeley, 2018.

[24] Marcin Jakubowski. Open Source Ecology. https://www.opensourceecology.org/, 2003.

[25] Shinpei Kato, Shota Tokunaga, Yuya Maruyama, Seiya Maeda, Manato Hirabayashi, Yuki Kitsukawa, Abraham Monrroy, Tomohito Ando, Yusuke Fujii, and Takuya Azumi. Autoware on board: Enabling autonomous vehicles with embedded systems. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, pages 287–296. IEEE, 2018. https://github.com/Autoware-AI/autoware.ai.

[26] Tobias Kessler, Julian Bernhard, Martin Buechel, Klemens Esterle, Patrick Hart, Daniel Malovetz, Michael Truong Le, Frederik Diehl, Thomas Brunner, and Alois Knoll. Bridging the gap between open source software and vehicle hardware for autonomous driving. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 1612–1619, 2019.

[27] Nathan Koenig and Andrew Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2149–2154. IEEE, 2004.

[28] Pablo Marin-Plaza, Ahmed Hussein, David Martin, and Arturo de la Escalera. Global and local path planning study in a ros-based research platform for autonomous vehicles. *Journal of Advanced Transportation*, 2018, 2018.

[29] Elie Michel. Maps Models Importer. https://github.com/eliemichel/MapsModelsImporter.

[30] William F Milliken, Douglas L Milliken, et al. *Race car vehicle dynamics*, volume 400. Society of Automotive Engineers Warrendale, PA, 1995.

[31] Naohiro Nakamoto and Hiroyuki Kobayashi. Development of an open-source educational and research platform for autonomous cars. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 6871–6876, 2019.

[32] ODrive Robotics. ODrive. https://odriverobotics.com/.

[33] Open Motors. Tabby EVO. https://www.openmotors.co/evplatform/.

[34] Jonathan Oxer and Hugh Blemings. *Practical Arduino: cool projects for open source hardware*. Apress, 2011.

[35] Peachtree Corners. Curiosity Lab. https://www.curiositylabptc.com/.

[36] Scott Pendleton, Tawit Uthaicharoenpong, Zhuang Jie Chong, Guo Ming James Fu, Baoxing Qin, Wei Liu, Xiaotong Shen, Zhiyong Weng, Cody Kamin, Mark Adam Ang, et al. Autonomous golf cars for public trial of mobility-on-demand service. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1164–1171. IEEE, 2015.

[37] PixMoving. Pixbot. https://gitlab.com/pixmoving/pixbot.

[38] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009.

[39] Craig Quiter. Deepdrive. https://github.com/deepdrive/deepdrive.

[40] Robot Power. Open Source Motor Control (OSMC). http://www.robotpower.com/products/osmc_info.html.

[41] Guodong Rong, Byung Hyun Shin, Hadi Tabatabaee, Qiang Lu, Steve Lemke, Mārtiņš Možeiko, Eric Boise, Geehoon Uhm, Mark Gerow, Shalin Mehta, et al. Lgsvl simulator: A high fidelity simulator for autonomous driving. *arXiv preprint arXiv:2005.03778*, 2020.

[42] Christoph Rösmann, Wendelin Feiten, Thomas Wösch, Frank Hoffmann, and Torsten Bertram. Efficient trajectory optimization using a sparse model. In *2013 European Conference on Mobile Robots*, pages 138–143. IEEE, 2013.

[43] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics*, 2017.

[44] Shoprider. Shoprider flagship luxury scooter model: Te-889xlsn user manual. https://www.usermanual.uk/shoprider/te-889xlsn/manual, 2016.

[45] Sebastian Thrun. Probabilistic robotics. *Communications of the ACM*, 45(3):52–57, 2002.

[46] Bastien Vincke, Sergio Rodriguez Florez, and Pascal Aubert. An open-source scale model platform for teaching autonomous vehicle technologies. *Sensors*, 21(11), 2021. https://github.com/BastienV-SATIE/AutonomousCar/.

[47] Cathy Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M. Bayen. Flow: Architecture and benchmarking for reinforcement learning in traffic control. *CoRR*, abs/1710.05465, 2017. https://flow-project.github.io/.

[48] Z. Yan, S. Schreiberhuber, G. Halmetschlager, T. Duckett, M. Vincze, and N. Bellotto. Robot perception of static and dynamic objects with an autonomous floor scrubber. *Intelligent Service Robotics*, 13(3):403–417, 2020. https://github.com/LCAS/FLOBOT.

[49] Barry Loh Tze Yuen, Khairul Salleh Mohamed Sahari, and Zubaidi Faiesal Mohamad Rafaai. Improved map generation by addition of gaussian noise for indoor slam using ros. *Journal of Robotics, Networking and Artificial Life*, 4(2):118–123, 2017.

Copyright notice

Chapter 8

Discussion and Conclusions

"I dream of a better tomorrow, where chickens can cross the road and not be questioned about their motives." Ralph Waldo Emerson (1803–1882)

1 Summary

In this thesis, we have shown how to infer and operate pedestrian behaviour on autonomous vehicles, basing this work on the research gaps identified in the introduction (cf. Chapter 1). In Chapters 2 and 3, the two narrative reviews found that low-level pedestrian models are mature enough to be operational but highlighted the need for the development and operation of more high-level pedestrian models such as game theoretic models. In Chapter 4, we implemented the Sequential Chicken game theoretic model on an autonomous vehicle in a VR environment for road-crossing experiments with human participants. We inferred pedestrian interaction preference parameter values from these experiments using Gaussian Process regression. The results were found to be more realistic than previous empirical experiments and the parameter values could be inserted into future experiments or real-world AV controllers. In Chapter 5, we proposed novel quantitative approaches for proxemics and physical trust requirement (PTR) for operation into game theoretic AV controllers. We found that a hyperbolic function best describes pedestrian proxemics utility and showed that our PTR results closely match Hall's empirical zone sizes. In Chapter 6, we addressed some limitations of the PTR model and extended it to model more general human-human and human-robot interactions. Finally, in Chapter 7, we developed OpenPodcar, a low cost and open source hardware and software self-driving platform providing SLAM, path planning and pedestrian detection and tracking algorithms on-board, for the operation of real-world pedestrian interaction models including the models proposed in Chapters 4 to 6.

2 Limitations and Future Work

There are some limitations with this work. First, the literature reviews presented in Chapters 2 and 3 need to be updated regularly in order to take into account the latest models introduced in the fast-moving autonomous driving field. For example, since the publication of the reviews, recurrent and graph neural networks have recently been applied to tracking, benefiting from the recent availability of GPU computing and outperforming probabilistic methods in many cases [9, 13, 14]. Moreover in these reviews, only a selected number of papers have been reviewed and we did not go into the implementation details of the models including possible and different data types available from sensors like lidar. However, it is worth noting that a survey on human motion prediction [11], that appeared on arXiv shortly after our reviews were submitted, corroborated our main findings and highlighted the need for game theoretic approaches and human social cues to improve autonomous systems' decision-making.

The experimental data used is relatively small for the VR experiments with just a few participants and less than a hundred of distinct pedestrian-vehicle interactions in the proxemics work, therefore future work should consider a much larger number of participants and interaction data in order to make the results more robust. This work has mainly focused on a pair of interacting agents, but pedestrian behaviour can be affected by that of other road users, similar to human drivers, so methods to manage multiple and simultaneous interactions should be explored, similar to [6, 7]. This will require the model to become more complex and will also need a lot more computation but this is necessary in order to embed more human intelligence into autonomous vehicles. Also, faster methods to compute pedestrian preferences could be useful, because the Gaussian Process regression and the gradient descent approach used in the VR experiments can take some time to compute, so newer and faster methods should be explored in the future, and to estimate the AV driving parameters which may be learned directly from human drivers and then insert into the AV controller.

The proxemic utility functions used in this work may need in the future to include speed considerations, because an AV driving at 1m/s that invades someone's personal space won't be perceived in the same way as an AV driving at 10m/s. We assumed that agents move along straight paths, but the game theory and proxemics models should allow lateral movements, as was hinted in Option 4 from Chapter 6, but this will need to be fully implemented. Also, this work was undertaken from a UK point of view, but it was shown that cultural differences affect driving, crossing and proxemic behaviours, therefore the results may not translate directly from a country to another [4].

In this thesis, we mainly focused on inferring and operating pedestrian behaviour from positional data, i.e. there was no communication between the agents, we thus did not consider pedestrian visual features such as age, head orientation or hand gestures, which have been shown to influence crossing decisions, so they should form part of future work to improve the accuracy of the predictions, similar to the approach used by Ma et al. [5]. We may also consider signalling methods such as using flashing headlights and horns reviewed in Chapter 3 to communicate the AV's intent to pedestrians, similar to the approach used by [15] for inter-pedestrian signalling. Some of our work [1, 2] excluded from this thesis focused on interaction sequence analysis, where observed pedestrian-vehicle interactions were decomposed into a set of discrete actions or features, and we looked at how each feature could be predictive of a crossing behaviour. Integrating this into the game theory model could take the form of an online feature tracker and pedestrian behaviour predictor. For instance, a hand gesture would represent a feature with a specific utility value, and when the AV observes such a feature, this would give some hints about the pedestrian crossing intention and the AV may then take some actions in response, such as slowing down or speeding up. However, such utility values would first need to be inferred and measured from real-world pedestrian-vehicle interactions or experiments.

Furthermore, we aim to integrate all the models described in this thesis into a single and operational model for autonomous vehicle control. This would consists in combining the game theory model with the PTR model and using the OpenPodcar platform to test the overall performance of this approach as well as testing advanced open source software such as Autoware [3] in real-world settings. Even further, we plan to make the OpenPodcar platform more open source, i.e. it currently relies on a proprietary mobility scooter with some closed parts such as the motor driver and controller, which could be replaced by the brushed open source motor controller (OSMC) [10] or brushless ODrive v3.5 [8]. Also, OpenPodcar should move from using ROS1 to ROS2 for its secure, real-time and distributed features, and the open source nature of the OpenPodcar allows anyone who is interested to contribute to and tailor the project to the general needs of the community.

Finally, the work presented in this thesis forms a step towards the concrete operation of game theoretic autonomous vehicles with pedestrian proxemics and trust behaviour. This may have some policy implications. For example, operating autonomous vehicles with a collision probability may be considered as unethical or illegal and is under debate [12]. Hence we hope that the approach used in this thesis, that replaces the large negative utility of a crash by human natural proxemic behaviour to inflict a much smaller negative utility, contributes to this debate and may make autonomous vehicles safer and more efficient on the roads. Future work should therefore investigate the ethical and legal considerations as well as societal impacts of such models.

3 Concluding Remarks

The use of game theory and human nonverbal cues is emerging in transport and robotics research. For this reason, we restate the key contributions from this thesis:

- a comprehensive review of pedestrian behaviour models for autonomous driving;
- a new unifying taxonomy based on probability theory linking pedestrian models to SAE levels of automation for the first time;
- the operation of the Sequential Chicken game theoretic model on a autonomous vehicle in a VR environment;
- methods to infer pedestrian interaction preferences;
- a novel Bayesian method to infer pedestrian proxemic utility functions;
- a novel concept of physical trust requirement (PTR) linking proxemics and trust for the first time;
- results of the PTR model closely matching Hall's empirical proxemic zone sizes;
- the extension of the PTR model to include more general human-human and human-robot interactions;
- OpenPodcar, an open source hardware and software platform for autonomous vehicle research providing a higher-level stack where pedestrian models could be integrated to.
- a set of experiments, simulations and use cases to test this work.

We hope that the new links that this work creates between game theory and nonverbal behaviour will contribute to their use more widely and make them mature enough for safe operation in autonomous vehicles. We also hope that other fields such as social robotics, gaming, behavioural and social sciences will benefit from these new bridges.

References

[1] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proceedings of the 15th International Conference on Intelligent Autonomous Systems (IAS-15) workshops*, 2018.

[2] Fanta Camara, Oscar Giles, Ruth Madigan, Markus Rothmüller, Pernille Holm Rasmussen, Signe Alexandra Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and Charles W. Fox. Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. In *Proceedings of IEEE 21st International Conference on Intelligent Transportation Systems*, 2018.

[3] Shinpei Kato, Shota Tokunaga, Yuya Maruyama, Seiya Maeda, Manato Hirabayashi, Yuki Kitsukawa, Abraham Monrroy, Tomohito Ando, Yusuke Fujii, and Takuya Azumi. Autoware on board: Enabling autonomous vehicles with embedded systems. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, pages 287–296. IEEE, 2018. https://github.com/Autoware-AI/autoware.ai.

[4] Yee Mun Lee, Ruth Madigan, Oscar Giles, Laura Garach-Morcillo, Gustav Markkula, Charles Fox, Fanta Camara, Markus Rothmueller, Signe Alexandra Vendelbo-Larsen, Pernille Holm Rasmussen, et al. Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology & Work*, 2020.

[5] Wei-Chiu Ma, De-An Huang, Namhoon Lee, and Kris M. Kitani. Forecasting interactive dynamics of pedestrians with fictitious play. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4636–4644, 2017.

[6] Christoforos Mavrogiannis, Patrícia Alves-Oliveira, Wil Thomason, and Ross A Knepper. Social momentum: Design and evaluation of a framework for socially competent robot navigation. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(2):1–37, 2022.

[7] Christoforos I Mavrogiannis and Ross A Knepper. Multi-agent path topology in sup-

port of socially competent navigation planning. *The International Journal of Robotics Research*, 38(2-3):338–356, 2019.

[8] ODrive Robotics. ODrive. https://odriverobotics.com/.

[9] Ratheesh Ravindran, Michael J Santora, and Mohsin M Jamali. Multi-object detection and tracking, based on dnn, for autonomous vehicles: A review. *IEEE Sensors Journal*, 21(5):5668–5677, 2020.

[10] Robot Power. Open Source Motor Control (OSMC). http://www.robotpower.com/products/osmc_info.html.

[11] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. Human motion trajectory prediction: a survey. *The International Journal of Robotics Research*, 39(8):895–935, 2020.

[12] UK Law Commission. Automated Vehicles: a joint preliminary consultation paper, 2019. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/11/6.5066_LC_AV-Consultation-Paper-5-November_061118_WEB-1.pdf.

[13] Yongxin Wang, Kris Kitani, and Xinshuo Weng. Joint object detection and multi-object tracking with graph neural networks. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13708–13715. IEEE, 2021.

[14] Xinshuo Weng, Yongxin Wang, Yunze Man, and Kris M Kitani. Gnn3dmot: Graph neural network for 3d multi-object tracking with 2d-3d multi-feature learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6499–6508, 2020.

[15] Yifei Wu and Hansong Li. Signalling security: An observational and game theory approach to inter-pedestrian psychology. *Transportation Research Part F: Traffic Psychology and Behaviour*, 86:238–251, 2022.

Appendix A

Appendix to Chapter 2

1 Quality of Citations

These linked papers (Part I and II) review over 450 papers from high quality journals and conferences such as *CVPR, ICRA, PAMI, IROS, ITSC, ECCV, IV*. It is common in Computer Science fields including machine vision and machine learning for conferences to be considered higher quality or similar quality to journals, while psychology and sociology fields typically consider journals to be more authoritative. The following figures give some ideas about the quality of the cited papers.

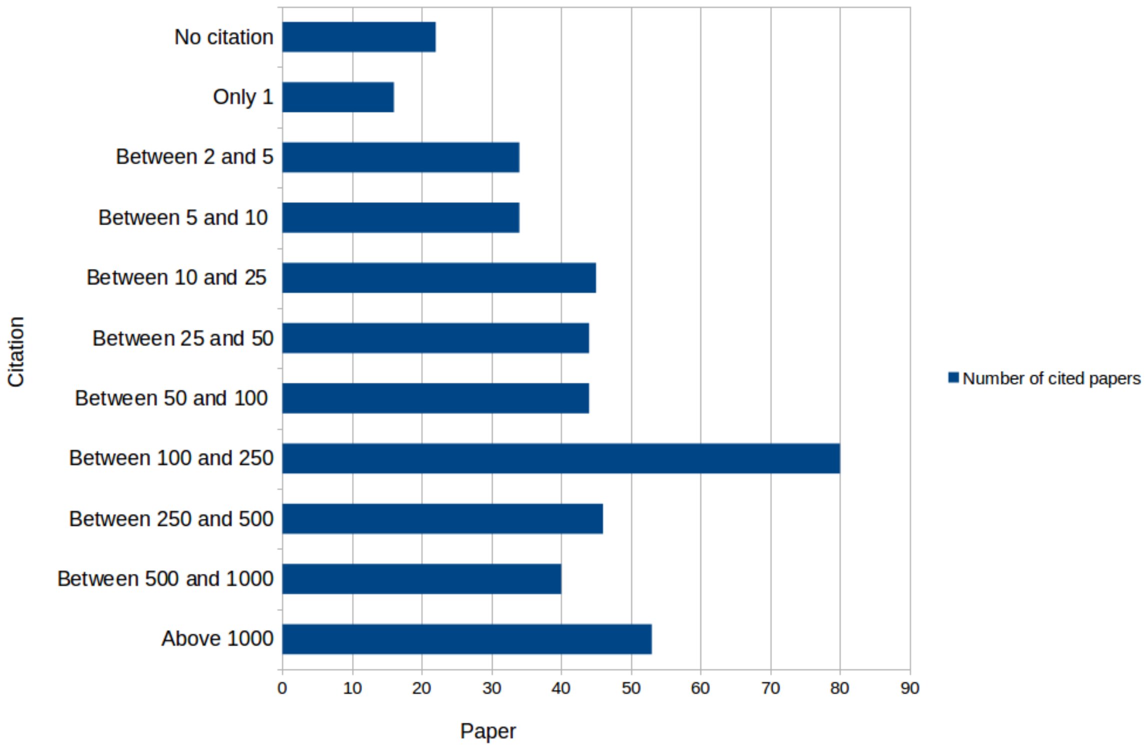


Figure A.1: Number of citations per paper

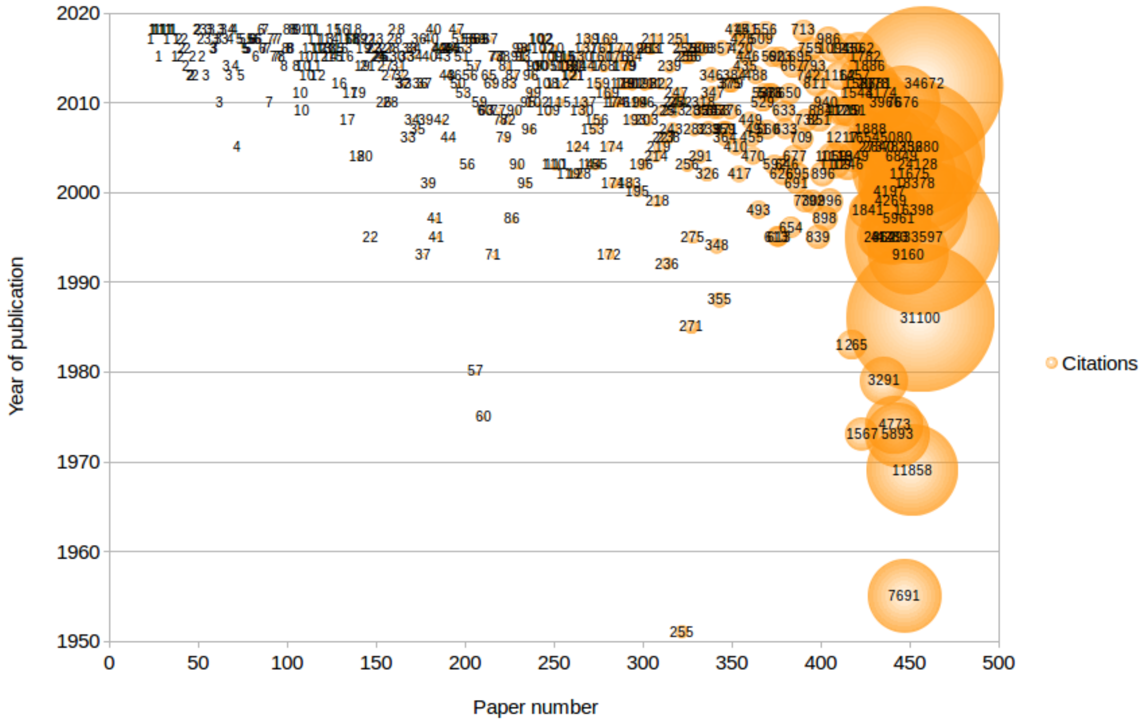


Figure A.2: Number of citations per paper and per year of publication

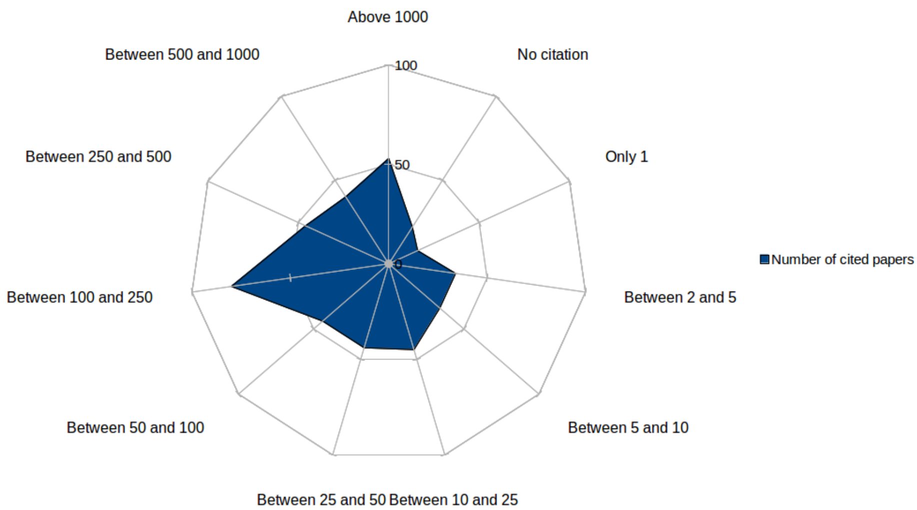


Figure A.3: Number of citations per paper

2 Summary of Pedestrian Recognition Models

Table A.1: Summary of the recognition models

Study/Paper	Input/Evaluation	Method	Recognition Models	Additional Info	SAE Level
Cao et al. [28] [29]	Images	CNN model with Part Affinity Fields (PAF)	Pose estimation	OpenPose: open-source software	Level 2
Shotton et al. [92]	Motion capture and synthetic data	Body parts representation model	3D human pose estimation		Level 2
Iqbal et al. [57]	Video data	Graphical model	Pose estimation and tracking	Release of PoseTrack a new dataset	Level [2,3]
Tompson et al. [98]	Monocular images	Deep CNN model with Markov Random Field	Pose estimation		Level 2
Fragkiadaki et al. [48]	Motion capture data: H3.6M dataset [Ionescu et al. 2014]	Encoder-Recurrent-Decoder (ERD)	Body pose estimation		Level [2]
Martinez et al. [73]	Motion capture data: H3.6M dataset [Ionescu et al. 2014]	Recurrent neural network with a gated recurrent unit (GRU)	Body pose estimation		Level 2
Tang et al. [94]	Motion capture data: H3.6M dataset [Ionescu et al. 2014]	Deep neural network (modified High-way Unit (MHU))	Pose estimation		Level 2
Ghosh et al. [52]	Motion capture data: datasets in [Ionescu et al. 2014] and [Holden et al. 2016]	Dropout AutoEncoder LSTM (DAE-LSTM)	Pose estimation		Level 2
Ma et al. [72]	Images	CNN model	Body heading	No annotations needed	Level 2
Kohari et al. [62]	Video	CNN model	Body orientation	Service robot	Level 2
Darrell et al. [41]	Images	Statistical model	Head direction	From a mobile robot	Level [2,3]
Schulz et al. [90]	Grayscale images	Multi-classifiers	Head pose		Level 2
Benfold et al. [23]	Video	HOG and colour features	Gaze tracking		Level 2
Baltrusaitis et al. [20]	Video	Deep learning model	Head pose and eye-gaze estimation	Openface: open source software	Level 2

Cornejo et al. [39] [40]	Images	Principal Component Analysis + Gabor wavelets or CENTRIST features	Emotion recognition		Level 2
Cambria et al. [27] [26]	Images	Review paper	Sentiment analysis		[Level 2,3]
Poria et al. [83]	Videos	CNN model with recurrent multilayer kernel learning	Emotion recognition		Level 2
Horng et al. [54]	Video	Dynamic template matching	Driver fatigue detection		Level 2
Denuyl et al. [43]	Video		Face expression recognition	FaceReader: commercial product	Level 2
Ahmed et al. [17]	Images	Deep neural networks	Re-identification		Level 2
Zheng et al. [108]	Images	Bag of Words model	Re-identification		Level 2
Zheng et al. [109]	Images	CNN model	Re-identification	Unlabeled images	Level 2
Li et al. [66]	Images	Filter airing neural network (FPNN) model	Re-identification	Occlusion handling	Level 2
Chen et al. [32] [24] [65]	Images	Hidden Markov model (HMM)	Gesture recognition		Level 2
Freeman et al. [49]	Images	Orientation histograms	Gesture recognition	10 different hand gestures recognition	Level 2
Ren et al. [86]	Images	Template matching with Finger-earth Mover's Distance (FEMD)	Gesture recognition		Level 2
Quintero et al. [84]	Images	Hidden Markov models (HMM)	Body language Recognition		Level [2,3]
Wang et al. [100]	Images	Background subtraction + PCA	Body language Recognition		Level 2
Chaaraoui et al. [30]	Videos	Contour points	Activity recognition	Real-time method	Level 2
Dollár et al. [45]		Spatio-temporal features	Activity recognition		Level 2
Vail et al. [99]	Videos	Hidden Markov models and Conditional random field	Activity recognition		Level 2

Liu et al. [68]	RGB data	Coupled conditional random field	Activity recognition		Level 2
Coppola et al. [38]	RGB-D data	Dynamic Bayesian mixture model (DBMM)	Activity recognition		Level 2

3 Summary of Pedestrian Tracking Models

Table A.2: Summary of pedestrian tracking models

Study/Paper	Input/Evaluation	Method	Tracking Models	Additional Info	SAE Level
Del Pino et al. [42]	Low resolution LiDAR data	Multi-Hypothesis EKF (MHEKF)	Point Tracking		Level 2
Bellotto et al. [22]	Robot with laser and camera	EKF, UKF, SIR Particle filter	Point Tracking	Trade-off between performance and computation cost	Level 2
Arulampalam et al. [18]	Example	Particle filter implementations	Point Tracking		Level 2
Fen et al. [46]	Video data	Color histogram based particle filter	Point Tracking		Level 2
Jurie et al. [58]	Video data	Template matching with SSD	Kernel-based Human Tracking	Real-time method and robustness to occlusions and illuminations	Level 2
Lipton et al. [67]	Video data	Frame differencing + Template matching	Kernel-based Human Tracking	Real-time method	Level 2
Kaneko et al. [59]	Image sequences	Template matching with a feature selection method	Kernel-based Human Tracking		Level 2
Comaniciu et al. [37]	Moving camera data	Mean-shift algorithm with Bhattacharyya coefficient	Kernel-based Human Tracking		Level 2
Collins et al. [35]	Video data	Mean-shift algorithm with 2d blob tracking	Kernel-based Human Tracking		Level 2

Tao et al. [95]	Airborne vehicle tracking system	Dynamic layering method + MAP using EM algorithm	Kernel-based Human Tracking		Level 2
Yalcin et al. [101]	Image sequences	Layering method with optical flow	Kernel-based Human Tracking		Level [2,3]
Geiger et al. [51]	Image sequences	Contour matching method based on Dynamic programming	Tracking pedestrian body state		Level 2
Techmer et al. [97]	Real-world images	Contour tracking with distance transformations of contour images	Tracking pedestrian body state		Level 2
Baumberg [21] Yilmaz [103]	Image sequences	Dynamic Kalman filter with active shape model	Tracking pedestrian body state	Method sensitive to initialization	Level 2
Adam et al. [16]	Image sequences	Region color histogram method	Tracking pedestrian body state	Occlusion and pose changes handling	Level 2
Meyer et al. [74]	Image sequences	Recursive algorithm using image regions information	Tracking pedestrian body state		Level 2
Collins et al. [36]	Gait databases: CMU, MIT, UMD, USH	Silhouette based model	Tracking pedestrian body state	To identify people from their body and gait	Level 2
Schwarz et al. [91]	Kinect data	Graph method with skeleton fitting	Tracking pedestrian body state	Full-body tracker	Level 2
Sinthanayothin et al. [93]	Kinect data	Skeleton tracking	Tracking pedestrian body state	Review paper	Level 2
Konstantinova et al. [63]	5 test matrices	Global Nearest Neighbor with Munkres algorithm	Multi-Target Tracking (MTT)		Level [2, 3, 4]
Azari et al. [19]	IBM, PETS2000 and PETS2001 databases	Kalman filter with Global Nearest Neighbor (GNN)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3,4]
Reid et al. [85]	Monte Carlo simulation	Iterative algorithm with Multi-Hypothesis Tracking(MHT)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]

Luber et al. [71]	Two datasets collected in indoor and outdoor environments	Social force with Multiple Hypothesis Tracking (MHT)	Multi-Target Tracking (MTT)		Level [3,4]
Kim et al. [61]	PETS and MOTChallenge benchmarks	Multiple Hypothesis Tracking (MHT)	Multi-Target Tracking (MTT)		Level [2,3]
Zhou et al. [110]	Computer simulations	Joint probabilistic data association filter (JPDAF) with a depth-search approach	Multi-Target Tracking (MTT)		Level [2,3]
Chen et al. [33]	Video data	Contour based tracker with JPDAF and HMM	Multi-Target Tracking (MTT)	Real-time method	Level [2,3]
Liu et al. [69]	Simulations and real robot	Sample-based JPDAF and multi-sensor fusion	Multi-Target Tracking (MTT)	Real-time method	Level [2,3]
Horridge et al. [55]	Simulations	JPDAF based tracker	Multi-Target Tracking (MTT)	400 tracks in real-time	Level [2,3,4,5]
Rezatofighi et al. [87]	Fluorescence microscopy sequences and surveillance camera data	JPDAF based tracker	Multi-Target Tracking (MTT)		Level [2,3]
Zhang et al. [107]	Simulations	Gaussian Mixture Measurement PHD tracker (GMM-PHD)	Multi-Target Tracking (MTT)	Handle bearing measurements	Level [2,3]
Khazaei et al. [60]	Data from a distributed network of cameras	Probabilistic Hypothesis Density (PHD) filter based tracker	Multi-Target Tracking (MTT)		Level [2,3]
Feng et al. [47]	Simulations with sequences from CAVIAR dataset	Variational Bayesian PHD filter with deep learning updates	Multi-Target Tracking (MTT)		Level [2,3]
Correa et al. [40]	Tested on a real-time crowded environment	PHD filter	Multi-Target Tracking (MTT)		Level [4, 5]
Yoon et al. [104]	ETH dataset	Sequential Monte Carlo PHD filter (SMC-PHD)	Multi-Target Tracking (MTT)	Can handle missing detections	Level [2,3]

Oh et al. [79] [78]	Simulations	Markov Chain Monte Carlo Data Association (MCMCDA) Metropolis-Hastings method	Multi-Target Tracking (MTT)		Level [2,3]
Yu et al. [105]	Simulations and video data	Data-driven Markov Chain Monte Carlo data association (DD-MCMCDA)	MTMulti-Target Tracking (MTT)		Level [2,3]
Chen et al. [56]	Video data	Dynamical graph matching	Multi-Target Tracking (MTT)	Tracker can deal with interactions	Level [2,3]
Pirsiavash et al. [82]	Video data	Greedy algorithm based on Dynamic Programming	Multi-Target Tracking (MTT)		Level [2,3]
Zhang et al. [106]	CAVIAR and ETHMS datasets	Min-Cost Flow algorithm with an explicit occlusion model (EOM)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]
Chari et al. [31]	PETS and TUD datasets	Min-Cost Max-Flow network optimization with pair-wise costs	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]
Taycher et al. [96]	Video data	Conditional Random Field (CRF) state estimation and grid filter	Multi-Target Tracking (MTT)	Real-time capability	Level [2,3]
Milan et al. [76]	PETS 2010 Benchmark and TUD-Stadtmitte dataset	CRF-based multiple tracker with HOG-SVM detector	Multi-Target Tracking (MTT)		Level [2,3]
Milan et al. [75]	PETS, TUD, ETHMS datasets	CRF-based multi-target tracker using discrete continuous minimization	Multi-Target Tracking (MTT)	Trajectory estimation of targets	Level [2,3,4]
Brendel et al. [25]	ETHZ Central, TUD Crossing, i-Lids AB, UBC Hockey and ETHZ Soccer datasets	Maximum-weight independent set (MWIS) based tracker	Multi-Target Tracking (MTT)	Long-term occlusion handling	Level [2,3]

He et al. [53]	PETS09, TUD Statmitte, TUD Crossing and ETHMS datasets	Connected component model with MWIS	Multi-Target Tracking (MTT)		Level [2,3]
Gaidon et al. [50]	KITTI Benchmark and PASCAL-to-KITTI dataset	Online Domain Adaptation for Multi-Object Tracking	Multi-Target Tracking (MTT)	Generic detector and video adaptation fro tracking	Level [2,3]
Ondruska et al. [80]	Simulated data	End-to-end recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)	No data association required	Level [2,3]
Dequaire et al. [44]	Real-world environment	End-to-end recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)		Level [2,3]
Milan et al. [77]	MOTChallenge 2015 benchmark	Online recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)		Level [2,3]
Ristani et al. [88]	Multi-cameras system data	Neural networks	Multi-Target Tracking (MTT)	Features learnt multi-cameras and Re-identification	Level [3, 4]
Liu et al. [70]	Multi-camera systems data	Generalized Maximum Multi-Clique optimization	Multi-Target Tracking (MTT)		Level [2,3]
Park et al. [81]	Monocular camera data	3D object tracking	Multi-Target Tracking (MTT)	3D object Tracking for augmented reality applications	Level [2,3]
Scheidegger et al. [89]	Single camera data	Multi-Bernoulli mixture tracking filter	Multi-Target Tracking (MTT)		Level [2,3,4]
Yan et al. [102]	Lidar data	3D LIDAR based tracking with Support Vector Machine (SVM) classifier	Multi-Target Tracking (MTT)	Online classification of humans	Level [2,3]
Leal-taixe et al. [64]	Camera data	Interaction feature strings with Random Forest method	Multi-Target Tracking (MTT)	Scene understanding	Level [2,3]
Choi et al. [34]	Video data	Discriminative model	Multi-Target Tracking (MTT)	Group activity recognition	Level [2,3]

Appendix B

Appendix to Chapter 3

1 Quality of Citations

These linked papers (Part I and II) review over 450 papers from high quality journals and conferences such as *CVPR, ICRA, PAMI, IROS, ITSC, ECCV, IV*. It is common in Computer Science fields including machine vision and machine learning for conferences to be considered higher quality or similar quality to journals, while psychology and sociology fields typically consider journals to be more authoritative. The following figures give some ideas about the quality of the cited papers.

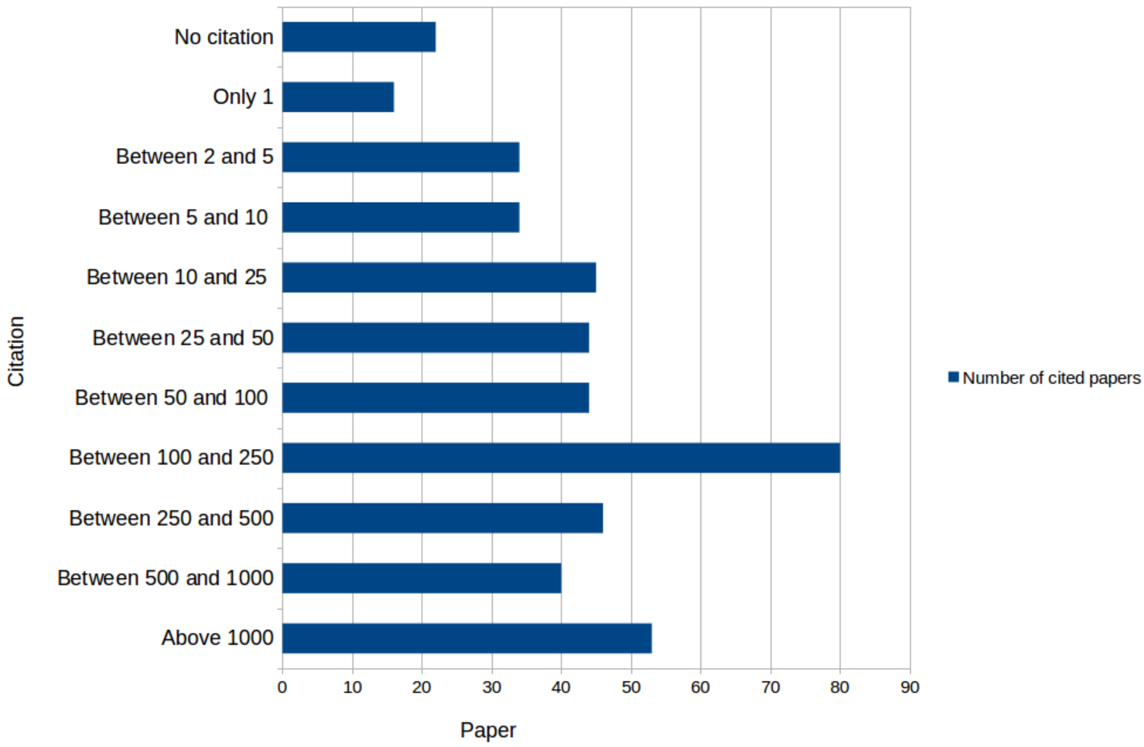


Figure B.1: Number of citations per paper

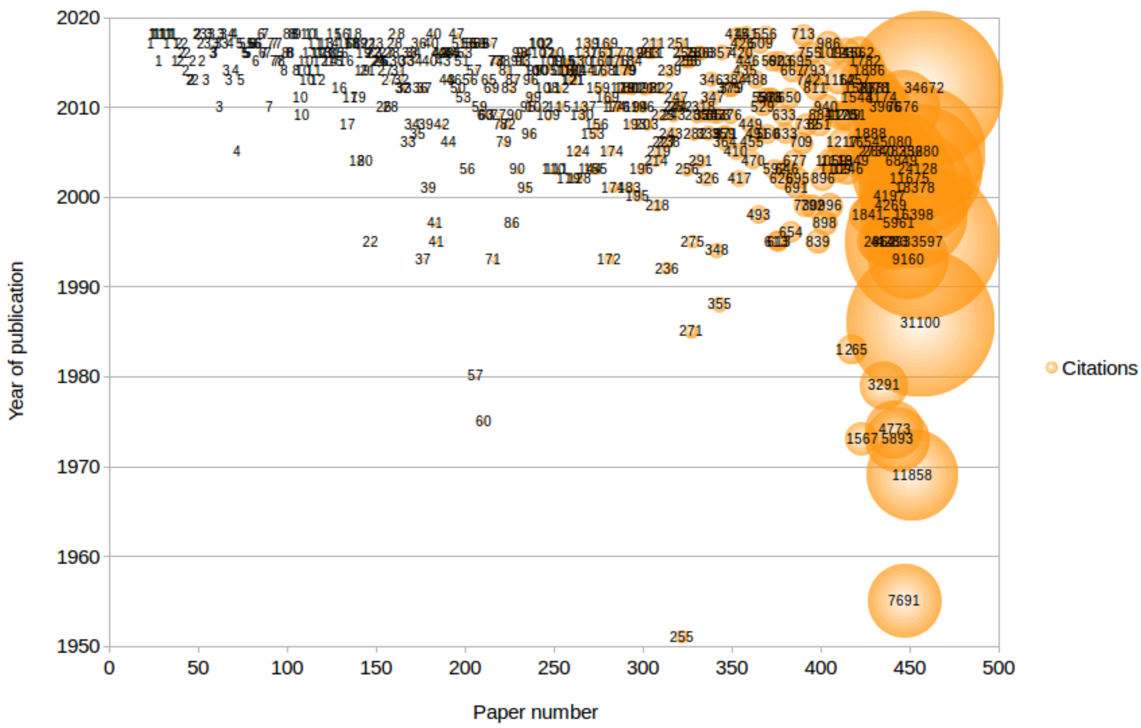


Figure B.2: Number of citations per paper and per year of publication

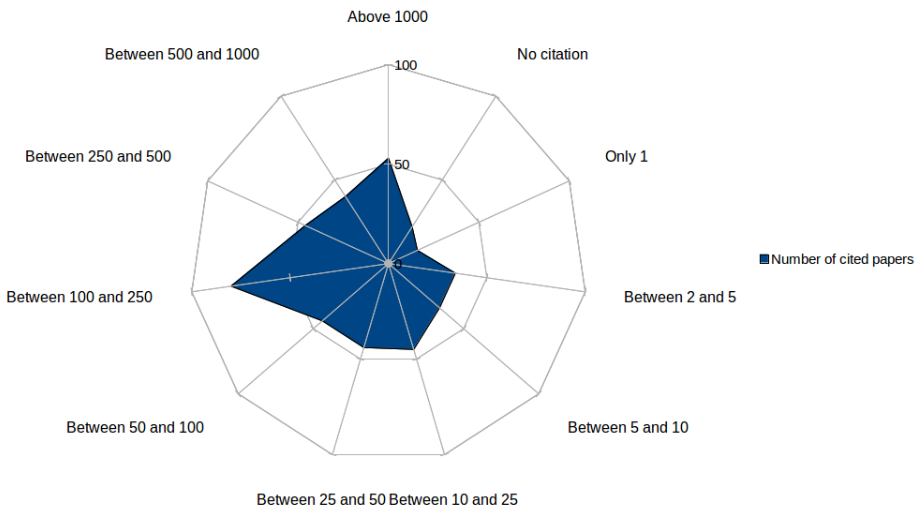


Figure B.3: Number of citations per paper

2 Pedestrian Trajectory and Interaction Models

Table B.1: Summary of pedestrian trajectory prediction and interaction models

Study/Paper	**Input/Evaluation**	**Method**	**Trajectory Prediction Models**	**SAE Level**
Hoogendoorn et al. [162]	Simulated trajectories	Optimal control theory	Unobstructed walking paths	Level [4,5]
Antonini et al. [114]	Pedestrian movements (video sequence)	Discrete choice model	Unobstructed walking paths Two agents' interaction	Level 3
Borgers et al. [126]	Pedestrian movements (manual)	Discrete choice	Unobstructed walking paths	Level 3
Puydupin-jamin et al. [200]	Pedestrian trajectories dataset [115]	Unicycle model with inverse optimal control	Unobstructed walking paths	Level 3
Habibi et al. [159]	Pedestrian trajectories dataset from two intersections [159]	Gaussian Process with a Transferable ANSC algorithm	Route prediction around obstacles Gaussian Process methods	Level [3,4]
Kitani et al. [171]	92 videos (80% for training)	Inverse reinforcement learning with inverse optimal control theory	Uncertain destination models Dynamic graphical models	Level [3,4]
Ziebart et al. [240]	Predicted trajectories used in an incremental motion planner	Maximum entropy with inverse optimal control	Uncertain destination models Dynamic graphical models	Level [3,4]
Vasquez et al. [226]	14 pedestrian trajectories dataset [171]	Markov decision process (MDP) with a Fast Marching Method (FMM)	Uncertain destination models Dynamic graphical models	Level 4
Gockley et al. [154] Topp et al. [218]	Laser data	Direction-following and Path-following with Curvature velocity method	Route prediction around obstacles	Level 3
Bennewitz et al. [119, 118]	Laser range data from Pioneer robots	Clustering with Expectation Maximization (EM) algorithm	Uncertain destination models Dynamic graphical models	Level 3
Wu et al. [230]	Pedestrian trajectories from Rutesheim dataset	Markov chains with an heuristic method	Uncertain destination models Dynamic graphical models	Level 3
Karasev et al. [168]	Pedestrian dataset [168]	Markov decision process (MDP) with Rao-Blackwellized filter	Uncertain destination models Dynamic graphical models	Level [4,5]

Bai et al. [116]	Autonomous golf car	Partially observable Markov decision process (POMDP)	Uncertain destination models Dynamic graphical models	Level [4,5]
Rehder et al. [206]	Pedestrian trajectories (stereo video dataset)	Inverse reinforcement learning	Uncertain destination models Deep learning methods	Level [4, 5]
Garzon et al. [150]	Simulations and real-world data	Fast Marching Method (FMM) and A Star (A*) algorithm	Uncertain destination models	Level 4
Kooij et al. [172]	Pedestrian trajectories dataset [172]	Dynamic Bayesian network (DBN) with a switching linear dynamic system (SLDS)	Event/activity models Dynamic graphical methods	Level [4,5]
Schulz et al. [213]	Pedestrian dataset [212]	Interacting Multiple Model (IMM) with a latent-dynamic conditional random field (LDCRF)	Event/activity models Dynamic graphical methods	Level [4,5]
Dondrup et al. [142]	Laser and RGB-D data	Qualitative Spatial Relations (QSR) with Hidden Markov model (HMM)	Event/activity models Dynamic graphical methods	Level [4,5]
Bonnin et al. [124]	Pedestrian datase [124]	Inner-city model with 'Context Model Tree' approach	Event/activity models	Level [4,5]
Borgers et al. [125]	Pedestrian dataset from the city of Maastricht	Discrete Choice Model	Event/activity models	Level 3
Camara et al. [131] [130]	Pedestrian-vehicle interactions dataset [131]	Regression models Filtration analysis	Event/activity models	Level [4,5]
Völz et al. [227]	LIDAR pedestrian trajectories	Support vector machine (SVM)	Event/activity models	Level 3
Duckworth et al. [143] [144]	Pedestrian dataset from a mobile robot [144]	Qualitative Spatial Analysis (QSR) with a graph representation	Event/activity models	Level [4,5]
Mögelmose et al. [192]	Pedestrian trajectories from a monocular camera	Particle filter	Route prediction around obstacles Dynamic graphical methods	[Level 3,4]
Schneider et al. [212]	Pedestrian dataset [212]	Extended Kalman filter (EKF) and Interacting Multiple Model (IMM)	Event/activity models Dynamic graphical methods	Level [3,4]

Quintero et al. [202, 201]	CMU Dataset with 129 video sequences	Balanced Gaussian process dynamical models (B-GPDMs) and Naive Bayesian classifiers	Event/activity models Dynamic graphical methods	Level[3,4,5]
Goldhammer et al. [155]	Camera data	Multilayer perceptron (MLP) with polynomial least square approximation	Uncertain destination models Deep learning methods	Level [4,5]
Kruse et al. [176]	Camera data	Statistical analysis	Route prediction around obstacles	Level [3]
Cosgun et al. [137]	Real robot	Motion planning with curvature velocity method	Uncertain destination models	Level [3,4]
Koschi et al. [173]	Real world data from a moving vehicle	Set-based method Reachability analysis	Uncertain destination models	Level [4,5]
Bock et al. [123]	Dataset in [123]	LSTM	Event/activity models Deep learning methods	level 5
Hug et al. [163]	Synthetic test conditions	LSTM with a mixture density output layer (LSTM-MDL) model and particle filter method	Uncertain destination models Deep learning methods	Level [4,5]
Cheng et al. [135]	Pedestrian datasets: ETH [197] and UCY [179]	Social-Grid LSTM based on RNN architecture	Uncertain destination models Deep learning methods	Level 5
Bhattacharyya et al. [122]	CityScapes dataset [136]	Two-stream recurrent neural network (RNN)	Uncertain destination models Deep learning methods	Level [4,5]
Broz et al. [127]	Simulated data	Time-state aggregated partially observable Markov model (POMDP)	Two agents' interaction	Level 4
Rudenko et al. [210]	Simulated and real data	Markov decision process (MDP) with a joint random walk stochastic policy sampling	Two agents' interaction Graphic dynamical models	Level 5
Kretzschmar et al. [175]	Turing test with human participants	Markov chain Monte Carlo (MCMC) sampling	Two agents' interaction Graphic dynamical models	Level [4,5]
Kawamoto et al. [170]	Pedestrian datasets: ETH [197] and [117]	Kriging (Gaussian process) model	Two agents' interaction Gaussian Process methods	Level [3,4]

Alahi et al. [111]	Pedestrian datasets: ETH [197] and UCY [179]	Social LSTM	Two agents' interaction Deep learning methods	Level [4,5]
Hoogendoorn et al. [161]	Simulations	Optimal control theory	Two agents' interaction	Level [4,5]
Ikeda et al. [164]	Shopping mall data	Social force and sub-goal concept	Two agents' interaction	Level [3,4]
Chen et al. [134]	Experimental vehicle ALSVIN	Extended Kalman filter (EKF)	Two agents' interaction Graphic dynamical models	Level [3,4]
Bera et al. [121] [120]	Indoor and outdoor crowded videos	Ensemble Kalman filter (EnKF)	Group interaction Graphic dynamical models	Level [4,5]
Deo et al. [140] [228]	Crowded unsignalized intersection dataset	Variational Gaussian mixture models (VGMM)	Group interaction Graphic dynamical models	Level 5
Pellegrini et al. [197] [198]	Pedestrian dataset with birds-eye view images [197]	Linear Trajectory avoidance model (LTA)	Small Group interaction Graphic dynamical models	Level [4,5]
Sun et al. [217]	L-CAS Pedestrian dataset [217]	Temporal 3DOF-pose LSTM (T-pose LSTM)	Group interaction Deep learning methods	Level [4,5]
Yi et al. [235]	Crowded scenes video data	Behaviour convolutional neural network (CNN)	Group interaction Deep learning methods	Level [4,4]
Radwan et al. [203]	6 public datasets comprising ETH, UCY, L-CAS	Interaction-aware trajectory convolutional neural network (IA-TCNN)	Group interaction Deep learning methods	Level [4,5]
Moussaid et al. [193]	Pedestrian trajectories [193]	Heuristic model	Group interaction	Level 5
Turner et al. [220]	Simulations	Exosomatic visual architecture	Group interaction	Level [4,5]
Vasishta et al. [225]	Real world scenes dataset [225]	Natural vision model	Group interaction	Level [4,5]
Zhou et al. [239]	Pedestrian dataset from New York Central station	Mixture model of dynamic pedestrian-agents (MDA)	Group interaction Graphic dynamical models	Level [4,5]
Shi et al. [215]	2D laser sensor dataset	LSTM	Group interaction Deep learning	Level[4,5]
Amirian et al. [113]	Synthetic dataset	GAN based method with hand-desgined interaction features	Group interaction Deep learning	Level [4,5]

Lee et al. [178]	KITTI [151][152] and Stanford Drone Dataset [209]	GAN and RNNs	Group interaction Deep learning	Level[4,5]
Gupta et al. [156]	ETH and UCY datasets	GAN and RNNs	Group interaction Deep learning	Level[4,5]
Sadeghian et al. [211]	ETH and UCY	GAN	Group interaction Deep learning	Level[4,5]
Henry et al. [160]	Crowd flow simulator	Inverse reinforcement learning (IRL) and Gaussian process (GP)	Crowd behaviour models	Level 5
Trautman et al. [219]	Pedestrian dataset: ETH [197]	Gaussian process (GP)	Group interaction Gaussian Process methods	Level 5
Ali et al. [112]	Video from Google videos and National Geographic documentary	Lagrangian particle dynamics model	Macroscopic models Crowd behaviour models	Level 5
Mehran et al. [189]	Dataset of escape panic scenarios and web videos	Particle advection with social force model	Macroscopic models Crowd behaviour models	Level 5
Ma et al. [184]	UCY Zara Dataset, the Town Centre Dataset and the LIDAR Trajectory Dataset.	Fictitious game and reinforcement learning	Game theoretic models Two agents' interaction	Level 5
Isaacs [166]	/	Homicidal taxi driver problem	Game theoretic models Two agents' interaction	Level 5
Turnwald et al. [221] [222] [188]	Pedestrian trajectories from motion capture system [221]	Finite set of single-shot games	Game theoretic models Two agents' interaction	Level 5
Fox et al. [148] [129]	Simulations and dataset in [131]	Game of Chicken	Game theoretic models Two agents' interaction	Level 5
Vascon et al. [224]	Public datasets	Game theory	Game theoretic models Small group interaction	Level 5
Johora et al. [167]	Simulations	Stackelberg games	Game theoretic models Small group interaction	Level 5
Mesmer et al. [190]	Experiments	Game theory with velocity vector	Game theoretic models Crowd interaction	Level 5
Shi et al. [214]	Experiments	Modified lattice model	Game theoretic models Crowd interaction	Level 5
Dimitris et al. [195]	Video data	Two-dimensional classification model	Signalling models	Level 5
Katz et al. [169]	Controlled experiment	Statistical analysis	Signalling models	Level 5

Guéguen et al. [158]	Controlled experiment	Statistical analysis	Signalling models	Level 5

3 Datasets

Table B.2: Summary of pedestrian datasets

Dataset	**Data type**	**Viewpoint**	**Applications**	**Quantity of data**
The Caltech Pedestrian Benchmark [141]	Urban video data (Resolution: 640x480)	Moving car	Detection, Tracking, Trajectory Prediction	10 hours of 30Hz video with 250,000 annotated frames, 350,000 labeled bounding boxes and 2300 unique pedestrians
ETHZ Benchmark [146]	Urban video data using a stereo pair of cameras (Resolution: 640x480)	Children's stroller	Detection, tracking, Trajectory prediction	2,293 frames with 10958 annotations
TUD-Brussels [229]	Image pairs	Hand-held camera and Moving car	Detection	Training set: 1,092 positive image pairs (resolution: 720x576) with 1,776 annotations and 192 negative image pairs (resolution: 720x576). Additional 26 image pairs with 183 annotations Test set: 508 image pairs (resolution: 640x480) with 1,326 annotations
Daimler Benchmark [145]	Grayscale camera images	Moving car	Detection	Training set: 15,660 positive samples (resolution: 72 pixels height) and 6,744 negative samples Test set: 21,790 images with 56,492 annotations including 259 trajectories of fully visible pedestrians
INRIA Pedestrian Dataset [138]	Camera images	Any	Detection	1805 images (resolution: 64x128)
CityPersons [236]	Camera images	Moving car	Detection	5k images with 35k bounding boxes of pedestrians and 20k unique persons

Edinburgh Informatics Forum pedestrians overhead dataset [187]	Pedestrian Trajectories	Surveillance camera	Trajectory prediction	Over 92k pedestrian trajectories
ETHZ BIWI Walking Pedestrian dataset [197]	Video	Bird-eye view	Detection, Tracking and Trajectory prediction	650 tracks over 25 minutes
UCY Zara pedestrian dataset [179]	Synthesized crowd data	Bird-eye view	Tracking and Trajectory Prediction	1 video 2-min long with 5-6 persons per frame 1 video with 40 persons per frame Trajectories
Town Center Dataset [117]	Video data	Bird-eye view	Detection, Tracking and Trajectory Prediction	Video (resolution: 1920x1080) with 71500 annotations
MARKET-1501 [237]	Camera images	Moving car	Detection and Re-identification	32k bounding boxes with 1,501 individuals and 500k non-pedestrian (street windows)
VIPER Benchmark [207]	Video data	Moving car	Optical flow, semantic instance segmentation, object detection and tracking, object-level 3D scene layout, visual odometry	250k video frames
CUHK01 [181]	Camera images	Surveillance camera	Detection and Re-Identification	971 persons with 2 camera views
CUHK02 [180]	Camera images	Surveillance camera	Detection and Re-Identification	1,816 unique persons with 5 pair of camera views
CUHK03 [182]	Camera images	Surveillance camera	Detection and Re-Identification	13k images with 1,360 pedestrians
DUKEMTMC dataset [208]	Video and trajectories	Surveillance camera	Detection, Tracking, Trajectory Prediction, Re-Identification	6,791 trajectories for 2,834 unique persons over 85 minutes video per camera (8 cameras in total) (resolution: 1080p)
DUKEMTMC-reID dataset [238]	Video and bounding boxes	Surveillance camera	Detection and Re-Identification	Over 36k bounding boxes of 1,812 unique individuals
MOTChallenge Benchmark [177] [191]	Video	Any	Detection, Tracking, Re-Identification, Trajectory Prediction	Composed of parts of other datasets and new data (videos, bounding boxes) website: motchallenge.net/
Daimler Pedestrian Benchmark [172]	Annotations	Moving car	Trajectory prediction [crossing, stopping]	58 annotated pedestrian-vehicle interactions data
PETA dataset [139]	Images	Any	Detection and Recognition	19,000 images with 8,705 persons

L-CAS 3DOF Pedestrian Trajectory Prediction Dataset [217]	Pedestrian trajectories	Mobile robot	Trajectory prediction	50 trajectories
L-CAS 3d-point-cloud-people-dataset [232]	3D LiDAR point clouds	Mobile robot	Pedestrian detection and tracking	5,492 annotated frames with 6,140 unique persons and 3,054 groups of peopleâ€<
IAS-Lab People Tracking dataset [194]	RGB-D video sequences + ground truth given by a motion capture system	Mobile Pioneer P3AT robot	People detection and tracking	4,671 frames with 12,272 persons
Porch experiment dataset [115]	Motion capture system	Any	Trajectory prediction	1,500 person trajectories
UCLA Pedestrian dataset [168]	Video data	Moving car	Trajectory Prediction	17 annotated video sequences, ranging from 30 to 900 frames, and containing 67 pedestrian trajectories
LIDAR Trajectory dataset [184]	Lidar data	Top view	Trajectory Prediction	20 interacting person trajectories
Joint Attention in Autonomous driving (JAAD) [205]	Videos and annotations	Moving car	Detection, Tracking, Trajectory Prediction	346 video clips with annotations extracted from 240 hours of driving videos
MoCap database [216]	Motion capture system	Any	Detection, Recognition	500k frames with persons in many different poses
The Multi-Person PoseTrack Dataset [165]	Video data	Any	Detection, Recognition, Tracking	60 videos with 16k annotated persons with different poses
CMU dataset [171]	Video data	Any	Detection, Tracking, Trajectory Prediction	92 videos
Pedestrian-Vehicle Interactions dataset [131] [130]	Annotations	Human observers	Trajectory Prediction (crossing, stopping)	204 annotated pedestrian-vehicle interactions at an unsignalized intersection
CITR adn DUT datasets [233]	Trajectories	Top view	Trajectory prediction and interaction	over 2k
NERC Agricultural Person Detection Dataset [199]	stereo video	Mobile platforms	Detection and Tracking	76k labelled person images and 19k person-free images
Action Recognition Dataset [149]	Stereo camera and thermal images + Lidar point clouds	Mobile robot	Action and gesture recognition	10 actors performing 9 gestures and 4 acivities

Table B.3: Summary of vehicle datasets

Dataset	Data type	Applications	Quantity of data
Berkeley DeepDrive Video dataset (BDDV) [231]	Video and GPS/IMU data	Detection, Tracking, Identification	10k hours of driving videos around the world
EPFL multiview car database [196]	Images	Detection, Identification	2k images with 20 different car models
KITTI dataset [151] [152]	Video data	Detection, Tracking, Identification, Localisation	About 1 hour in one city in daytime
Cityscape [136]	Video data	Detection, Tracking, Identification	About 100 hours videos in multiple cities in daytime
Commai.ai [234]	Video data	Detection, Tracking, Identification	7.3 hours videos in highway during daytime and night
The Oxford RobotCar Dataset [185]	Video data	Detection, Tracking, Identification	214 hours videos in Oxford in daytime
Princeton TORCS DeepDriving [133]	Synthetic video data	Detection, Tracking, Identification	13.5 hours videos in highways
Honda Research Institute Driving Dataset (HDD) [204]	Video data	Detection, Tracking, Identification	104 hours videos in one city
Udacity [223]	Video data	Detection, Tracking, Identification	8 hours of videos
nuScenes [128]	RGB, LIDAR and RADAR data	Scene understanding	1k different scenes form 2 cities
Fieldsafe dataset [174]	Multiple sensors	Obstacle detection	2h raw sensor data from a mobile platform

Table B.4: Summary of pedestrian and vehicle simulators

Simulator	Type	Applications
Technical University of Munich (TUM)	• Pedestrian simulator: Head-mounted display with a motion capture system • Driving Simulator software	• Pedestrian behaviour understanding • Driver behaviour analysis

Institute for Transport Studies (ITS), University of Leeds	• HIKER Lab : pedestrian simulator • Driving Simulator • Truck Simulator	• Pedestrian behaviour understanding • Pedestrian interaction with the environment • Driver behaviour understanding
Japan Automobile Research Institute (JARI)	• JARI-ARV (Augmented Reality Vehicle) • JARI-OVDS (Omni Directional View Driving Simulator)	• Road running driving simulator • Driving simulator with 360-degree spherical screen and rocking device
French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR)	• Driving Simulator • Immersive Simulator • Driving Simulator with human assistive devices • Bicycle Simulator	• Driver behaviour analysis • Road user behaviour understanding
University of Iowa	Driving Simulator	Driver behaviour understanding
Pedsim [153]	Synthetic simulator	Crowd behaviour understanding
OnFoot [186]	VR pedestrian simulator	Pedestrian behaviour understanding
Vehicle-Pedestrian simulators [147] [183]	Cellular automata models	Vehicle-pedestrian interactions
Macroscopic traffic simulator [132]	Force model	Traffic and road user behaviour understanding
AV-Pedestrian negotiations simulator [157]	Different pedestrian behaviour models	Pedestrian behaviour understanding for AVs