DEVELOPMENT OF A MULTI-AGENT BASED SIMULATION MODEL FOR CYCLIST-PEDESTRIAN INTERACTIONS IN SHARED SPACES

by

Rushdi Mah'd Alsaleh

B.Sc., Kuwait University, 2013

M.Sc., Kuwait University, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2021

© Rushdi Mah'd Alsaleh, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Development of a Multi-Agent Based Simulation Model for Cyclist-Pedestrian Interactions in Shared Spaces

submitted by Rushdi Mah'd Alsaleh in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Civil Engineering

Examining Committee:

Prof. Tarek Sayed, Professor, Department of Civil Engineering, UBC
Supervisor

Prof. Omar Swei, Assistant Professor, Department of Civil Engineering, UBC
Supervisory Committee Member

Prof. Amy Kim, Associate Professor, Department of Civil Engineering, UBC
University Examiner

Prof. Peter Cripton, Professor, School of Biomedical Engineering, UBC
University Examiner

Prof. Salvatore Cafiso, Professor, Department of Civil Engineering, University of Catania
External Examiner

Additional Supervisory Committee Members:

Prof. Ziad Shawwash, Assistant Professor, Department of Civil Engineering, UBC
Supervisory Committee Member

Abstract

Understanding and modeling cyclist and pedestrian dynamics and their microscopic interaction behaviour in shared space facilities are crucial for several applications, including safety and performance evaluations. Recently, a few studies have developed models to simulate road user interactions in shared space facilities. However, existing models suffer from several shortcomings and show significant discrepancies with real-world behaviour. As such, this thesis presents a novel microsimulation-oriented framework for modeling cyclist and pedestrian interactions in such facilities. Advanced Artificial-Intelligent techniques were used to model road users' behaviour and their interactions as reward-based intelligent agents. This thesis bridges the gap in modeling road user interactions by accounting for their rationality, intelligence, and sequential decision-making process by implementing the Markov Decision Process (MDP) modeling framework. Furthermore, this thesis proposes a multi-agent modeling framework to model cyclist and pedestrian interactions in shared spaces. Unlike the traditional game-theoretic framework that models multi-agent systems as a single time-step payoff, the proposed approach is based on Markov Games (MG), which models road users' sequential decisions concurrently. Moreover, this thesis investigates the ability of different equilibrium behavioral theories (i.e., Nash-Equilibrium (NE) and Logistic-Stochastic-Best-Response-Equilibrium (LSBRE)) in predicting road user operational-level decisions and evasive-action mechanisms. Road user trajectories from three shared space facilities located in Vancouver, Canada, and New York City, USA, were extracted by means of computer vision algorithms. Single and multi-agent inverse reinforcement learning approaches were utilized to estimate road user reward functions using examples of their demonstration (i.e., trajectories). Reward function weights infer road users' goals and preferences and can form the key component in developing agent-based

microsimulation models. Single-agent and multi-agent simulation platforms were developed, relying on deep reinforcement learning approaches, to emulate and validate road user interactions in shared spaces. The utilized multi-agent modeling approaches led to a significantly more accurate prediction of road user behaviour and their evasive action mechanisms. Moreover, the recovered reward functions based on the single-agent modeling approach failed to capture the equilibrium solution concept compared to the multi-agent approach. This thesis determines a behavior-based consistent paradigm to model equilibrium in multi-agent transportation systems, such as road user interactions in shared space facilities.

Lay Summary

Shared space facilities have recently emerged as a strategy to promote active modes of transportation and supports cities' sustainability goals. The shared space is a unique approach of urban design that supports pedestrian and cyclist movements by reducing the dominance of motorized vehicles. However, the majority of these facilities were constructed with little prior evaluation for their operation, performance, or safety. This is likely due to the lack of available tools for such evaluation as a result of the more complex interaction behaviour and the shared right of way. The majority of the existing road user behaviour models relies on unrealistic methodologies and is based on the single-agent modeling framework. However, this assumption is unrealistic and could limit the models' accuracy. This thesis presents several advances toward modeling road user interactions in shared spaces by considering their rationality, intelligence, sequential decision-making process, and multi-agent interaction nature using advanced artificial-intelligence techniques.

Preface

The thesis chapters are published as 4 papers in reputable journals and 1 paper as a presentation at a refereed conference. In addition, 1 paper is under review in a prominent journal. These are as follows:

Journal Papers

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, and a version of chapter 6 have been published in:

- **Alsaleh, R., & Sayed, T.** (2021). Markov-game modeling of cyclist-pedestrian interactions in shared spaces: A multi-agent adversarial inverse reinforcement learning approach. ***Transportation Research Part C: Emerging Technologies***, 128, 103191.

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T.

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, and a version of chapter 5 have been published in:

- **Alsaleh, R., & Sayed, T.** (2021). Microscopic modeling of cyclists interactions with pedestrians in shared spaces: a Gaussian process inverse reinforcement learning approach. ***Transportmetrica A: Transport Science***, 1-27.

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T.

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, portions of chapter 3, and a version of chapter 4 have been published in:

- **Alsaleh, R., & Sayed, T.** (2020). Modeling pedestrian-cyclist interactions in shared space using inverse reinforcement learning. ***Transportation research part F: traffic psychology and behaviour***, 70, 37-57.

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T.

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, and a version of chapter 3 have been published in:

- **Alsaleh, R., Hussein, M., & Sayed, T.** (2020). Microscopic behavioural analysis of cyclist and pedestrian interactions in shared spaces. ***Canadian Journal of Civil Engineering***, 47(1), 50-62.

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T., and Dr. Hussein. M.

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, and a version of chapter 7 are under review in a prominent journal as follows:

- "Do Road Users Play Nash Equilibrium? A Comparison between Nash and Logistic Stochastic Equilibriums for Multiagent Modeling of Road User Interactions in Shared Spaces."

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T.

Conference Papers

Portions of the introductory text in chapter 1, portions of the literature review in chapter 2, and portions of chapter 3 have been presented as follows:

- **Alsaleh, R., Hussein, M., & Sayed, T.** (2019). Microscopic Interaction Behavioral Analysis Between Cyclists and Pedestrians in Shared Space Area: A Study Based on Automated Video Analysis. ***In The Proceedings of Transportation Research Board 98th Annual Meeting, Washington DC***.

I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing, and manuscript composition. This work was all under the supervision of Prof. Sayed, T. and Dr. Hussein. M.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A2C	synchronous Advantage Actor-Critic
ABM	Agent-Based Modeling
ACKTR	Actor-Critic with Kronecker factors Trust Region
AI	Artificial Intelligent
CA	Cellular Automata
CE	Correlated Equilibrium
CV	Computer Vision
DRL	Deep Reinforcement Learning
FM	Feature Matching
GD	Gaussian Distribution
GP	Gaussian Process
GPIRL	Gaussian Process Inverse Reinforcement Learning
HausD	Hausdorff Distance
HCM	Highway Capacity Manual
IEBCA	Improved Extended Burgers Cellular Automata
IRL	Inverse Reinforcement Learning
K-FAC	Kronecker-Factored Approximate Curvature
KLT	Kanade–Lucas–Tomasi
LG	Lattice Gas
LQRE	Logistic Quantal Response Equilibrium
LSBRE	Logistic Stochastic Best Response Equilibrium
MA	Multi-Agent

MA-AIRL	Multi-Agent Adversarial Inverse Reinforcement Learning
MACK	Multi-agent Actor-Critic with Kronecker factors
MA-DRL	Multi-Agent Deep Reinforcement Learning
MAE	Mean Absolute Error
MA-GAIL	Multi-Agent Generative Adversarial Imitation Learning
MAIRL	Multi-Agent Inverse Reinforcement Learning
MARF	Multi-Agent Reward Function
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Process
ME	Maximum Entropy
MG	Markov Game
NE	Nash Equilibrium
OECD	Organization for Economic Co-operation and Development
PMV	Personal Mobility Vehicle
RF	Reward Function
RL	Reinforcement Learning
SFM	Social Force Model
SPNE	Sub-game Perfect Nash Equilibrium
TD	Temporal Difference

Acknowledgements

I would like to express my enduring gratitude to my advisor, Prof. Tarek Sayed, for his valuable guidance, encouragement, continuous support, and providing me with an excellent atmosphere for doing research. He has been a tremendous mentor from whom I have learned a lot during my work at UBC. I will always feel fortunate that I had him as my supervisor.

I would like to thank my supervisory committee for their positive feedback and valuable suggestions for my thesis. I would also like to thank the professors in the Civil Engineering Department at UBC who impacted my graduate study throughout the coursework. I would like to thank all my fellow colleagues in the *BITSAFS lab* from whom I have learned a lot.

Finally, but foremost, I am deeply grateful to my parents, my brothers, and my sisters for their continuous support and sacrifice throughout this long journey.

Dedication

To my parents and family

Chapter 1: Introduction

This chapter provides a general introduction to the thesis, and it consists of four sections. The first section presents background information that describes the importance of the research problem. Then, the existing research gaps and issues are discussed in the next section. The third section states the research objectives and goals. Lastly, the chapter concludes with an outline for the thesis structure.

1.1 Background

Surface transportation is witnessing major challenges worldwide, particularly in urban areas. Road traffic congestion and accidents are considered significant challenges for transport policy-makers and urban governance. Many urban streets worldwide are witnessing enormous growth in motor vehicle traffic and its infrastructure, resulting in growing traffic congestion and safety concerns. The social, economic, health and environmental impacts of traffic congestion and accidents are enormous. The health impact cost of air pollution attributed to road transport was estimated at about US $850 billion annually for the OECD[1] countries (OECD, 2014). Moreover, road traffic accidents have become the eighth leading cause of death globally and the leading cause of death among children and young adults aged 5–29 years. The fatalities and injuries resulting from road collisions form a global epidemic, causing 1.35 million losses of lives and 20

[1] OECD refers to the Organization for Economic Co-operation and Development, and it includes 34 countries (by 2014), including Canada, Australia, United States, and many European countries.

to 50 million injuries with many incurring disabilities annually (World Health Organization, 2018). The annual estimated social and economic losses of traffic injuries are tremendous and estimated at about US $518 billion globally (Jacobs, et al., 2000; World Health Organization, 2004).

The concept of sustainable transportation, particularly in urban areas, has recently gained considerable interest from the public and policy-makers. Active transportation modes such as walking and cycling have a vital role in supporting cities' sustainability goals and reducing traffic-induced air pollution and traffic congestion. Active modes of transportation help road users to adopt a healthy lifestyle and increase their physical activity level. Increasing the population's physical activity level minimizes the risk of several medical conditions, including cardiovascular diseases, diabetes, and obesity. Globally, physical inactivity is attributed to 3.2 million deaths annually (World Health Organization, 2009). The shift to such active transportation modes reduces traffic accident risks and casualties and imposes economic benefits (Rabl & De Nazelle, 2012; Campbell & Wittgens, 2004). In Canada, the economic benefit of shifting to active transportation modes, at a mode share level as low as 15%, is estimated at about $7 billion annually (Campbell & Wittgens, 2004).

Active transportation modes such as walking and cycling suffer from an elevated collision risk and discomfort. This is attributed to the higher vulnerability of pedestrians and cyclists, where they are at higher risk of physical damages and severe injuries when involved in road collisions compared to vehicular drivers. This may discourage using active transportation modes as they could be perceived as less safe than driving (Lawson, et al., 2013). Although active

transportation modes are usually involved in a low number of accidents, they account for most of the fatalities. Globally, around a quarter (26%) of all road traffic deaths are among the less protected transportation modes (i.e., pedestrians and cyclists) (World Health Organization, 2018). Locally, in the City of Vancouver, active transportation modes (cycling and walking) account for about 2% of all road collisions involving motor vehicles. However, they represent about 65% of the traffic-related fatalities between the years 2007 and 2012 (City of Vancouver, 2015).

Many cities and ministries worldwide have recognized the vital role of active transportation modes and have been adopting policies that aim at promoting these modes while increasing their safety and comfort levels. For example, the City of Vancouver adopted the "Transportation 2040" plan, which presents a long-term strategic vision for the city (City of Vancouver, 2012). The plan includes targets to increase the proportion of active transportation mode trips, including walking, cycling, and transit, to be two-thirds of all trips and achieving zero traffic-related fatalities by 2040. The plan gives special emphasis to vulnerable road users by creating a safe and convenient active transportation network.

Moreover, many cities have adopted policies that involve redesigning selected streets or public places into non-motorized shared space facilities for social and recreational activities. The shared space design paradigm has received considerable attention as an alternative approach to the classic design of streets. Shared space programs have been implemented in many cities worldwide, including Vancouver, Calgary, Vienna, Auckland, and many other cities in Germany and the Netherlands. Furthermore, the "Open Streets" program presents another policy that

encourages active road users of all ages and abilities to share the road. This program temporarily repurposes specific cities' roads into motorized-free areas for active transportation modes. By 2016, around 122 cities in the United States have hosted this program, including New York and Los Angeles with more than 100 thousand participants per event (Hipp, et al., 2017).

Although many studies demonstrated the benefits of walking and cycling, active transportation modes' research area is generally less studied and developed compared to vehicular traffic. Moreover, despite that many cities have implemented shared space facilities worldwide, only a few studies have focused on understanding, analyzing, and modeling multimodal active road users' movement and interaction behaviour in such facilities. The majority of these facilities were constructed with little prior evaluation for their operation, performance, or safety (Karndacharuk, et al., 2016; Chong, et al., 2010). This is likely due to the lack of available data and tools for such evaluation as a result of the more complex interaction behaviour and the shared right of way. Unlike conventional roads, shared space regulations allow road users to freely navigate the whole area without being restricted to follow predefined paths (e.g., sidewalk or bike lanes). Active road users' behaviour analysis and modeling require collecting a considerable amount of their behavioural data and extracting their microscopic behavioural elements. However, manual collection of cyclist and pedestrian behavioural data and extraction of their microscopic behavioural elements are extremely costly and challenging due to their complex interaction behaviour.

Over the past decades, several simulation models have been developed to address several active road user applications, including designing active road user facilities and planning active road

user routes. However, despite the relative success of the existing models in addressing road user behaviour, they are suffering from several shortcomings that limit their applications, especially in shared space facilities. Firstly, the majority of the existing microsimulation models were developed to model a single mode of transportation, e.g., vehicular traffic (Wiedemann, 1974; Gipps, 1981), pedestrian flow (Helbing & Molnar, 1995; Burstedde, et al., 2001) or cyclist flow (Jiang, et al., 2004; Liang, et al., 2012). Specifically, these models are concerned with road/ path segments with physical separations between different transportation modes, making them challenging to model heterogeneous road users (i.e., cyclists and pedestrians) in shared space facilities.

Secondly, most of the developed microsimulation models for active road users rely on unrealistic methodologies. Such models ignore the fundamentals of road user behaviour, including their intelligence and rationality (i.e., their ability to make rational decisions based on their surrounding environments, experiences, and characteristics). For example, in the well-known road user models such as the Social Force Model (SFM) (Helbing & Molnar, 1995), road users are modeled as particles that can attract or repel each other by physical forces. Moreover, in the cellular automata (CA) model (Nagel & Schreckenberg, 1992; Jiang, et al., 2004), road users' movement is controlled by transition rules, which are based mainly on the probabilities of selecting the target cells. A road user movement in the CA model is usually limited by specified cell size, making it hard to capture the heterogeneity in road user behaviour. The assumptions in the SFM and CA models clearly neglect road users' rationality and their ability to think and make logical decisions based on their surroundings and goals. In particular, these models simulate road user trajectories without relying on explicit rules or utility functions that consider their objectives

and goals. Thus, these models are challenging to predict road user interactions accurately and infer their intentions and preferences in shared space facilities.

Thirdly, most of the developed road users' interaction models are based on the single-agent modeling framework. In this framework, a single road user's operational level actions (e.g., acceleration) are considered for modeling while assuming the other interacting road users' actions are known over time (i.e., part of the passive environment). However, this assumption is unrealistic and could limit the models' accuracy and transferability in non-stationary road user environments (e.g., multi-agent environments). Moreover, road user interactions are better represented as multi-agent interactions. In the multi-agent modeling framework, the actions of all interacting agents are modeled. Moreover, agents can be cooperative or competitive and can coordinate their actions to achieve specific goals (e.g., avoid collisions). Lastly, the existing calibration and validation approaches of active road users' simulation models are mainly based on macro-level evaluation by comparing the model results with macro-level Measures-Of-Effectiveness (MOE), such as average speed and traffic flow. Although macro-level validation is important, it does not guarantee realistic and accurate behavior at a micro-level, which could limit the applications of the model. Therefore, there is a need to develop cyclist and pedestrian microsimulation simulation models that overcome the shortcomings of the existing models. The models should accurately reproduce active road users' trajectories, which present their microscopic behaviour during their interactions in shared space facilities. Such models would be beneficial for understanding and analyzing cyclist and pedestrian dynamics and their microscopic interaction behaviour in shared space facilities. Moreover, the models would be

beneficial in several applications that require accurate prediction of road user trajectories, such as safety evaluation and predictions of traffic conflict indicators.

1.2 Problem Statement

The study of active transportation modes in shared space facilities draws its importance from the public and political awareness of environmental challenges and the vital role cycling and walking play in creating a sustainable transportation system. Despite this awareness, considerable efforts are still needed to understand, plan and design safe and convenient non-motorized shared space facilities. Although some studies have investigated active road users' behaviour and developed models to simulate their interactions in shared space facilities, these studies suffer from some limitations, and several research gaps still need to be addressed. This thesis tackles several research problems and presents novel developments in investigating, understanding, and modeling active road user dynamics and their interactions in shared space facilities. This thesis covers the following research problems:

1.2.1 Problem One: Microscopic Analysis of Cyclist and Pedestrian Interactions

The analysis of the behaviour of cyclists and pedestrians in shared space facilities has been gaining increased interest in the literature in recent years. The majority of the studies focused on the macroscopic characteristics of cyclist and pedestrian interactions in such facilities (e.g., the relationship between road user speed and flow rate). However, little consideration is paid for their interaction behaviour at the microscopic level. This is attributed to the complexity of analyzing the microscopic behaviour of road user interactions in shared space facilities and the shared right of way. Building a simulation model based solely on the macroscopic results does

not guarantee that the model can produce accurate predictions of road user trajectories at the microscopic level, which could limit the applications of the model. This emphasizes the need to investigate the various interaction types between cyclists and pedestrians in shared space facilities and conduct a detailed analysis of their interaction behavior at the microscopic level. Such interactions can be characterized based on road user micro-level parameters (e.g., speed, acceleration, longitudinal distance, and interaction angle profiles). These micro-level parameters can be extracted using state-of-the-art computer vision platforms.

1.2.2 Problem Two: Development of a Single-Agent Based Simulation Model for Cyclist and Pedestrian Interactions

The majority of the developed microsimulation models for active road users are mainly based on physical force models, e.g., social force model (SFM) (Helbing & Molnar, 1995), or rule-based models, e.g., cellular automata (CA) (Nagel & Schreckenberg, 1992). Although some studies extended these models to model road users' interactions in shared space facilities, these models still show significant discrepancies with real-world behaviour (Liu, et al., 2020). Moreover, most of the agent-based models developed to simulate road user interactions are based on ad-hoc or heuristic rules (Papadimitriou, et al., 2009) and are concerned with a single mode of transportation. However, agents (e.g., road users) can be prone to undesired behaviour due to rules or utility function miss-specification (Amodei, et al., 2016). Furthermore, the validation of most of the active road user models is based on the macro-level evaluation. This validation approach does not guarantee accurate predictions for microscopic behaviour. Validating model results based on the microscopic level would expand the applicability of the model in applications that require accurate road users' trajectories predictions, such as traffic safety

applications. Therefore, there is a need for developing a novel agent-based cyclist and pedestrian (i.e., multi-modal) interaction microsimulation model that overcomes the shortcoming of the existing models. The model should recover cyclist and pedestrian interaction utility functions based on real-world demonstrations by utilizing advanced Artificial Intelligent (AI) techniques. The model should incorporate detailed microscopic road user behaviours. The proposed model should be validated using real-world trajectories data relying on road user microscopic behavioral parameters. A shared space simulation platform should be developed to integrate the developed interaction model in order to simulate cyclist and pedestrian trajectories.

1.2.3 Problem Three: Selection of Utility Functional Form for Cyclist and Pedestrian Interactions

The aforementioned agent-based modeling approach can utilize various advanced AI techniques to recover various cyclist and pedestrian interaction utility functional forms. Implementing various algorithms requires different computational powers and may lead to varying prediction accuracies. Thus, different utility functional forms of cyclist and pedestrian interaction should be investigated. These functional forms include discrete and continuous functions (in terms of movement plan/ microscopic variables) as well as linear and nonlinear (e.g., Gaussian Process) functions. Therefore, there is a need to utilize various advanced AI techniques to recover various utility functional forms of cyclist and pedestrian interactions in shared space facilities. The recovered utility functions should be assessed in terms of the microscopic road user behaviour parameters prediction accuracy (i.e., road user trajectories) and the inferred interaction behaviour accuracy.

1.2.4 Problem Four: Development of a Multi-Agent Based Simulation Model for Cyclist and Pedestrian Interactions

The majority of the developed microsimulation models for active road users, e.g., (Dias, et al., 2018; Hussein & Sayed, 2017; Anvari, et al., 2015), are concerned with the single-agent modeling framework. However, many modeling problems like autonomous vehicles, robot swarms, and road user interactions are better represented as multi-agent interactions. In the single-agent modeling framework, the sequence of actions for only a single agent (e.g., road user) is considered for modeling while assuming the other interacting agents to be part of the passive environment. However, this assumption is unrealistic and may limit the accuracy of the estimated reward functions and thus the predicted trajectories. The use of the multi-agent modeling framework has several advantages. First, it presents a more realistic approach for modeling road user interactions and traffic conflicts due to its ability to capture social negotiation in road user interactions. Second, it can handle non-stationary environments and capture the equilibrium information compared to the single-agent modeling framework, making it suitable for modeling road user interactions in shared space facilities. Third, it can handle optimal mixed-strategy (i.e., probabilistic) policies, which are difficult to be handled in the single-agent modeling framework. Therefore, there is a need for developing a novel multi-agent microsimulation model for cyclist and pedestrian interactions in shared space facilities. The model should accounts for road users' sequential decisions concurrently. The model should incorporate detailed microscopic road users' behaviours and be validated based on their microscopic parameters. A multi-agent shared space simulation platform should be developed to integrate the developed multi-agent interaction model in order to simulate cyclist and pedestrian trajectories.

1.2.5 Problem Five: Selection of Equilibrium Solution Concept for the Multi-Agent Model

Another critical component for developing a multi-agent simulation model is selecting the appropriate equilibrium solution concept that can represent the nature of the multi-agent system being modeled. Unlike the single-agent modeling framework, each agent's optimal policy is influenced by the other interacting agents' policies in multi-agent system modeling. Thus, the optimality notion in the multi-agent system is defined by an equilibrium solution concept, such as the Nash Equilibrium (NE) (Hu & Wellman, 1998; Nash, 1951) and the Logistic Stochastic Best Response Equilibrium (LSBRE) (Yu, et al., 2019; McKelvey & Palfrey, 1998; McKelvey & Palfrey, 1995). Selecting the appropriate equilibrium solution in the multi-agent system modeling is vital in specifying the agent policies and model accuracy. For example, under the NE, agents are assumed perfectly rational and aware of all aspects of their environment (i.e., complete information). Thus, each agent's strategy (i.e., policy) is the best response to other agents' strategies (i.e., perfectly optimal agent behavior). However, under the LSBRE, agents' select strategies based on the relative expected utilities. Thus, better strategies are more likely to be selected than the worse ones, but the best strategies are not selected with certainty. The LSBRE can handle bounded rationality agents and sub-optimal agent behavior. Using the NE solution for bounded rationality agent systems could lead to obtaining nonsensical results (Harsanyi & Selten, 1988; Harsanyi, 1968). Despite the importance of investigating the appropriate equilibrium solution concept in modeling multi-agent cyclist and pedestrian interactions, none of the studies investigated this issue. Therefore, there is a need to determine the appropriate equilibrium solution concept for the multi-agent modeling of cyclist and

pedestrian interactions in shared space facilities. This would create a behaviour-based consistent paradigm to model equilibrium in multi-agent cyclist and pedestrian systems.

1.3 Research Objectives

The main goal of this thesis is to develop a novel microscopic simulation model for cyclist and pedestrian interactions in shared space facilities. The model should 1) produce accurate predictions of their behaviour (i.e., trajectories) in complex interactions, and 2) infer their behavioural preferences during their interactions in shared space facilities. The model is supported by the computer vision system that automatically extracts road user microscopic behavioural elements required for model development and validation. The following specific objectives are considered in this thesis:

1. Collecting cyclist and pedestrian video data in shared space facilities to create a video library required for conducting detailed cyclist and pedestrian interaction behaviour analysis. The video data should capture their natural movements and interactions while minimizing the risk of disturbing their behaviour. The video data would be used to extract the microscopic behaviour parameters required for model development and validation. The video data should cover a variety of cyclist and pedestrian interactions in shared space facilities.

2. Conducting a detailed microscopic behaviour analysis of cyclist and pedestrian interactions in shared space facilities. The analysis would rely on the video data for their interactions collected in the previous phase. Cyclist and pedestrian trajectories would be

extracted by means of the automated computer vision platform developed at the University of British Columbia. The extracted trajectories would be utilized to calculate elements of road users' behaviour, including their speed, acceleration, and relative distance profiles. Such extracted profiles represent road users' micro-level parameters that would be used in analyzing their interaction behavior. Furthermore, the study would identify a set of parameters that can be used to model such interaction behaviour in microsimulation models.

3. Developing novel microscopic simulation models for cyclist and pedestrian interactions in shared space facilities. The models should account for road users' intelligence and rationality and infer their preferences during their complex interactions in such facilities. The models should utilize the state-of-the-art agent-based modeling framework. The models should use advanced Artificial Intelligent (AI) techniques to recover road users' utility/ reward functions in their interactions in such facilities. The models should incorporate the microscopic parameters of road user behaviour in order to simulate trajectories at high accuracy. A simulation platform should be developed to integrate the developed interaction models in order to predict road user trajectories. The predicted road user trajectories should be evaluated against real-world data relying on the microscopic behaviour parameters.

4. Utilizing the advanced AI techniques to recover various cyclist and pedestrian interaction utility functional forms in shared space facilities. The modeling approach should investigate various utility functional forms, including discrete and continuous functions

as well as linear and nonlinear functions. The recovered utility functions would be assessed in terms of road user trajectories' prediction accuracy and the inferred interaction preferences.

5. Developing a novel multi-agent microsimulation model for cyclist and pedestrian interactions in shared space facilities. The model should account for road users' sequential decisions concurrently while considering their intelligence and rationality. Furthermore, the model should infer road users' preferences during their interactions in such facilities. The model should utilize the multi-agent-based modeling framework and use the advanced Artificial Intelligent (AI) techniques to recover road users' multi-agent utility/ reward functions. The detailed microscopic road users' behaviour parameters should be incorporated in the model in order to simulate their trajectories at high accuracy. The developed multi-agent interaction model should be integrated into a multi-agent simulation platform to predict road users' trajectories and evaluate their accuracy against real-world data.

6. Investigating the ability of various equilibrium solution concepts in explaining/ predicting cyclist and pedestrian behaviour during their interactions in multi-agent settings (i.e., shared space facilities) and creating a behaviour-based consistent paradigm for modeling equilibrium in such a system. The equilibrium solution concepts should be evaluated in terms of road user trajectories' prediction accuracy and the inferred interaction behaviour.

1.4 Thesis Structure

The thesis is divided into eight chapters. The first chapter presents an introduction to the thesis, background information related to the research, the research problem, and the main objectives of the thesis. Chapter two provides a detailed literature review related to the topics addressed in this thesis. Chapter three provides detailed information about the methodology utilized in this thesis, mainly the computer vision platform used to extract road user trajectories and the calculation of the microscopic elements of road user behaviour. The details of the microscopic behaviour analysis study of cyclist and pedestrian interactions in shared space facilities is presented in this chapter. Chapter four focuses on developing single-agent discrete agent-based cyclist simulation models for cyclist and pedestrian interactions in shared space facilities. Chapter five provides details on developing the single-agent continuous, linear and Gaussian Process nonlinear, agent-based cyclist simulation models for cyclist and pedestrian interactions in shared space facilities. Chapter six focuses on developing the multi-agent-based cyclist and pedestrian simulation model for their interactions in shared space facilities. Chapter seven investigates the impact of various equilibrium solution concepts on the multi-agent modeling of cyclist and pedestrian interactions in shared space facilities. Finally, chapter eight summarizes the key research findings and conclusion, the limitation, and the suggested future research directions.

Chapter 2: Literature Review

This chapter provides the background literature for the research covered in this thesis. The literature review focuses on the essential studies with a significant impact in the field of cyclist and pedestrian interaction modeling. Moreover, this chapter presents a critical review of previous studies that dealt with road user behaviour in shared space facilities. The literature review presented in this chapter covers four main topics. First, the chapter reviews the concept of shared space and provides a summary of previous studies that focused on its application and benefits. Second, a review is provided on the previous studies that explored cyclist and pedestrian behaviour in shared space facilities. Third, a comprehensive literature review is provided on the existing cyclist and pedestrian modeling methodologies and the various issues associated with them. Lastly, the chapter reviews a number of studies that are considered a milestone in the development of cyclist and pedestrian simulation models.

2.1 Shared Space Facilities

2.1.1 Shared Space Concept

In conventional traffic planning and engineering, the dominant *mobility* function of road designs is related to the efficiency of vehicular traffic flow. The other conflicting function in the conventional road hierarchies is *accessibility*, which describes the ability to access surrounding land-use activates. This function has an inverse relationship with the roads' mobility function (Karndacharuk, et al., 2014; AASHTO, 2011). Other functions such as environmental amenity and social aspects of pedestrians have been considered in conventional road designs as contributing factors to the overall function of abutting land use instead of creating a place for

social interaction and activities. Moreover, the *place* function is essential in urban streets, and it describes them as places where activities take place, such as loading/unloading areas. These three functions, mobility, accessibility, and place, are important for designing, using, and managing urban streets. The shared space design paradigm brings together these functions to better use urban road spaces (Karndacharuk, et al., 2014).

The shared space is an emerging concept of urban design that supports pedestrian and cyclist movements with slower vehicle speeds. This scheme of street design encourages the integration of road users and creates a pedestrian-friendly environment by reducing the segregation between road users and decreasing the dominance of motorized vehicles (Kaparias, et al., 2012). In shared space areas, there is no obvious segregation between road users so that all road users share the right of way. The access of motorized vehicles to the shared space can be prohibited to create a non-motorized shared space, which provides a safe and comfortable environment for non-motorized road users.

The application of the shared space concept in urban activity centers and the removal of traffic control devices, including the traffic signage, signals, pedestrian barriers, and road marking, impose several traffic operation challenges and safety concerns. The mixed road users in shared spaces and their freedom to move in the whole facility area without being restricted to predefine paths can lead to greater competing demand and more frequent road user interactions. However, most of these interactions occur at low design speeds and are associated with a lower risk of crash injury. In fact, the philosophy of the shared space design is based on the perceived risk, which necessitates road users to pay more attention to each other and react carefully

(Karndacharuk, et al., 2014). This concept is established based on the recent behavioral and environmental psychology findings and the evolution of the risk compensation theory (Adams, 1995; Hamilton-Baillie, 2008). The uncertainty of the right of way in shared spaces encourages road users to adopt a more cautious behaviour due to the perceived dangers (Adams, 2010). Such uncertainty of the right of way in shared spaces creates a shared and cooperative behaviour between road users.

2.1.2 Shared Space in Practice: Applications and Benefits

The development process of shared space facilities includes reducing vehicular traffic dominance while enhancing the active road user priority. Most of the shared space facilities incorporate a unique street environment for all road users, which is required to initiate uncertainty in their priorities and reduce the dominance of vehicular traffic. Such a design scheme involves removing curbs and the demarcation between various road users to encourage sharing behaviour. The mixed-use shared space design scheme has been implemented in several cities worldwide, including London UK, Auckland New Zealand, Gratz Austria, Bendigo Australia, Drachten Netherland, Vancouver Canada, among others (Karndacharuk, et al., 2014). Several previous studies documented the benefits of shared space facilities, including increased road users' safety and pedestrian activity levels. For example, Elliott Street, which is a one-way active road in the central business district of Auckland, New Zealand, has been transformed into a shared space facility. The design of the Elliot shared space is characterized by the continuous and level stone paving surface across the whole space area. The shared space has a minimum usage of traffic signage and marking with a 10 km/hr vehicular speed limit. The shard space has furniture for pedestrians, including seating and lighting. Before and after operation and safety evaluations

were conducted for the facility. The results show that vehicle flow drops approximately 40%, and the mean and 85th percentile of vehicle speed decreased by 20% to about 16 and 21 km/hr, respectively. Moreover, trajectories analysis shows that pedestrians move more freely within the facility. The safety evaluation shows that pedestrian-vehicle interactions were frequently observed at the facility with lower vehicle speed. However, no injuries were reported (Karndacharuk, et al., 2013; Karndacharuk, et al., 2014; Karndacharuk, et al., 2011; Karndacharuk, et al., 2014).

Moreover, the Sonnenfelsplatz complex roundabout in Graz, Austria has been converted into a shared space area. The design of the Sonnenfelsplatz shared spaced is characterized by the level surface with a minimum usage of traffic signage and marking. The shard space has furniture for pedestrians' occupancy and temporary trading activity. A before and after operation evaluation was conducted for the shared space facility. The results show that the shared space design has led to a lower average speed and a narrower speed distribution for all road users, which can be explained by the smoother movements and fewer stop and go conditions in the shared space area. The average vehicle speed at the shared space facility was around 15 km/hr. Moreover, the trajectory analysis shows that cyclists and pedestrians follow more direct paths across the shared space facility (Schonauer, et al., 2012; Rudloff, et al., 2013). Furthermore, the intersection of Hargreaves Street and Bull Street in Bendifo, Australia has been converted into a shared space facility. Similar to the aforementioned shared space facilities, the design of this facility is characterized by the removal of pedestrian barriers, pedestrian crossings, curbs, and traffic islands to create a level paved surface. Moreover, minimal traffic signs and markings were used, and a safe zone and street furniture for pedestrian occupancy were installed in the facility. In

addition, the design used 90-degree parking to slow vehicle speeds and reduce road widths, as well as stone rumble strip on entry ramps. A before and after traffic operation evaluation shows that vehicle flow drops approximately 30%, and the 85th percentile of vehicle speeds reduced by about 30% to 26-29 km/hr. Moreover, pedestrians were observed to move more freely in the facility with increased formal and informal activities after creating the shared space facilities (Department of Transport, Planning and Local Infrastructure, 2012; Government of South Australia, 2012).

Furthermore, the conversion of Kensington High Street, a busy route in London, to a shared space area significantly improved pedestrian safety (pedestrian collisions were reduced by about 64%) and increased pedestrian activity levels (Swinburne, 2005). Moreover, the conversion of a large five-way intersection in Oosterwolde, Netherlands, to a paved shared space area for all road users resulted in reductions in traffic speed and severe collisions despite the increase in traffic volume (Hamilton-Baillie, 2008). As well, the conversion of the Exhibition Road, a cultural center located in the Royal Borough of Kensington and Chelsea in London, to a shared space facility significantly reduced pedestrian-vehicle conflict rates, especially for slight (less-severe) conflicts (Kaparias, et al., 2013). The safety evaluation at the intersection of Park Avenue South and East 29th Street during the Open Street Program in New York City shows that pedestrian–bicycle and bicycle–bicycle conflict rates were significantly lower than the other normal operation condition, despite the increase in conflict frequencies (Hussein, et al., 2016).

Moreover, the benefits of shared space schemes include improving the urban realm attractiveness (Hamilton-Baillie, 2008; Hamilton-Baillie, 2008) and reducing traffic-related air pollution (Shu,

et al., 2016). Economic and urban revitalization and investment growth were clearly observed after remodeling the Grey's Monument area, Newcastle, into a public space area (Akkar, 2005). Despite the benefits of shared space areas, concerns were raised about the severity of interactions between non-motorized road users. Collisions between the vulnerable road users (cyclist-pedestrian collision) can lead to severe consequences, especially for pedestrians, who are more likely to suffer from severe injuries that can lead to death (Graw & König, 2002).

2.2 Road User Behaviour in Shared Space Facilities

Cyclist and pedestrian speeds are considered essential parameters of their flow, which are used to study their behaviour in several situations. Understanding and analyzing road user behaviour and developing tools to replicate their behaviour is essential for several applications in the transportation field, such as the design of shared space facilities and the safety and performance evaluations of such facilities. Such tools are beneficial in conducting virtual experiments to examine the impact of various layouts on traffic operations and evaluate the level of service under various demands. Earlier studies used manual approaches to extract road use behaviour in different situations. For example, Kiyota et al. (2000) (Kiyota, et al., 2000) studied the behaviour and the safety of pedestrians and cyclists in a non-motorized shared space in Japan. The study used the manual method to extract road use speeds. The study found that the speed of bicycles decreases as the pedestrian density increases. Moreover, the study conducted a survey to investigate the level of risk perception by shared space users under different circumstances. The study found that the perceived risk by pedestrians becomes higher as the passing distance between the cyclists and pedestrians decreases. Zacharias (1999) (Zacharias, 1999) studied pedestrian and cyclist behavior in the shared path of Leidsestraat street in Amsterdam. The study

used the manual method to extract road user behaviour. The study found that the distance between pedestrians decreases as pedestrian density increases. However, no significant change in the distance between cyclists and pedestrians was observed with changes in pedestrian density.

Moreover, Atkins (2012) (Atkins , 2012) conducted a behavioural study to investigate pedestrian and cyclist interactions on segregated and unsegregated shared facilities. The study used the manual method to extract road user speeds. The study conducted a survey to understand cyclist and pedestrian behaviour in segregated and unsegregated shared facilities. The study found that the average cyclist speed is not significantly higher on segregated road user routes compared with unsegregated routes, and the maximum cyclist speed decreases as pedestrian flow increases on shared space facilities. The cyclists usually adjust their cycling speed and moving direction when adopting a swerving maneuver around pedestrians. Moreover, the survey results show that both road users are generally comfortable using segregated and unsegregated shared facilities. Bernardi and Rupi (2015) (Bernardi & Rupi, 2015) studied cyclist behaviour on separated cycling facilities and mixed traffic conditions in Bologna, Italy. The study investigated the role of disturbances encountered in these facilities. The study used the manual method to extract road user speeds. The study found that pedestrian disturbance in the separated cycling facilities is more frequent and associated with moderate cyclist speed reduction. However, the disturbance in the mixed traffic is fewer and associated with more severe cyclist speed reduction. Kang et al. (2013) (Kang, et al., 2013) conducted a video-clip based survey to investigate pedestrians' perceived level of service in sidewalks shared with bicycles. The study developed random parameters ordered Probit model to investigate the effects of several contributing factors on pedestrians' perceived level of service. The study found that the presence of bicycles and the

higher bicycle speed decrease the pedestrians' perceived level of service. Moreover, the presence of a lateral separation (e.g., barrier) between the sidewalk and the motorized traffic increases their perceived level of service.

Furthermore, Kaparias et al. (2016) (Kaparias, et al., 2016) studied pedestrian gap acceptance behaviour before and after converting London's Exhibition Road into street layouts with elements of shared space. The study used the manual method to extract road use behaviour from the collected video data. The study found that pedestrians accept shorter gaps in traffic and are more at ease when crossing in shard spaces. Moreover, pedestrians were more comfortable and confident in their interaction with vehicles in such facilities. Hatfield and Prabhakharan (2016) (Hatfield & Prabhakharan, 2016) studied the behaviour and safety of cyclists and pedestrians on shared paths in Sydney, Australia. The characteristics of cyclist and pedestrian passing events and their incidents (e.g., near misses, crashes) were manually recorded in this study. A survey was conducted regarding cyclist and pedestrian beliefs, behaviour, and experience in shared paths. The study found that cyclists strongly prefer left-hand travel in shared spaces (on Australian roads), and their estimated cycling speed was found above 10 km/hr, which is the speed limit suggested considering pedestrian safety. However, cyclists closer to the centerline of the shared paths are associated with lower cycling speeds. Moreover, cyclists often pass pedestrians on the left, pass too close, and pass without slowing or warning. Cyclist and pedestrian near misses and collisions were reported at the shared paths and were mainly attributed to the road user distraction and violating the rules of thumb (i.e., keep on the left and overtake on the right).

Recently, automated video analysis techniques have been used to analyze road user behaviour. For example, Hussein et al. (2016) (Hussein, et al., 2016) studied the safety and behaviour of pedestrians and cyclists in the shared space of Park Avenue during the Summer Street program in New York City. The study utilized computer vision techniques to extract and analyze road user trajectories and their speeds and gait parameters. The study found that pedestrian walking speed was highest during the Summer Street operation, while it was lowest during the normal operation. However, cyclist speed was lowest during the Summer Street operation due to the frequent interaction with pedestrians, while higher cyclist speed was observed during the normal traffic operation. Moreover, the conflict rate between pedestrians and cyclists was found to be the lowest during the summer street operation. Beitel et al. (2018) (Beitel, et al., 2018) studied the behaviour and safety of pedestrians and cyclists in a non-motorized shared space area on the McGill University campus in Montreal, Canada. The study utilized computer vision techniques to extract road user trajectories. The study classified the conflicts between cyclists and pedestrians based on two criteria: the angle between their intersecting trajectories and who reaches first to the conflict point. The study found that bicycle speed decreases as pedestrian density increases. Moreover, the conflict rate between cyclists and pedestrians increases as pedestrian density increases.

Moreover, Shahhoseini and Sarvi (2019) (Shahhoseini & Sarvi, 2019) conducted a series of experiments to study the collective motion of pedestrian crowds in restricted shared space facilities. The study investigated the impact of shared space design and the movement speed of pedestrian crowds on the evacuation performance of the facilities. The experiments replicated the emergency escape of pedestrian crowds using different merging passes layouts. The study found

that the traffic efficiency (e.g., traffic flow) is impacted by the physical features of the facility layout and the desired speed level of the road users. Moreover, the trajectories analysis of the flows shows that the symmetric design layouts allow road user flows to keep their separate ways after entering the shared space. However, other layouts lead to a mixture of road user movement trajectories to allow the two flows to travel in the shared space facility.

2.3 Cyclist and Pedestrian Modeling Approaches

Microscopic simulation modeling of road users' dynamics has been advocated as a promising tool for analyzing their behavior in different facilities. Adopting a valid simulation model to analyze road user interactions in shared space areas provides a powerful tool that helps planners and engineers to assess the level of service and safety of such facilities. However, limited studies have investigated the development of microsimulation models of road users' behavior and their interactions in shared space facilities. The majority of road user simulation models are concerned with vehicular traffic and road/ path segments with physical separations between different transportation modes. Moreover, microsimulation models of active road users are less developed compared to vehicular traffic. Previous works dealing with the modeling and simulation of pedestrian and cyclist behaviour are mainly based on two approaches: the physical-based modeling approach and the cellular automata modeling approach.

In the physical-based modeling approach, road users are modeled as particles, and their interactions are modeled using physical forces. These forces represent the internal motivation of road users to perform certain movements due to the interactions with themselves or with the environment. Road user movement is updated in each simulation step based on the direction and

magnitude of the resultant force. For example, the social force model (SFM) developed by Helbing & Molnar (Helbing & Molnar, 1995) represents the first utilization of the physical-based modeling approach in addressing pedestrian movement dynamics. In this model, two forces control the movement of each pedestrian, the attraction and the repulsive forces. The attraction forces describe the desire of the pedestrian to reach a certain destination and the acceleration towards the desired velocity. However, the repulsive forces present the desire of a pedestrian to keep a certain distance from other pedestrians depending on pedestrian density and desired speed. These models are mainly developed for road segments with physical separations between different transportation modes.

The cellular automata (CA) modeling approach was initially proposed for simulating traffic flow and modeling transportation networks (Nagel & Schreckenberg, 1992). In this approach, the road user environment is spatially discretized into fixed-size cells, and the movement of road users is limited to the fixed cell size in each simulation step. Each cell can only be occupied by one road user each time step. In this approach, road users move from one cell to another based on predefined transition rules that mainly depend on the probability of choosing the target cell. The movement of the road user is recognized through the change of the cell's state (e.g., vacant or occupied) that the road user is occupying and the neighboring cells. The CA model was utilized to emulate pedestrian flow (Blue & Adler, 1998; Blue & Adler, 1999) and cyclist flow (Jiang, et al., 2004). The rules that control the movement of road users between the cells and the lane changing behaviour were developed in these studies (Blue & Adler, 1998; Blue & Adler, 1999; Jiang, et al., 2004). The developed pedestrian transition rules were adequate in presenting fundamental pedestrian flow as described in the Highway Capacity Manual (HCM) (Highway

Capacity Manual, 1985). Moreover, the developed bicycle transition rules were suitable to describe bicycle flow, and the obtained analytical results were in agreement with the simulation results.

Recently, a few studies have extended the SFM and CA models to simulate road user interactions in shared space facilities and mixed traffic conditions. However, these modeling approaches show significant discrepancies with real-world road user behaviour in shared spaces behaviour (Liu, et al., 2020). These classical approaches for modeling active road user dynamics and their microscopic interaction behaviour ignore the fundamentals of their behaviour and suffer from several shortcomings. The major challenges in these models are: 1) the reliance on unrealistic methodologies, 2) the lack of ability to consider the intelligence and rationality of road users, and their ability to make logical decisions depending on their surroundings, 3) the lack of ability to infer road users' goals and intentions. For example, the assumption of road users acting as particles that are affected by forces due to the interactions with other road users in the SFM is not ideal to describe road user behaviour, as discussed by several studies (Liu, et al., 2020; Shiwakoti, et al., 2008; Zhang, et al., 2020; Hussein & Sayed, 2017). The SFM, which is based on attractive and repulsive forces, can capture simple interactions but cannot model complex behaviour. Moreover, road users do not move based on predefined probability rules as the CA model assumes. A road user movement in CA models is usually limited by specified cell size, making it hard to capture the heterogeneity in road user behaviour. As well, the CA model is a discrete simulation model, which is of low fidelity, despite the continuous two-dimensional nature of road user movements. Thus, there is a need to develop a microsimulation model based

on a realistic modeling approach that considers road user intelligence and rationality and infers road user intentions in different situations.

Agent-Based Modeling (ABM) is a relatively new approach for modeling complex systems such as interacting agents and human motion behaviour. This modeling approach has been adopted by studies in various domains, including biology, economics, and social science. In this modeling approach, the system's individuals of interest are modeled as agents situated in the environment. Agents are autonomous and can interact between themselves and the environment based on predefined decision rules. They encapsulate states and can make logical decisions based on their experience and surrounding environments. Developing agent-based models requires modeling agents' goals and strategies (Wooldridge, 1997; Jennings, 2000). A few studies developed models for road user interactions using the ABM approach (e.g., (Fujii, et al., 2017; Hussein & Sayed, 2017)). However, most of the existing models for road user interactions are based on heuristic or ad-hoc rules (Papadimitriou, et al., 2009) and suffer from the inaccurate and unrealistic representation of the behavioural rules that govern their interactions. Designing agents' behavioral rules can be challenging, and agents can be prone to undesired behavior due to behavioral rule miss-specification. Moreover, some studies conducted controlled experiments to study road user behaviour and develop models to reproduce their flow dynamics (Guo, et al., 2020). However, this is problematic as road users' behaviour in controlled experiments may differ comparing with their natural behaviour in normal conditions.

Furthermore, the majority of the developed road users' interaction models for shared space facilities and mixed traffic conditions (Dias, et al., 2018; Hussein & Sayed, 2017; Anvari, et al.,

2015; Luo, et al., 2015) are based on the single-agent modeling framework. However, many modeling problems like autonomous vehicles, robot swarms, and road user interactions are better represented as multi-agent interactions. In the single-agent modeling framework, the actions or the sequence of actions for only a single agent is considered for modeling while assuming the other interacting agents to be part of the passive environment (i.e., their actions are assumed to be known over time). However, in the multi-agent modeling framework, the actions or the sequences of actions of all interacting agents are modeled. Moreover, agents can be cooperative or competitive and can coordinate their actions to achieve specific goals (i.e., avoid collision). The use of the multi-agent modeling framework has several advantages. First, it presents a more realistic approach for modeling road user interactions and traffic conflicts due to its ability to capture social negotiation in road user interactions. Second, it can handle non-stationary environments and capture the equilibrium information compared to the single-agent modeling framework, making it suitable for modeling road user interactions in shared space facilities. Third, it can handle optimal mixed-strategy policies, which are challenging to be handled in the single-agent modeling framework. Previous studies that compared the single-agent and multi-agent modeling performance in different domains, particularly in social and computer science, found that the multi-agent modeling framework outperformed the single-agent modeling framework (Georgila, et al., 2014; Lin, et al., 2014). The studies found that the single-agent modeling is unsuitable for concurrent agents' learning and fails to learn the optimal agent policy since it is designed for stationary environments. The performance of the single-agent modeling approach was found to deteriorate compared to the multi-agent approach as the number of agents increases. Moreover, under the multi-agent framework, agents' utility functions estimation

algorithms could recover more accurate agents' rewards, which are substantially closer to the ground truth rewards, and yield better policies than the single-agent framework.

Few studies adopted the multi-agent modeling framework to develop simulation models for driver lane-changing behaviors and road user behaviour (Talebpour, et al., 2015; Levinson, 2005; Elvik, 2014). However, the majority of these models are based on the traditional game-theoretic framework that models multi-agent systems as a single time-step payoff (i.e., the sequential decision-making process with the state transition concept cannot be handled). Moreover, it is challenging to handle bounded rationality agents (e.g., limited information access) in the traditional game-theoretic framework. Moreover, unlike the single-gent framework, each agent's optimal policy is influenced by the other interacting agents' policies in the multi-agent modeling framework. Thus, the optimality notion is defined by an equilibrium solution concept, such as the Nash Equilibrium (NE) (Nash, 1951) or the Logistic Quantal Response Equilibrium (LQRE) (McKelvey & Palfrey, 1995; McKelvey & Palfrey, 1998). Selecting an appropriate equilibrium solution in MG is vital in specifying the agent policies and model accuracy. Under the NE, agents are assumed perfectly rational and aware of all aspects of their environment (i.e., complete information). Thus, each agent's strategy (i.e., policy) is the best response to other agents' strategies (i.e., perfectly optimal agent behavior). However, under the LSBRE, agents' select strategies based on the relative expected utilities. Thus, better strategies are more likely to be selected than the worse ones, but the best strategies are not selected with certainty. The LSBRE can handle bounded rationality agents and sub-optimal agent behavior. Using NE solution for bounded rationality agent systems leads to obtaining nonsensical results (Harsanyi & Selten, 1988; Harsanyi, 1968). However, limited studies have investigated the ability of

different equilibrium behavioural theories in explaining/ predicting road user operational level decisions (e.g., acceleration and steering/yaw rate) and evasive action mechanisms during their interactions in multi-agent settings. Therefore, there is a need to develop a microsimulation model for road user interactions in shared space facilities, which helps to understand their behaviour in different interactions. The model should be based on the ABM modeling approach that considers road user rationality and intelligence and their ability to assess their surrounding environment and take corresponding decisions. Accurate recovery of road user utility functions would be essential to ensure the accurate representation of their interaction behaviour. The model should be based on the multi-agent modeling framework and should simultaneously model the sequential decision process of road user interactions. The model should investigate and adopt the appropriate equilibrium behavioural theory to ensure the model's accuracy.

2.4 Cyclist and Pedestrian Simulation Models

Over the past few decades, several microsimulation models have been developed to study road users' movement behaviour and address a wide range of road user applications. Most of the existing microsimulation models were developed to model a single mode of transportation, e.g., vehicular traffic (Wiedemann, 1974; Gipps, 1981), pedestrian flow (Helbing & Molnar, 1995; Burstedde, et al., 2001), or cyclist flow (Jiang, et al., 2004; Liang, et al., 2012). Nevertheless, microsimulation models of active road users (i.e., pedestrians and cyclists) are less developed compared to vehicular traffic. Microsimulation models of pedestrians and cyclists are mainly based on the SFM (Helbing & Molnar, 1995) and the CA model (Nagel & Schreckenberg, 1992). Moreover, these models are mainly developed for road segments with physical separations between different transportation modes. For example, Zhou et al. (Zhou, et al., 2019) developed

a microscopic simulation model to study pedestrian crossing behaviour during the flashing green signal. The study developed a modified SFM to describe the mechanism of pedestrian crossing behavior. The desired speed of pedestrians during the flashing green phase is incorporated in the model. The model's parameters were either directly estimated using empirical data or derived indirectly using the maximum likelihood estimation. The developed model was validated against empirical data. The study found that the developed model can produce pedestrian crossing behaviour during the flashing green signal. Taherifar et al. (Taherifar, et al., 2019) utilized a modified SFM for emulating microscopic pedestrians' emergent self-organization and lane formation phenomena. The study proposed a macroscopic calibration framework for bi-directional pedestrian flow in high-density areas with complex phenomena based on the pedestrian area-wide fundamental diagram concept. The developed model was validated against empirical data. The study found that the proposed calibration framework and the modified SFM can replicate pedestrian's fundamental diagrams and microscopic self-organization and lane formation phenomena.

Furthermore, Xue et al. (Xue, et al., 2017) developed a simulation model to study the dynamic characteristics of bicycle flow. The study utilized an improved Burgers CA model to investigate the fundamental diagram of bicycle flow. The model used the following move mechanism, which enabled the smooth movement of bicycles in the model and increased the critical density of bicycles to a reasonable value. The calibration and validation of the model were conducted using experimental and field-measured data. The study found that the improved model can emulate the bicycle flow. Zhang et al. (Zhang, et al., 2020) proposed a multi-grid potential field CA model to emulate overcrowded pedestrian flow conditions. The study enhances the spatial

and temporal continuity of the CA models by dividing a normal cell into nine small cells. This cell formulation allows accounting for the high pedestrian density exceeding 10 ped/m^2. In such formulation, pedestrian compressibility is defined such that one pedestrian occupies a central cell can share a side or corner cell with at most one or three other pedestrians. The compressibility between a pedestrian and a wall is half of that between pedestrians. Moreover, the study modified the CA rules by proposing a smaller movement probability to control the frequency of pedestrian movement at varying densities. The study compared the simulated pedestrian fundamental diagrams and evacuation process with field studies in the literature and found that they are generally in agreement. The model successfully reproduces the self-organized lane formation phenomenon in pedestrian counter-flow.

Several studies extended these models to simulate road user interactions in shared space facilities and mixed traffic conditions. For example, Zhao et al. (Zhao, et al., 2013) developed a cyclist behaviour model that utilizes the CA modeling approach. The model was used to address and emulate passing bicycle events in mixed traffic consisting of bicycles and e-bicycles on separate bicycle paths. The study showed that the CA model was effective in simulating the features of bicycle passing events. Luo et al. (Luo, et al., 2015) proposed a CA model with a modified occupancy rule that accounts for variable lateral distance in the mixed traffic condition of bicycles and cars on urban roads. The proposed occupancy rule specifies lateral distances based on vehicles' speeds. The model captured the complex interactions between bicycles and cars, and the fundamental diagrams of the mixed traffic condition were devised. The microscopic simulation model was used to study the bicycle’s spilling behavior and the impact of the constant occupancy rule. The study explicitly modeled the interference transformation from friction state

to block state to account for the interference of a bicycle on a car. The study concluded that using the fixed and constant occupancy rule might lead to overestimating car flux under heterogeneous traffic. Jin et al. (Jin, et al., 2015) proposed a modified multi-value CA model, called the improved extended Burgers cellular automata (IEBCA) model, to emulate the characteristics of a heterogeneous traffic flow consisting of bikes and e-bikes. The study improved the updating rules for bikes and e-bikes by considering the maximum speeds of two and three cells per second, respectively. In addition, the study conducted numerical simulations and analyzed the bicycle traffic fundamental diagrams. The results showed that the developed model effectively addressed bicycle flow characteristics in heterogeneous traffic, and the results were more realistic and consistent with the field observations than previous models.

Furthermore, Anvari et al. (Anvari, et al., 2015) proposed a three-layered microscopic mathematical model to emulate pedestrian-vehicle interactions in shared spaces. The first layer is a planning layer that computes the routes based on the static obstacles in the environment. This layer calculates the shortest path towards the road user's destination by generating some intermediate targets. The second layer utilized an extended SFM to produces feasible trajectories for mixed traffic conditions. The study extended the SFM to account for the vehicles' velocity angle constraints. In the third layer, rule-based constraints were developed to resolve the conflicts between shared space users. The study utilized an optimization algorithm to calibrate the SFM parameters using field-measure data. Huang et al. (Huang, et al., 2016) developed a microscopic simulation model for cyclist crossing behaviour in heterogeneous traffic at un-signalized intersections. The heterogeneous traffic consists of pedestrians, cyclists, and vehicles. The study proposed a modified SFM and fuzzy logic model to simulate cyclist interactions with

heterogeneous traffic. The model's parameters were either directly estimated using empirical data or derived indirectly using the maximum likelihood estimation. The study showed that the developed model is effective in simulating cyclist crossing behaviour at un-signalized intersections with heterogeneous traffic.

Liu et al. (Liu, et al., 2017) developed a microscopic simulation model to emulate pedestrian dynamics at crosswalks. The study utilized a two-layered modified SFM to simulate pedestrian interactions with pedestrians and vehicles. The first layer is the tactical layer, which determines the desired movement direction for pedestrians. The second layer is the operational layer, which determines the microscopic pedestrian behavior when interacting with other agents. The developed SFM considers pedestrian evasion behavior with counter-flow pedestrians, the following behavior with the leader pedestrians, and the collision avoidance behavior with vehicles. The calibration of the model was carried out using field-measured data, and the model parameters were estimated using the maximum likelihood. The study found that the improved SFM model can replicate microscopic pedestrian behaviour, fundamental diagrams, and lane formation phenomenon. Dias et al. (Dias, et al., 2018) developed a microscopic simulation model to emulate the interactions between pedestrians and personal mobility vehicles (PMVs) such as segways in a shared sidewalk. The study utilized the SFM originally developed for pedestrians and assumed similar characteristics of segways as pedestrians. The model parameters were calibrated using empirical data extracted from controlled experiments under various scenarios. The developed model was validated against the empirical data obtained from the controlled experiments. The study found that the SFM, with proper calibration, can potentially be used to emulate pedestrians and personal mobility vehicles mixed traffic.

In the last decade, the agent-based modeling framework has been frequently used to model road user interactions while considering their intelligence and rationality. In this approach, behavioural rules that control agents' movement and interaction are used. For example, Fujii et al. (Fujii, et al., 2017) proposed an agent-based simulation framework to emulate road user interactions in mixed traffic consisting of cars, trams, and pedestrians. The developed model was used to assess the impact of a tramway extension plan for Okayama City in Japan. The study extended the layered road structure, originally designed for car traffic simulations, to interact with a one-dimensional car-following model and a two-dimensional pedestrian's discrete choice model. The study developed a set of heuristic interaction rules to emulate the interactions of pedestrians with cars. The study conducted several simulations and analyzed the pedestrian and car traffic fundamental diagrams. The developed model was validated against empirical data from the literature. The study found that the developed interaction rules show agreement with the observed empirical values of road user flow characteristics. Hussein and Sayed (Hussein & Sayed, 2017) developed an agent-based simulation model to emulate pedestrian interactions at crosswalks. The study extracted a set of heuristic interaction rules that govern their movement during interactions with other pedestrians in a bi-directional flow. The behavioural rules were extracted based on a detailed pedestrian behavioural study conducted in Vancouver, BC. The model parameters were calibrated either directly, by measuring them from an observed dataset, or indirectly, by the Genetic algorithm aiming at minimizing the error between simulated and actual trajectories. The developed model was validated against collected field-measured data and empirical data from the literature. The study found that the developed model can replicate pedestrians' microscopic behaviour, collision avoidance strategies, and fundamental diagrams.

Dia et al. (Dai, et al., 2013) developed an agent-based simulation model to emulate pedestrian interactions at a bottleneck with complex conflict. The study developed a set of heuristic interaction rules that govern their movement during interactions with other pedestrians in a bi-directional flow evacuation. The speed of the pedestrian is determined based on the speed of the leading pedestrian, the local pedestrian density, and the desired speed of the pedestrian. The walking direction of each pedestrian is determined by a driving force, which represents the sum of four forces, including gradient force, repulsive force, resistance force, and random force. In such an agent-based model, the driving force interprets pedestrian walking direction only, as a distinction from the SFM, where the forces control pedestrians' motion. The model parameter was calibrated, and the simulation results were validated using an observed dataset for passenger alighting and boarding in Kecun station in Guangzhou Metro System in China. The study found that the model replicates pedestrian trajectories and captures the complexity of bi-directional flow behavior near egresses. Moreover, the study compared the developed agent-based model with a floor field CA model and lattice gas (LG) model and found that the agent-based model outperformed the CA and LG models in predicting pedestrian trajectories.

Recently, few studies have adopted the multi-agent modeling framework to develop simulation models for road user interactions. The majority of the developed models are based on the traditional game-theoretic framework that models multi-agent systems as a single time-step payoff. For example, Amini (Amini, et al., 2021) developed a microscopic simulation model to emulate road user behaviour and their interactions in mixed traffic conditions. The study utilized the game-theoretic approach to model road user movements and their decisions when

encountering other road users in roadway crossing under heterogeneous traffic of pedestrians, two-wheeler, three-wheeler, small car, and heavy commercial vehicles. The game-theoretic approach is employed by assuming discrete road user space steps towards their destinations. The study adopted the strategic game of Stackelberg leadership competition to model the interactions between pedestrians and vehicular traffic (von Stackelberg, 1952). In this approach, the decisions of road users are not modeled simultaneously; however, they are modeled as an action and a reaction (i.e., the leader agent selects first the strategy, and then the follower agent selects a strategy according to the leader's announced strategy). The study specified the leader agent as the agent with the lower time to the theoretical collision point (i.e., a point where the two trajectories intersect). The study utilized a sub-game perfect Nash equilibrium (SPNE) to solve the game, where each road user decision is modeled as a single time-step payoff. The model parameters were calibrated using a small sample size of an observed road user interactions dataset from India. The study found that the utilized modeling approach has the potential to provide realistic modeling of road user strategies during their interactions in mixed traffic conditions.

Schonauer et al. (Schönauer, et al., 2012) developed a microscopic simulation model for road user movements and interactions in shared space facilities. The study used the social force model to emulate the interactions between pedestrians and vehicles in shared spaces. The study utilized a discrete single-track approach to model vehicle dynamics. A game-theoretic approach is used to model the interactions on the tactical level for resolving the conflicts between pedestrians and vehicles. The road user interactions were modeled as a non-cooperative game with perfect information and fully rational agents, using the strategic game of Stackelberg leadership

competition (von Stackelberg, 1952). The study assumed that pedestrians are the leader in the game. The Stackelberg is solved using the sub-game perfect Nash equilibrium (SPNE), where each road user decision is modeled as a single time-step payoff. The study developed heuristic utility payoff functions for the agents. The model parameters were calibrated using real trajectories for road user interactions in shared space in Austria. The study compared the simulated and real trajectories and found that the developed model can reproduce car trajectories in shared spaces.

Rahmati & Talebpour (Rahmati & Talebpour, 2018) developed a microscopic simulation model of pedestrian motions and interactions in dynamic environments. The study utilized the game-theoretic approach to model pedestrian decisions in their interactions. The non-zero-sum game is used to describe pedestrian interaction behaviour, where pedestrians have both aligning and conflicting interests. Pedestrian payoff functions were designed heuristically in this study, considering a set of variables, including the pedestrian desired direction, desired speed, and collision avoidance behaviour. The study assumed the desired speed of all pedestrians to be 1.3 m/s. The pedestrian optimal moving strategies were computed using the Nash equilibria calculations, where each pedestrian plays optimally given the other pedestrians' decisions. The study conducted a real-world experiment to extract pedestrian trajectories associated with a bidirectional flow passing through a corridor and exiting from an instructed direction. The extracted pedestrian trajectories were used to calibrate the model parameters. The study found that the utilized modeling approach can describe pedestrian walking behavior and their macroscopic walking parameters and fundamental diagrams.

Chapter 3: Microscopic Behaviour Analysis of Cyclist and Pedestrian Interactions in Shared Spaces

3.1 Background

Understanding and analyzing the behaviour of active road users (i.e., cyclists and pedestrians) in shared space facilities has been gaining increased interest in the literature in recent years. Solid understanding of active road user behaviour and their microscopic interactions aids designers and transportation planners in developing better designs of shared space facilities, which encourages more active road users to use the facilities and enhances their safety and the level of service of the facilities. However, the majority of the studies focused on the macroscopic characteristics of cyclist and pedestrian interactions in such facilities (e.g., the relationship between road user speed and flow rate), with little consideration for their interactions at the microscopic level. This is likely attributed to the lack of available data and tools for such evaluation as a result of the more complex interaction behaviour and the shared right of way. Building simulation models based solely on macroscopic results does not guarantee that the models can produce accurate predictions of shared space user trajectories at the microscopic level, which could limit the applications of the models. This emphasizes the need to investigate the various interaction types between cyclists and pedestrians in shared space facilities and conduct a detailed analysis of their interaction behavior at the microscopic level. Such interactions can be characterized based on road user micro-level parameters (e.g., speed, acceleration, longitudinal distance, and interaction angle profiles). These micro-level parameters can be extracted using state-of-the-art computer vision platforms.

This chapter presents the details of a comprehensive cyclist and pedestrian interaction behavioural study conducted in the City of Vancouver. Video data were collected at the busy shared space of Robson Square in downtown Vancouver. Computer Vision (CV) algorithms were used to detect and track pedestrians and cyclists from video footage, using the automated computer vision platform developed at the University of British Columbia (Saunier & Sayed, 2006). Trajectories of cyclists and pedestrians involved in interactions were extracted and used to calculate several variables that describe elements of road users' behaviour, including longitudinal and lateral distances, road user speed, interaction angle, and acceleration profiles. The interactions between cyclists and pedestrians were classified based on their interaction angles into (1) interactions with shared space users moving in the same direction, (2) interactions with opposing direction shared space users, and (3) interactions with crossing shared space users. The speed and acceleration profiles and relative spatial profiles were used to analyze the interaction behaviour of pedestrians and cyclists in different cases. Moreover, the study explored the collision avoidance mechanisms employed by road users to avoid collisions with other shared space users. In addition, the study identified a set of parameters that can be used to calibrate a microscopic cyclist-pedestrian modeling platform to represent the behaviour of pedestrians and cyclists in shared space environments. The study presented in this chapter has several objectives. First, investigate and demonstrate the role of the computer vision platform in collecting and analyzing cyclist and pedestrian interactions in shared space facilities. Second, extract and analyze important microscopic elements required to understand cyclist and pedestrian interactions and the collision avoidance mechanisms employed by them to avoid collisions with other users in the shared space facility. Moreover, identifying the factors affecting the selection of the collision avoidance mechanisms in such facilities. Lastly, identify the main set of

parameters required to model specific cyclist and pedestrian interaction in agent-based microsimulation models. Extracting and analyzing road user behaviour and the parameters from actual data is expected to develop a more realistic simulation model and enhance the accuracy of the trajectories produced from the model.

The following sections in this chapter provide a detailed description of the study location and video data collection, the methodology applied to analyze video data and extract road user trajectories, the calculation of variables that describe elements of road users' behaviour, the identification of cyclist and pedestrian interactions, and the results of analyzing road user interaction and collision avoidance mechanisms.

3.2 Study Location and Data Collection

Video data were collected at a busy non-motorized shared space located in Robson Square in downtown Vancouver, British Columbia, as shown in Figure 3.1. The city of Vancouver considered the permanent closure of Robson Street, between Hornby Street in the west and Howe Street in the east, for motorized traffic in order to provide a comfortable and safe plaza for vulnerable road users in the heart of downtown. This area is a place for many recreational and business facilities, including the Vancouver Art Gallery, the Robson Square ice rink, and the Vancouver Supreme Court. The area is an active environment for walking and cycling, especially in the summer season. Cyclist-pedestrian interactions were frequently observed in the shared space area. Eighteen hours of video data were collected in daylight over several days in August and September 2016. Video data were collected using a video camera mounted at the

shared space area near Hornby Street at the western entrance of the shared space, as shown in Figure 3.1(a).

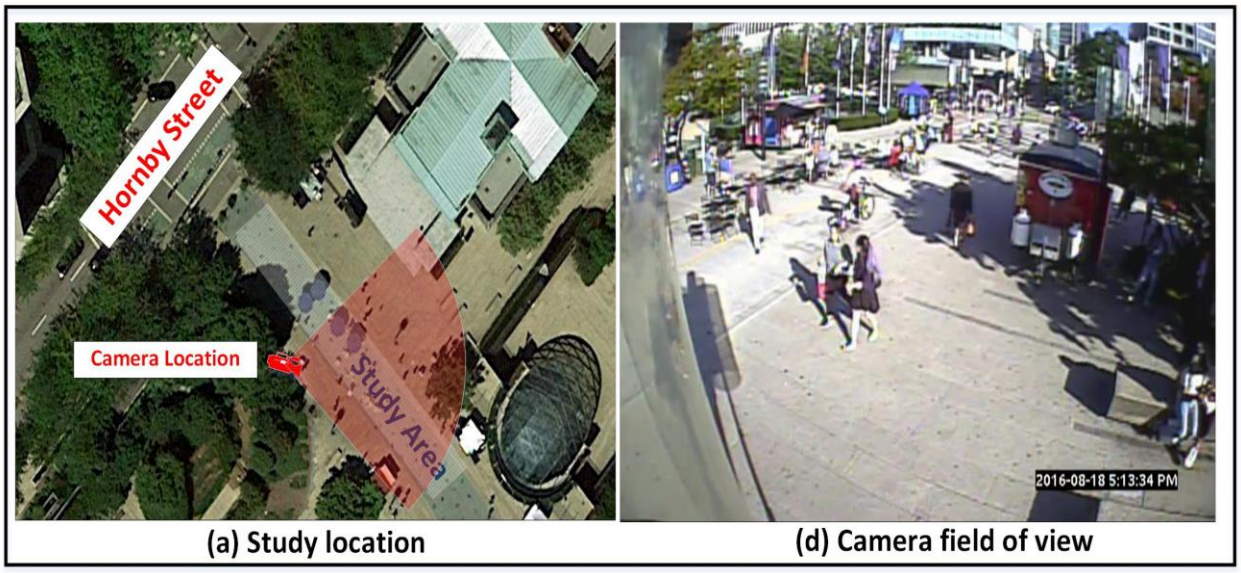


Figure 3.1 Data collection location

3.3 Road User Tracking: A Computer Vision Approach

The automated extraction of road user trajectories from video footage was conducted using a video analysis system that has been developed at the University of British Columbia (Saunier & Sayed, 2006). The system employs computer-vision algorithms to detect and track road users in traffic scenes. The detection of the road users is carried out using an implementation of the Kanade–Lucas–Tomasi (KLT) feature-tracking algorithm (Lucas & Kanade, 1981) (Tomasi & Kanade, 1991). The algorithm detects distinct points (features) on moving objects in the video scene. The algorithm is capable of differentiating between features that belong to road users (e.g., pedestrians, cyclists) and that are part of the environment.

Typically, multiple features are tracked on each road user in each video frame. Clustering of features is important to determine which group of features belongs to the same road user. The clustering algorithm uses different cues to cluster the tracked features, including the spatial proximity of features and their dynamical (i.e., movement pattern) similarities. The clustered road user features are then tracked over time to generate road user trajectories. In this study, road user classification was carried out manually to avoid road user misclassification in the shared space facilities.

However, an essential task prior to the aforementioned procedures is the camera calibration process, in which a homography matrix is generated to create a mapping between the two-dimensional video image coordinates and the real-world three-dimensional coordinates, as described in details in Ismail et al. (Ismail, et al., 2013). This enables transferring the spatial and temporal information of the tracked trajectories to the actual coordinate system of the location being analyzed. The accuracy of the system was validated in several previous studies and was found to be satisfactory high (Tageldin & Sayed, 2019; Ismail, et al., 2013; Ismail, et al., 2010). The overall analysis procedures are illustrated in Figure 3.2.

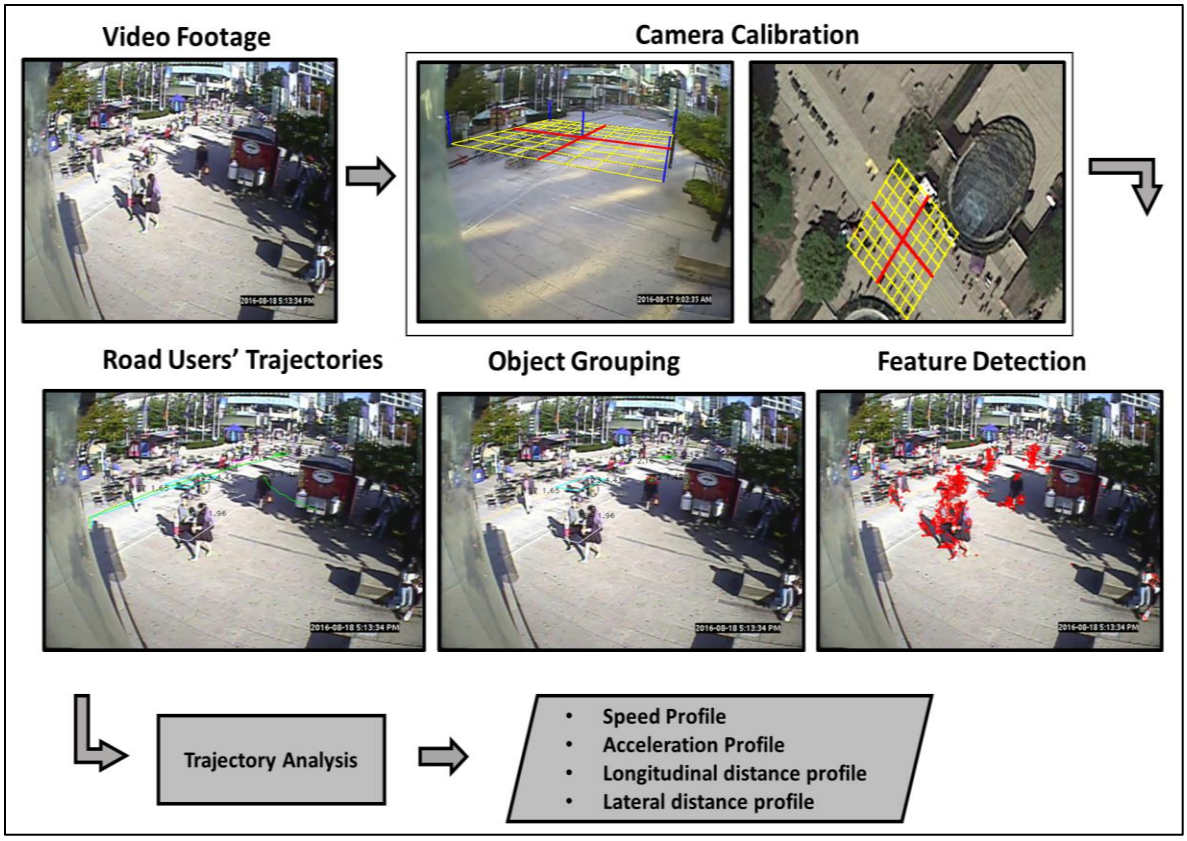


Figure 3.2 Trajectory and behavioural profiles extraction process

3.4 Extraction of Road User Behavioual Profiles

Road user trajectories capture the movement of each pedestrian and cyclist in the form of a sequence of spatial coordinates and instantaneous speed at each video frame (1/30 second). A road user trajectory (T) is defined along the trajectory lifetime (n video frame) as a finite set of tuples, as shown in Equation (3.1).

$$T = \{(X_1, Y_1, V_{X_1}, V_{Y_1}, \ \dots \ , X_i, Y_i, V_{X_i}, V_{Y_i}, \ \dots \ , X_n, Y_n, V_{X_n}, V_{Y_n})\} \tag{3.1}$$

where $i = \{1, \dots, n\}$ is a discrete temporal index, X_i and Y_i are the spatial coordinates of the road user at time frame (i), and V_{X_i}, V_{Y_i} are the corresponding velocities. The road user speed profile

(S) is computed by normalizing the speed components along the trajectory lifetime, as given by Equation (3.2). The acceleration profile (Acc) of each road user is derived from the smoothened speed profile as the rate of change of the instantaneous velocity (S) over time, as given by Equation (3.3).

$$S(t) = l^2\text{-}norm(V_x, V_y) \tag{3.2}$$

$$Acc(t) = \frac{d(S)}{dt} \tag{3.3}$$

where V_x and V_y are the velocity vectors of length n, for the X and Y coordinates, respectively, and t is time.

The cyclist heading angle (Ψ) and longitudinal and lateral distances in shared space facilities are defined as illustrated in Figure 3.3. The distance vector ($\vec{\boldsymbol{d}}$) is defined as the distance between the cyclist and pedestrian and directing toward the pedestrian. The angle θ is defined regarding the cyclist considering the pedestrian as a neighbour user. The angle θ is the angle between the velocity vector of the cyclist ($\vec{\boldsymbol{v}}$) and distance vector ($\vec{\boldsymbol{d}}$), and can be calculated as given by Equation (3.4). The longitudinal distance is the distance between the cyclist and pedestrian along the direction of the cyclist's movement. Equation (3.5) shows the calculation of the longitudinal distance ($d_{Longitudinal}$) between the cyclist and pedestrian. Initially, the longitudinal distance is positive as the cyclist is behind the pedestrian. Negative values of the longitudinal distances indicate that the cyclist becomes ahead of the pedestrian. The lateral distance is the absolute distance between the cyclist and pedestrian perpendicular to the direction of the cyclist's movement. Equation (3.6) shows the calculation of the lateral distance ($d_{Lateral}$) between the

cyclist and pedestrian. The interaction angle between road users is computed as a difference between their heading angles.

$$\theta = \cos^{-1}\left(\frac{\vec{d} \,.\, \vec{v}}{\|\vec{d}\| \,.\, \|\vec{v}\|}\right) \tag{3.4}$$

$$d_{Longitudinal} = \|\vec{d}\|.\cos\theta \tag{3.5}$$

$$d_{Lateral} = \left|\|\vec{d}\|.\sin\theta\right| \tag{3.6}$$

The accuracy of the road user trajectories' extraction process was validated by comparing a sample of 100 trajectories with the manually measured road user speeds based on the time required to traverse a known distance in the video scene. The mean absolute percentage error of the extracted road user speeds was 6.6 % (mean absolute error of 0.11 m/s for pedestrians and 0.18 m/s for cyclists), which is considered low and not to affect the findings.

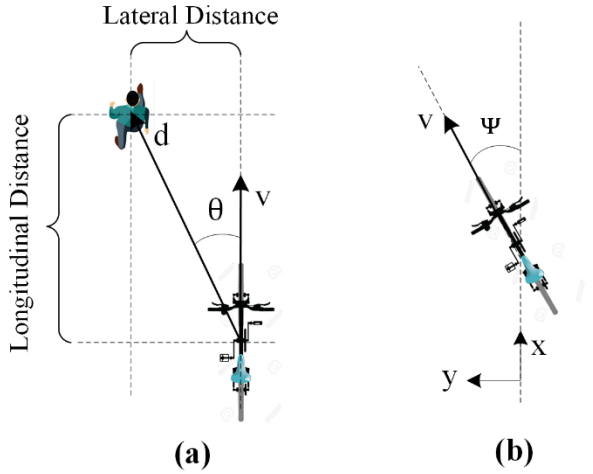


Figure 3.3 Illustrations of (a) cyclist longitudinal and lateral distances, and (b) cyclist heading angle

3.5 Interaction Identification and Analysis

Video data were manually reviewed in order to define different types of interactions between pedestrians and cyclists in the shared space. Observed interactions were classified into three main categories based on the interacting angle (Beitel, et al., 2018):

- Interactions with slower road users moving in the same direction: These interactions involve cyclists and pedestrians moving in the same directions (difference in movement direction = 0° ± 30°).
- Interactions with opposing road users: These interactions involve cyclists and pedestrians moving in the opposite directions (difference in movement direction = 180° ± 30°).
- Interactions with crossing road users: These interactions include all other cyclist-pedestrian interactions.

A total number of 208 interactions were observed during the 18 hours of the analyzed video data. The trajectories, speed and acceleration profiles, and longitudinal and lateral distance profiles of all cyclists and pedestrians involved in at least one of the previous interactions were considered for the analysis. In the analyzed interactions, evasive action was taken by at least one of the road users involved in an interaction to resolve the conflict. Generally, two types of evasive actions were observed: (1) changing cycling/walking speed, and (2) swerving maneuvers. The start and end of each interaction were identified manually from the road user trajectory and behavioural profiles, similar to (Hussein & Sayed, 2015). Test of significance (two-sample *t*-test) is used to investigate the significance of the change in road user behaviour during the interaction.

3.6 Results and Discussion

3.6.1 Interactions with Road Users Moving in the Same Direction

This type of interaction occurs when a faster road user is hindered by another slower user, moving in the same direction (difference in movement direction = 0° ± 30°). These interactions are frequently observed in shared space areas due to the significant difference in the operational speed between different categories of road users (Department of Transport and Main Roads of Queensland State, 2014). Faster road users (typically cyclists) tend to apply at least one of the following two collision avoidance mechanisms to avoid collisions with the slower road users (pedestrians): (1) reduce their speeds to keep adequate space with the slower road users (following maneuver), or (2) apply swerving maneuvers to overtake the slower road users (overtaking maneuvers). In most cases, the overtaking maneuvers are associated with a change in speed of the faster road user. The factors affecting the selection of the collision avoidance mechanism and a detailed description of the two strategies are discussed in more detail in the following subsections.

3.6.1.1 Selection of the Collision Avoidance Strategy

As discussed earlier, road users involved in interactions with other road users moving in the same direction may choose to follow slower road users or overtake them. Reviewing the interactions suggested that the selection of the collision avoidance mechanism taken by the faster road users mainly depends on two factors; (1) the available lateral space for the faster road user, and (2) the shared space density during the interaction. When the density of the shared space increases, it becomes more difficult for faster road users to swerve and overtake slower pedestrians. Similarly, when the available lateral space for the faster road users is limited due to

barriers or other street furniture, faster road users may be forced to reduce their speed and follow slower road users.

A binary logit model was developed to calculate the probability of selecting each of the two collision avoidance strategies. One hundred and five cyclist-pedestrian interactions in the same direction were reviewed manually, and the collision avoidance strategies adopted by the faster road users (i.e., cyclists) were recorded. The model is expressed as given by Equation (3.7).

$$P_{overtaking} = \frac{1}{1+e^{-(a_0+a_1.Dist+a_2.Den)}} \quad (3.7)$$

where $P_{overtaking}$ is the probability of performing an overtaking maneuver, $Dist$ is the maximum available lateral distance for the faster road user during the interaction (m), Den is the average shared space density (road user / m^2), and a_0, a_1, and a_2 are model parameters. The density is calculated over the shared space area that covers the interacting road users while considering the presence of shared space furniture (e.g., tables), similar to the previous studies (Hussein & Sayed, 2018; Liu, et al., 2014).

The estimated model parameters are presented in Table 3.1. The coefficients of the explanatory variables were significant at 95%. The estimated parameters confirm that the probability of the faster cyclist to execute an overtaking maneuver increases as the density of the shared space decreases and the available space increases as shown in Figure 3.1. The average value of the shared space density and available space for the two collision-avoidance strategies are presented in Table 3.2.

Variable	Estimate (P-value)
Intercept	5.045[b] (0.09)
Lateral Distance (m)	2.504[a] (<0.01)
Density (road user/m^2)	-29.888[a] (<0.01)
Goodness of fit	
DF	102
R^2	0.801

Note: Values in parentheses represent the p-value. 1 m = 3.28 ft.
[a]Statistically significant (at 95% confidence level).
[b]Statistically significant (at 90% confidence level).

Table 3.1 Logit model estimates for the selection of the collision avoidance mechanism (Alsaleh, et al., 2020)

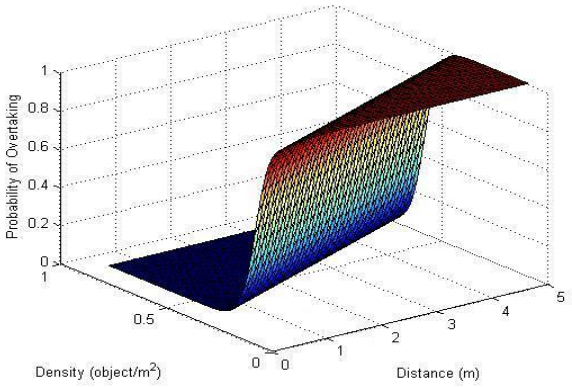


Figure 3.4 Probability of overtaking in cyclist-pedestrian interaction (Alsaleh, et al., 2020)

Parameters	Mean [SD], Collision Avoidance Mechanism	
	Following Strategy	Overtaking Strategy
Max. Available Distance (m)	1.28 [0.65]	3.53[a] [1.58] (<0.01)
Shared space Density (object/m^2)	0.52 [0.16]	0.25[a] [0.10] (<0.01)

Note: Values in parentheses represent the p-value of the t-test. SD = standard deviation.
[a]Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.

Table 3.2 Features significance across the different collision avoidance strategies (Alsaleh, et al., 2020)

3.6.1.2 Description of the Following Strategy

Sixty-five pedestrian-cyclist interactions that involved following maneuvers were analyzed in this study. The speeds of all road users involved following maneuvers were extracted, as well as the longitudinal distances between each pair of road users in the conflict. Figure 3.5 shows an example of the following maneuver that involves a faster cyclist and a slower pedestrian. The cyclist and pedestrian speed and acceleration profiles and the longitudinal distance profile between the cyclist and the pedestrian are shown in Figure 3.6. As shown in the figure, the interaction can be classified into three phases:

1. Phase 1: the cyclist is traveling at a normal cycling speed (4.33 m/s in this example) as he/she is approaching a slower walking pedestrian.
2. Phase 2: at a specific distance from the slower pedestrian (4.68 m in the current example), the cyclist started to reduce the speed to avoid collision with the slower pedestrian. The average longitudinal distance at which cyclists started to reduce their speeds for the 65 interactions analyzed was found to be 4.46 m, with a standard deviation of 1.29 m. The average deceleration value for the 65 interactions was found to be 0.81 m/s^2, with a standard deviation of 0.40 m/s^2. This phase can be considered an evasive action taken by the cyclists to avoid collision with the slower pedestrians during the high-density periods in the shared space.
3. Phase 3: the cyclist approaches a safe following distance (2.02 m in the current example) so that he/she keeps traveling at a reduced speed to maintain this distance. The average following distance was found to be 2.36 m, with a standard deviation of 0.50 m for the 65 interactions analyzed.

Table 3.3 shows the mean and standard deviation of the cycling and walking speeds during the following maneuver, compared to the corresponding values before the interaction (i.e., during unconstrained moving periods). As shown in the table, the cycling speed was significantly reduced by about 39% when following slower pedestrians in the shared space. On the other hand, no significant change in the walking speed was observed during the maneuver, which was expected since pedestrians did not apply any evasive actions and, in some cases, they were not even aware of the following cyclists. Based on the previous analysis, it is suggested to define three parameters to model this type of interaction in a microsimulation model:

1. The longitudinal distance between the pedestrians and the cyclists at which the cyclists start to reduce their speeds.
2. The desired distance (following distance) the cyclists prefer to keep from slower leading pedestrians during the following maneuver.
3. The cyclists' deceleration needed to reduce speed when approaching the slow pedestrians.

(a) Before interaction (b) During Interaction

Figure 3.5 Example of a typical following strategy by a cyclist hindered by a slower pedestrian moving in the same direction

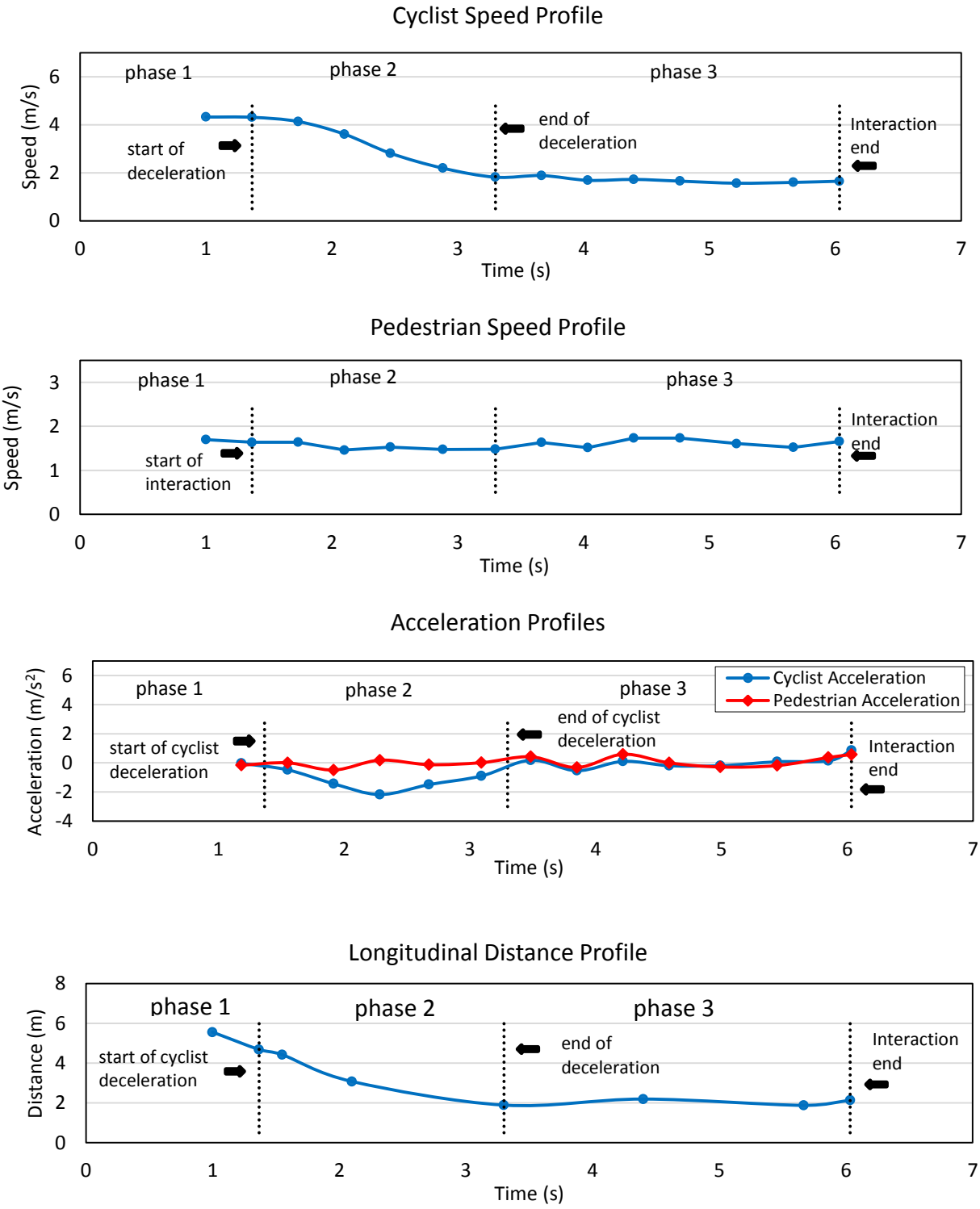


Figure 3.6 Speed, acceleration, and longitudinal distance profiles for a typical following strategy by a cyclist hindered by a slower pedestrian moving in the same direction (Alsaleh, et al., 2020)

Bicycle- Pedestrian Interaction Type	Road User Type	Parameter	Mean [SD], Bike Changing Speed	
			Normal/Before Interaction	During Interaction
Following Strategy	Bike	Speed (m/s)	3.12 [0.74]	1.85[a] [0.45] (<0.01)
	Pedestrian	Walking Speed (m/s)	1.67 [0.36]	1.68 [0.39]
Overtaking Strategy	Bike	Speed (m/s)	2.30 [1.10]	3.14[a] [0.63] (<0.01)
	Pedestrian	Walking Speed (m/s)	1.24 [0.69]	1.18 [0.67]

Note: Values in parentheses represent the p-value of the t-test. SD = standard deviation.
[a]Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.

Table 3.3 Comparison between cyclist and pedestrian speeds for normal cycling/walking behaviour and during the following and overtaking maneuvers (Alsaleh, et al., 2020)

3.6.1.3 Description of the Overtaking Strategy

Forty interactions that involve an overtaking maneuver were observed in the analyzed data set. This maneuver typically includes cyclists changing their movement directions at a specific distance from the slower pedestrians to avoid collisions. As well, an increase in cycling speed was observed during the overtaking phase. Figure 3.7 shows an example of a cyclist's overtaking maneuver. The cyclist and pedestrian speed and acceleration profiles, and the longitudinal distance profile between the cyclist and the pedestrian are shown in Figure 3.8. As shown in the figure, this type of interaction can be described by the following three phases:

1. Phase 1: the cyclist is traveling at a normal cycling speed (or at a reduced speed if the cycling is following a slower pedestrian, as was noticed in many cases).

2. Phase 2: the cyclist decides to execute a swerving maneuver to overtake the slower pedestrian. This usually occurs when adequate space becomes available for the cyclists to perform the maneuver. The cyclist accelerates and swerves in order to overtake the slower pedestrian. The average acceleration observed in the 40 interactions was found to be 0.87 m/s^2, with a standard deviation of 0.35 m/s^2. According to the data analyzed in this study, the swerving maneuver is usually initiated simultaneously with the cyclist's acceleration. The average longitudinal distance between the overtaking cyclists and the overtaken pedestrians at the start of the overtaking maneuver was found to be 3.00 m, with a standard deviation of 1.65 m.
3. Phase 3: the cyclist reaches the desired speed and the desired lateral distance that he/she prefers to keep from the slower pedestrian. Consequently, the cyclist maintained the desired speed and continue cycling in the original movement direction to overtake the slower pedestrian. The average minimum lateral distance between the overtaking cyclists and the overtaken pedestrians during the analyzed maneuvers was found to be 1.23 m, with a standard deviation of 0.25 m.

Table 3.3 shows the mean and the standard deviation of the cycling and walking speeds during the overtaking maneuver, compared to the corresponding values before the interaction. As shown in the table, the cycling speed increased significantly during the overtaking process. On average, cyclist speeds were increased by about 41% during the overtaking of the slower pedestrians in the shared space. Similar to the previous interaction, no significant change in the walking speed was observed during the maneuver, which was expected since pedestrians did not apply any

evasive action during the interaction and, in some cases, they were not even aware of the cyclists.

Based on the previous analysis, it is suggested to define three parameters to model this type of interaction in a micro-simulation model:

1. The longitudinal distance between the cyclists and the pedestrians at which the swerving maneuver starts.
2. The lateral distance between the overtaking cyclists and the overtaken pedestrians at the passing point.
3. The cyclists' acceleration needed to execute the overtaking maneuver.

(a) Before interaction (b) During Interaction

Figure 3.7 Example of a typical overtaking strategy by a cyclist hindered by a slower pedestrian moving in the same direction.

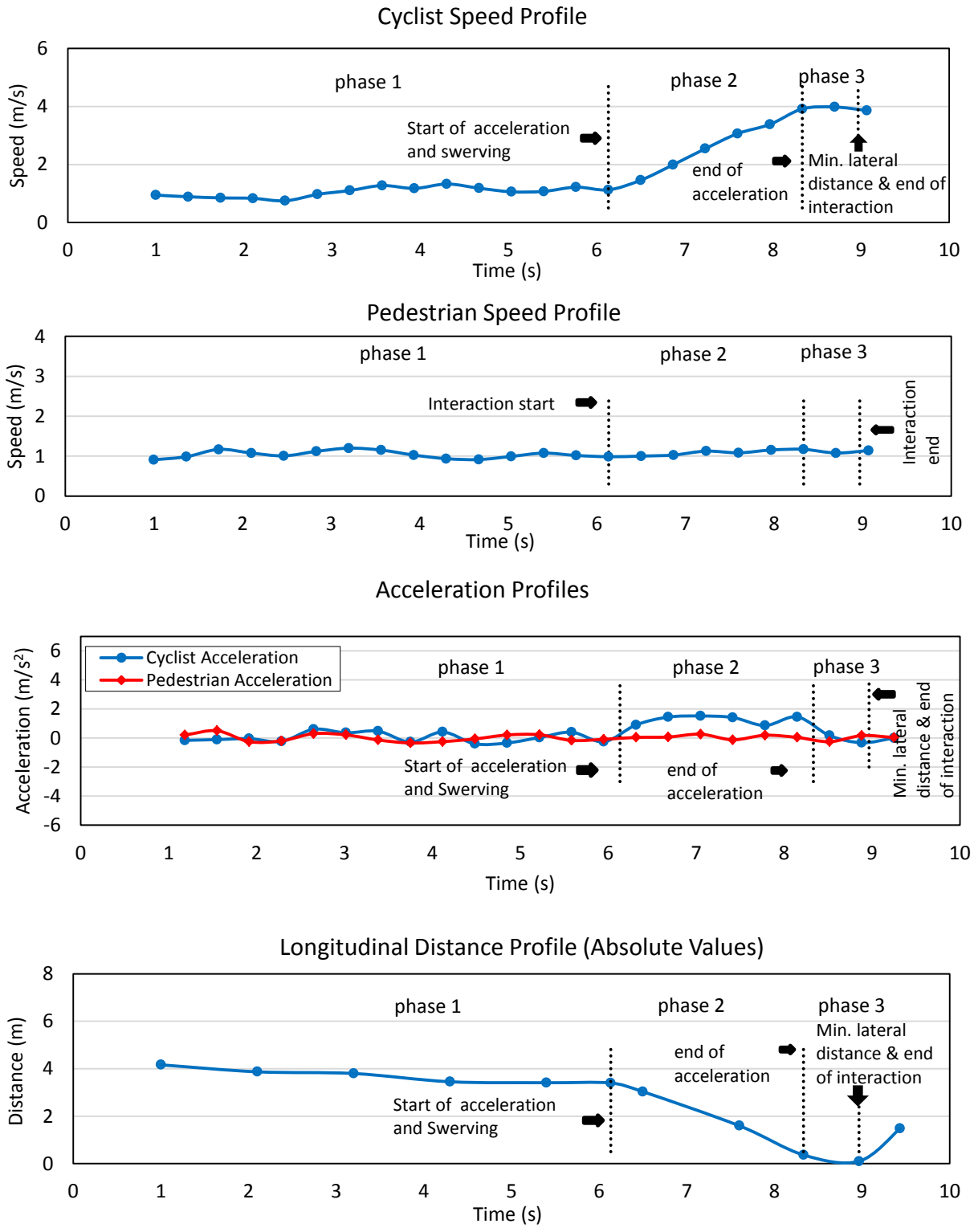


Figure 3.8 Speed, acceleration, and longitudinal distance profiles for a typical overtaking strategy by a cyclist hindered by a slower pedestrian moving in the same direction (Alsaleh, et al., 2020)

3.6.2 Interactions with Road Users Moving in Opposing Direction (Head-on Interactions)

This type of interaction occurs when a pedestrian and a cyclist approach each other, with an angle of 180° ± 30°, in time and space so that a collision would occur if none of them takes evasive action. In the analyzed video data, sixty-two head-on interactions were observed. This type of interaction involves at least one of the road users in conflict changing direction to avoid the collision. It was observed that most of the maneuvers are associated with reductions in cyclist speeds. Table 3.4 shows the mean and the standard deviation of the cycling and walking speeds during the interaction and unconstructed moving periods. As shown in the table, cycling speed is reduced significantly during the interaction, regardless of whether the cyclist executes a swerving maneuver or not. The average reduction in cyclist speeds during the interaction was found to be 20%. On the other hand, no significant change in walking speeds was observed during the interactions regardless of whether pedestrians change their movement directions or not. This agrees with the results reported in (Hussein & Sayed, 2015), which indicated that pedestrians do not change speed when changing direction to avoid opposing pedestrians. However, other studies found that pedestrians may change their speed while changing their direction (Daamen, et al., 2014). Figure 3.9 and Figure 3.10 show a typical example of a head-on interaction along with the corresponding speed and acceleration profiles of the two shared space users involved in the interaction. As shown in Figure 3.9, this type of interaction can be described by the following three phases:

1. Phase 1: the cyclist is traveling at a preferred speed, depending on the current density of the shared space environment.
2. Phase 2: the cyclist approaches the opposing pedestrian and decides to swerve and reduce speed to avoid a collision. The average deceleration value for the cyclists was found to be

1.14 m/s^2, with a standard deviation of 0.75 m/s^2. The average longitudinal distance at which the cyclist or pedestrian starts to execute the evasive action was found to be about 6.39 m, with a standard deviation of 2.31 m.

3. Phase 3: once the cyclist ensures that he/she has an adequate lateral distance from the opposing pedestrian, the cyclist starts to accelerate to the normal (i.e., preferred) cycling speed. The average acceleration value for the cyclists was found to be 1.11 m/s^2, with a standard deviation of 0.56 m/s^2. The average lateral distance between the cyclists and pedestrians at the crossing point was found to be 1.24 m, with a standard deviation of 0.32 m. This lateral distance is significantly higher than the corresponding distance in pedestrian head-on interactions. Hussein and Sayed reported that the average lateral distance between two opposing pedestrians on a typical crosswalk in Vancouver is about 0.73 m (Hussein & Sayed, 2017). This significant difference can be explained by the larger size and the higher speed of bicycles compared to pedestrians.

Based on the previous analysis, it is suggested to define three parameters to model the head-on interaction between cyclists and pedestrians in a microsimulation model:

1. The longitudinal distance between the cyclist and the pedestrian at which one of them starts to swerve to avoid the collision.
2. The lateral distance between the cyclist and the pedestrian at the passing point.
3. The cyclists' acceleration and deceleration needed to execute the swerving maneuver.

(a) Before interaction (b) During Interaction

Figure 3.9 Example of a typical head-on interaction between cyclist and pedestrian (opposing direction interaction)

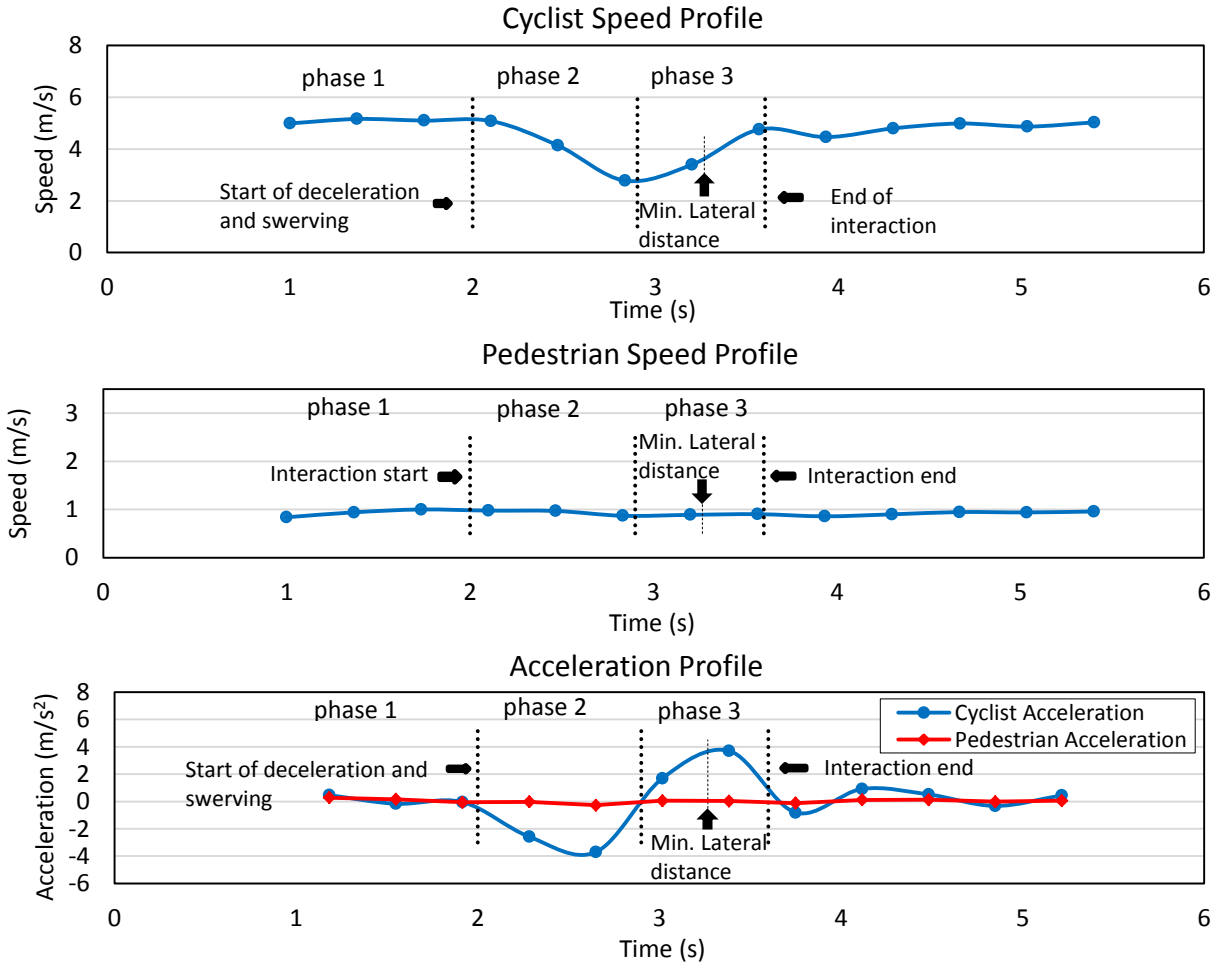


Figure 3.10 Example of a typical head-on interaction between cyclist and pedestrian expressed by their speed and acceleration profiles (Alsaleh, et al., 2020)

Road Users involved in Head-on Interaction	Parameter	Mean [SD], Bikes Swerve		Mean [SD], Ped Swerve		Mean [SD], Both Swerve		Mean [SD], Total	
		Normal/ Before interaction	During Interaction	Normal/ Before interaction	During Interaction	Normal/ Before interaction	During Interaction	Normal/ Before interaction	During Interaction
Bikes	Speed (m/s)	3.70 [1.40]	2.98[a] [1.07] (0.01)	3.00 [0.96]	2.33[a] [0.68] (0.01)	2.86 [0.62]	2.26[a] [0.71] (0.04)	3.45 [1.29]	2.74[a] [0.99] (<0.01)
Pedestrian	Walking Speed (m/s)	1.67 [0.43]	1.70 [0.48]	1.57 [0.41]	1.57 [0.43]	1.74 [0.31]	1.78 [0.32]	1.66 [0.40]	1.68 [0.44]

Note: Values in parentheses represent the p-value of the t-test. SD = standard deviation.

[a]*Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.*

Table 3.4 Comparison between cyclist and pedestrian speeds for normal cycling/walking behaviour and during the swerving maneuver in the head-on interaction (Alsaleh, et al., 2020)

3.6.3 Interaction with Crossing Road Users

This type of interaction occurs when a pedestrian and a cyclist approach each other from the side in time and space so that a collision can occur if either of the road users takes no action. The shared space does not provide a right of way for one user over the other. Instead, all users share the right-of-way so that it is not always clear which road user should yield during these interactions. Forty-one cyclist-pedestrian crossing interactions were observed in the data set. At least one shared space user involved in this type of interaction changed the speed and/or the moving direction to avoid a collision. The 41 interactions were classified into two categories; (1) interactions where the pedestrian crosses first (cyclist yields to the pedestrian), and (2) interactions where the cyclist crosses first (pedestrian yields for the cyclist). If the cyclist yields to pedestrians, pedestrians tend to increase their walking speed to cross and clear the conflict area. As shown in Table 3.5, the pedestrian speed increased significantly during the interaction if the cyclist yields to the pedestrian. On average, the speed of pedestrians increased by about 25.7% during the interaction. The cycling speed was reduced significantly by about 23.4% in order to provide right of way for the pedestrians. If the pedestrian yields to the cyclist, pedestrians reduce their walking speed. On average, pedestrians significantly reduced their speed by 63.5% in order to provide the right of way to the crossing bicycles. As well, the cyclist speed was reduced significantly during the interaction. The average reduction in cycling speed during the interaction was found to be 18.8%. Values reported in Table 3.5 suggest that cyclists tend to reduce speed when approaching a crossing pedestrian. However, the decision is likely made by the pedestrian whether to speed up and cross first or slow down and allow the cyclist to cross first. The pedestrian's decision likely depends on the approaching bike speed as shown in Table 3.6, where the approaching speeds of cyclists are significantly higher for the cases where

the pedestrian yields (bike crosses first) compared to the cases where pedestrian crosses first. If the approaching cyclist speed is high, pedestrians prefer to yield for the cyclist. Otherwise, pedestrians prefer to accelerate to cross before the cyclists.

Figure 3.11 and Figure 3.12 show a typical example of a crossing interaction involving a cyclist yields to the crossing pedestrian and their corresponding speed and acceleration profiles. As shown in the figure, the interaction can be described by the following three phases:

1. Phase 1: the cyclist and pedestrian are traveling at their preferred cycling and walking speeds, depending on the current density of the shared space environment.
2. Phase 2: the pedestrian decides to cross first. The pedestrian increases his/her walking speed to clear the conflict course safely. The cyclist decelerates in order to allow the pedestrian to cross safely. The average deceleration of cyclists yielded to pedestrians was found to be 0.92 m/s^2, with a standard deviation of 0.42 m/s^2. However, the average cyclists' deceleration when pedestrians yield to cyclists was found to be 1.21 m/s^2, with a standard deviation of 0.68 m/s^2. The average longitudinal distance between the cyclists and pedestrians along the crossing point at which the cyclist or pedestrian behavior starts to be affected by this interaction is about 5.95 m, with a standard deviation of 2.73 m. At the end of this phase, the pedestrian has safely crossed the conflict course.
3. Phase 3: as the pedestrian clears the conflict course, the cyclist starts to bike at a cycling speed suitable to the current density level of the shared space.

Based on the previous analysis, it is suggested to define three parameters to model the crossing interaction between pedestrians and cyclists in a microsimulation model. However, a more detailed analysis for a larger data set is required to determine the factors that affect pedestrians'

decision to yield or accelerate and cross first and the factors that affect the change of road user movement directions. The three recommended parameters are summarized as follows:

1. The percentage change in pedestrian's speed for the two possible evasive actions (pedestrian yields to bike or accelerates and crosses first).
2. The longitudinal distance between the pedestrian and the cyclist along the crossing point at which they start to modify speed to resolve the conflict.
3. The lateral distance between the cyclist and the pedestrian at the crossing point.
4. The cyclists' deceleration as they approach the crossing pedestrians for the two possible evasive actions.

(a) Before interaction (b) During Interaction

Figure 3.11 Example of a typical crossing interaction between cyclist and pedestrian (the case of pedestrian crosses first)

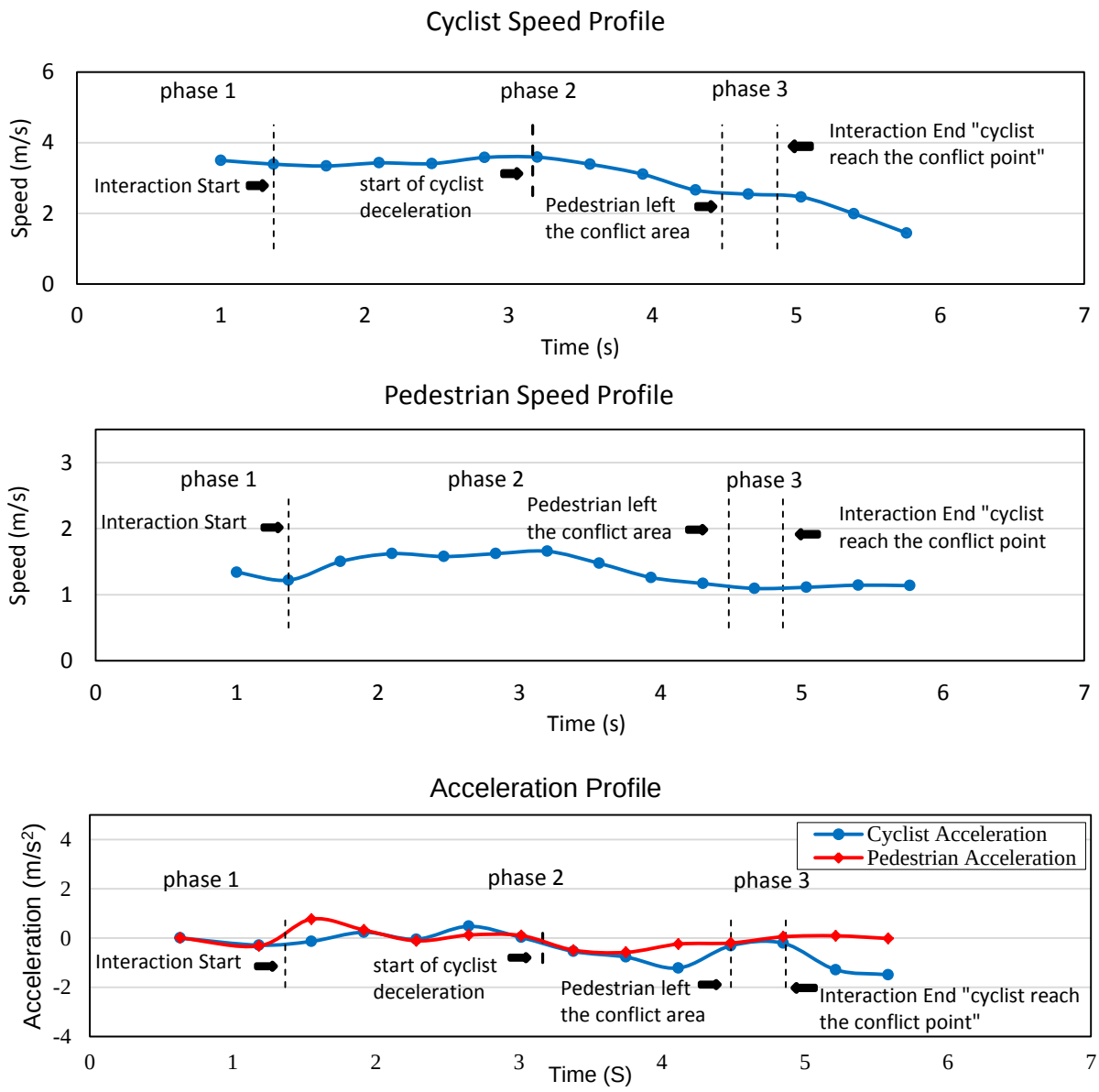


Figure 3.12 Example of a typical crossing interaction between cyclist and pedestrian expressed by their speed and acceleration profiles (the case of pedestrian crosses first) (Alsaleh, et al., 2020)

Road Users involved in Crossing Interaction	Parameter	Mean [SD], Pedestrians Cross First		Mean [SD], Bikes Cross First		Mean [SD], Bikes Swerve		Mean [SD], All	
		Normal/ Before Interaction	During Interaction	Normal/ Before Interaction	During Interaction	Normal/ Before Interaction	During Interaction	Normal/ Before Interaction	During Interaction
Bike	Speed (m/s)	3.36 [1.44]	2.57[b] [1.12] (0.08)	4.61 [1.60]	3.65[a] [1.00] (0.03)	3.52 [1.10]	2.80[a] [0.89] (0.05)	4.02 [1.63]	3.14[a] [1.18] (0.01)
Pedestrian	Walking Speed (m/s)	1.32 [0.28]	1.67[a] [0.44] (0.01)	1.40 [0.41]	0.51[a] [0.44] (<0.01)	1.24 [0.25]	0.98 [0.74]	1.37 [0.35]	1.07[a] [0.73] (0.02)

Note: Values in parentheses represent the p-value of the t-test. SD = standard deviation.

[a]*Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.*

[b]*Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.*

Table 3.5 Comparison between cyclist and pedestrian speeds for normal cycling/walking behaviour and during the swerving and yielding maneuvers in the crossing interaction (Alsaleh, et al., 2020)

Parameters	Mean [SD], Bikes		Mean [SD], Pedestrians	
	Pedestrians Cross First	Bikes Cross First	Pedestrians Cross First	Bikes Cross First
Approaching Speed (m/s^2)	3.36 [1.44]	4.61[a] [1.60] (0.01)	1.32 [0.28]	1.41 [0.4]

Note: Values in parentheses represent the p-value of the t-test. SD = standard deviation.

[a]*Statistically significant difference (at 95% confidence level) compared with the cell directly to the left.*

Table 3.6 Comparison between cyclist and pedestrian approaching speeds in the crossing interaction (Alsaleh, et al., 2020)

3.7 Summary and Conclusion

The main objective of the work presented in this chapter is to investigate the microscopic interaction behaviour of cyclists and pedestrians in a non-motorized shared space in Vancouver, British Columbia. Video data were collected at the Robson Square shared space in downtown Vancouver. Trajectories of cyclists and pedestrians involved in 208 interactions (416 Trajectories) in the shared space area were extracted using computer vision algorithms. The extracted trajectories were used to define road users' speed and acceleration profiles and the longitudinal and lateral distances between road users during the different phases of the interactions. The study demonstrates the use of speed and acceleration profiles and relative spatial profiles in describing the microscopic interaction behaviour between cyclists and pedestrians. The interactions between cyclists and pedestrians were categorized into three categories; (1) interactions with road users moving in the same direction, (2) interactions with road users moving in opposing direction (head-on interactions), and (3) bicycles interaction with crossing pedestrians. The road users applied different collision avoidance mechanisms to avoid collisions during the interactions. The collision avoidance mechanisms involved a change in either the movement direction or walking/cycling speed or both of them to avoid collision with other road users. The results showed that the collision avoidance mechanisms adopted by road users depend on several factors, including the shared space density and the space available for road users.

The results obtained in this study represent the basis for developing an agent-based microsimulation model of cyclist and pedestrian interactions in shared space areas. The study identifies several parameters that can be used to model the microscopic behaviour of cyclists and

pedestrians in the microsimulation models. Table 3.7 summarizes the parameters that describe cyclist-pedestrian interactions in shared space areas and the values reported in the current study. Modeling road user interactions can be beneficial in evaluating the safety and the facility performance under various designs of shared space areas.

Interaction	Evasive action	Parameters needed to model	Values obtained from the current study Mean [SD]	
(1) Same direction movement interaction (following strategy)	• Changing Speed (cyclist)	1. Longitudinal distance at start of reducing speed	4.46 [1.29] m	
		2. Following distance	2.36 [0.5] m	
		3. Cyclist deceleration profile	0.81 [0.40] m/s^2 (average value)	
(2) Same direction movement interaction (overtaking strategy)	• Changing movement direction (cyclist) • Changing speed (cyclist)	1. Longitudinal distance at the start of overtaking maneuver	3.00 [1.65] m	
		2. Lateral distance	1.23 [0.25] m	
		3. Cyclist acceleration profile	0.87 [0.35] m/s^2 (average value)	
(3) Opposing direction movement interaction (head-on interaction)	• Changing movement direction (cyclist or pedestrian or both) • Changing speed (cyclist)	1. Longitudinal distance at the start of the interaction	6.39 [2.31] m	
		2. Lateral distance	1.24 [0.32] m	
		3. Cyclist deceleration profile	1.14 [0.75] m/s^2 (average value)	
		4. Cyclist acceleration profile	1.11 [0.56] m/s^2 (average value)	
(4) Interaction with Crossing road users	• Changing movement direction (cyclist or pedestrian or both) • Changing in cyclist speed • Changing in pedestrian walking speed	1. Longitudinal distance at the start of the interaction	5.95 [2.73] m	
			Cyclist yield to Ped	*Ped yield to cyclist*
		2. Percentage change in pedestrian's speed	25.7%	-63.5%
		3. Cyclist deceleration profile	0.92 [0.42] m/s^2 (average value)	1.21 [0.68] m/s^2 (average value)

Table 3.7 Microscopic behavioural characteristics of cyclist and pedestrian interactions in shared spaces (Alsaleh, et al., 2020)

Chapter 4: Uni-directional Models for Cyclist-Pedestrian Interactions: A Discrete Inverse Reinforcement Learning Approach

4.1 Background

Shared spaces are areas with no clear segregation between the road users, meaning that the right of way is shared between them. Despite the emerging popularity and the wide implementation of shared spaces around the world, only a few studies have analyzed and modeled the behaviour of the road users and the interactions among them in such facilities. The majority of developed simulation models of road users' movement behaviour are concerned with vehicular traffic and road segments with physical separations between different transportation modes. However, little effort was made to develop simulation models for active transportation modes. The CA model and the SFM have been widely adopted in modeling active road users' behaviour (e.g., (Zhou, et al., 2019; Taherifar, et al., 2019; Zhang, et al., 2020)). A few studies extended these models to simulate road users' behaviour and their interactions in mixed traffic conditions (e.g., (Zhao, et al., 2013; Luo, et al., 2015; Jin, et al., 2015; Anvari, et al., 2015; Dias, et al., 2018)).

However, despite considerable efforts in extending these models for modeling road user interactions in shared spaces, these models still show significant discrepancies with real-world road user behaviour (Liu et al. 2020). Such models are limited and not ideal to describe road user behaviour as discussed by several studies (Liu, et al., 2020; Shiwakoti, et al., 2008; Zhang, et al., 2020; Hussein & Sayed, 2017), as they do not consider the fact that road users can logically assess the surrounding environment and take rational decisions. For example, in the social force

model, road users are modeled as particles, and their interactions are modeled using physical forces. In the cellular automata approach, road users move from one cell to another based on predefined rules that mainly depend on the probability of choosing the target cell.

Agent-Based Modeling (ABM) is a relatively new approach for modeling complex systems such as interacting agents and human motion behaviour. This approach considers road users' intelligence and rationality. Developing agent-based models requires modeling agents' goals and strategies (Wooldridge, 1997; Jennings, 2000). A few studies developed models for road user interactions using the ABM approach (e.g., (Fujii, et al., 2017; Hussein & Sayed, 2017)). However, most of the developed rules for road user interactions are heuristic or ad-hoc rules (Papadimitriou, et al., 2009) and suffer from inaccurate and unrealistic representations of road user behaviour. All of these issues highlight the need for more research on developing novel simulation models for road user interactions in shared space facilities.

The work presented in this chapter aims to develop novel microscopic simulation models for cyclist and pedestrian interactions in shared space facilities. Video data were collected at two non-motorized shared space locations of Robson Square in downtown Vancouver, British Columbia. Trajectories of cyclists and pedestrians involved in interactions were extracted, and several variables that describe elements of road users' behaviour were calculated using the methodology presented in the third chapter of this thesis. Two cyclist-pedestrian interactions are considered in the analysis; the following and the overtaking interactions. The road users' behaviour was modeled as utility-based intelligent and rational agents using the finite-state Markov Decision Process (MDP) modeling framework (Bellman, 1957; Sutton & Barto, 1998).

This work focuses on modeling cyclists' operational-level decisions during their interactions with pedestrians in shared space facilities. The operational-level decisions are detailed-level decisions that describe the cyclist's microscopic control decision of accelerating and steering. A microsimulation platform integrating the developed interaction models was developed to simulate road user trajectories in their interactions in shared space facilities. Several accuracy measures were estimated to validate the developed models against actual road user interaction trajectories, relying on their microscopic behaviour parameters.

The MDP models the decision maker's behavior as a sequential decision process in which the decision and the consequent action depend on the current state of the decision-maker and the optimal policy that aims at maximizing its utility or reward function. To solve the MDP and compute the optimal policy (i.e., decision and action sequence), the reward function must first be specified. However, specifying the reward function is very challenging and requires more effort than learning the policy itself. Learning from the demonstration of the task is easier. One approach to solving this problem without the need to extract the reward function is behavioural cloning as an approach of imitation learning, where the main aim is to mimic the action of the expert. However, this approach limits the models' applicability, generality, and accuracy (Bratko, Urbancic and Sammut 1995; Finn et al. 2016). The main shortcomings of this approach are that (1) if the demonstrated agent is not the same as the agent trying to perform the task, the goal now becomes to achieve the outcome the expert achieved instead of mimic the same actions; (2) behavioural cloning provides no reasoning about the outcomes (how the agent achieved the goal

at the end of the task); (3) the expert may have different degrees of freedom on how to accomplish that task (Bratko, et al., 1995).

Another approach for solving this problem is using Inverse Reinforcement Learning (IRL) (Ng & Russell, 2000) to recover/estimate the reward function given expert demonstrations (e.g., real-world road user trajectories that describe road users' decisions and behaviour). This approach provides a tool of reasoning what the expert (i.e., road user) is trying to achieve. Recovering road users' reward functions is crucial for many reasons. First, road users' reward functions infer their goals, objectives, and intentions and provide a tool for analyzing their behaviour and preferences. Second, recovering road users' reward functions helps achieve the generality and transferability of the developed models. It enables the developed models to estimate/ predict road users' preferences on unseen situations (e.g., states). Third, the reward functions are succinct descriptions of road users' strategies.

The contribution of the work presented in this chapter is the recovery and estimation of the reward (utility) functions and optimal policies of road users involved in cyclist-pedestrian interactions in non-motorized shared spaces. Estimating the reward functions is important for several applications, such as the development of Agent-Based microsimulation models (ABM) of cyclists. Two IRL algorithms were used to recover/estimate the reward functions: (1) the Feature Matching (FM) algorithm (Abbeel & Ng, 2004), which assumes optimal road user behaviour or decision process, and (2) the Maximum Entropy (ME) algorithm (Ziebart, et al., 2008), which

assumes sub-optimal behaviour. Moreover, this study inferred and gave insights into road users' preferences and behaviors during their interactions in non-motorized shared spaces.

The following sections in this chapter provide a detailed description of the study locations and the data used for models development and validation, the methodology applied to model cyclist-pedestrian interactions, the various measures used to assess the accuracy of the developed models, the development of the shared space simulation tool, and the results of estimating cyclists' reward functions and policies during their interactions with pedestrians in shared space facilities.

4.2 Study Locations and Data Collection

Video data were collected at two locations of a non-motorized shared space located in Robson Square in downtown Vancouver, British Columbia. Robson Square is an active environment for walking and cycling. The characteristics of Robson Square are described in detail in chapter 3 of this thesis. Video data were obtained for the two locations using two cameras mounted on the edges of the Robson shared space area. The first camera was mounted at the shared space area near Howe Street (Figure 4.1(a)-Figure 4.1(c)), and the video data were obtained for 21 hours over five days in May 2019. The second camera was mounted at the shared space area near Hornby Street (Figure 4.1(b)-Figure 4.1(d)), and the video data were obtained for 18 hours over nine days in August and September 2016, as described earlier in chapter 3.

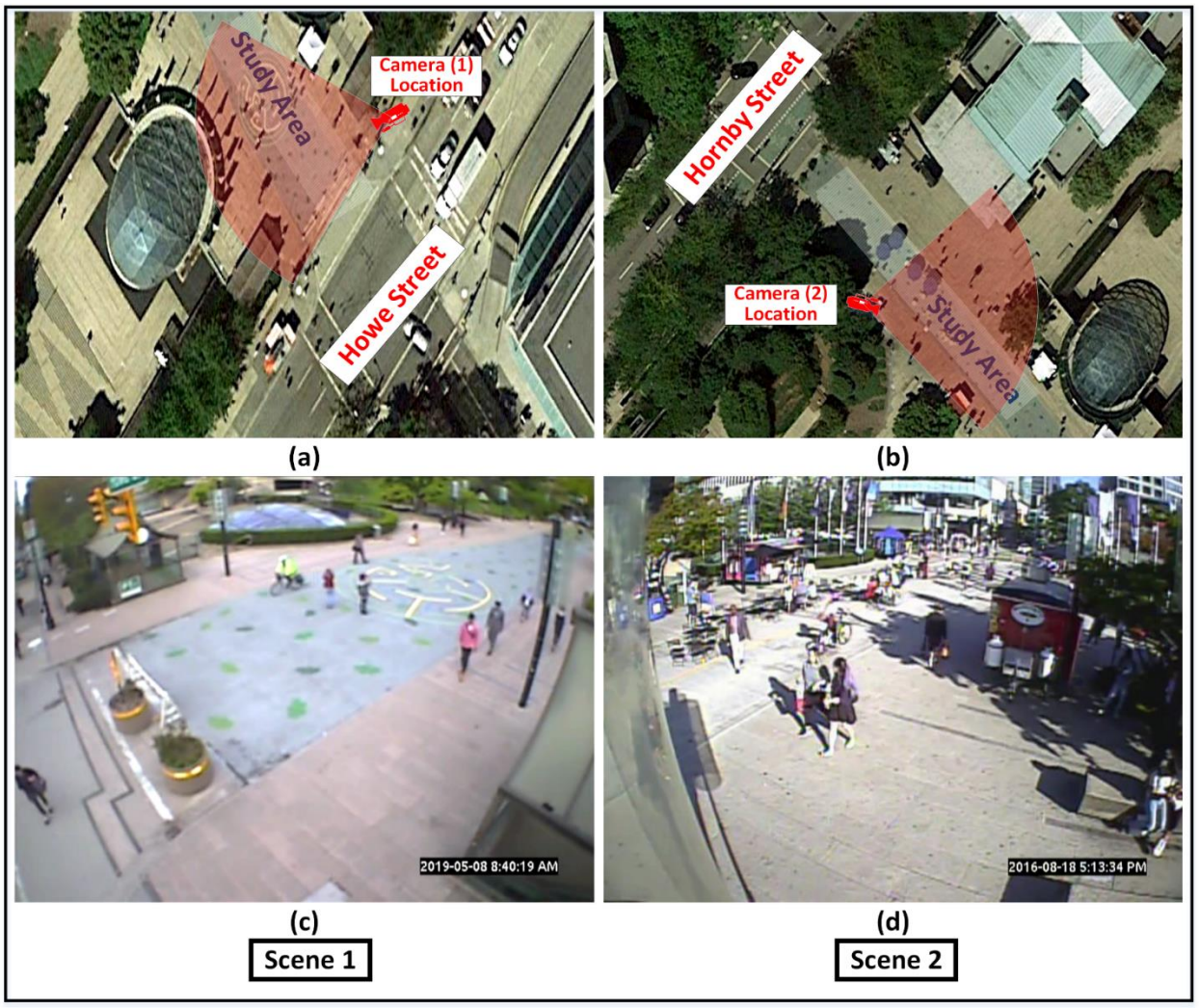


Figure 4.1 Study locations (a) world image for the first study location (scene 1); (b) world image for the second study location (scene 2); (c) camera view for scene 1; (d) camera image for scene 2

4.2.1 Data for Training and Testing

In this study, cyclist and pedestrian trajectories that are involved in the following and overtaking cyclist-pedestrian interactions were extracted during the 39 hours of the analyzed video data of the two locations at the Robson shared space. Cyclist and pedestrian trajectories were extracted using the computer vision methodology described in chapter 3. The demonstration of road user

tracking is shown in Figure 4.2. A total number of 228 and 276 cyclist and pedestrian trajectories involved in following and overtaking cyclist-pedestrian interactions were extracted. Trajectories were extracted each time frame (1/30 seconds) and were associated with 18,376 and 28,068 data points for the following and overtaking interactions, respectively. The extracted data points of each interaction were divided into two sets; the training dataset, which consists of around 80% of the data, while the remaining dataset, e.g., 20% of the data, was used as a testing dataset. For the following interactions, the training dataset consists of 14,698 data points that are associated with 170 trajectories, while the testing dataset consists of 3,678 data points that are associated with 58 trajectories. For the overtaking interactions, the training dataset consists of 22,636 data points that are associated with 222 trajectories, while the testing dataset consists of 5,434 data points that are associated with 56 trajectories.

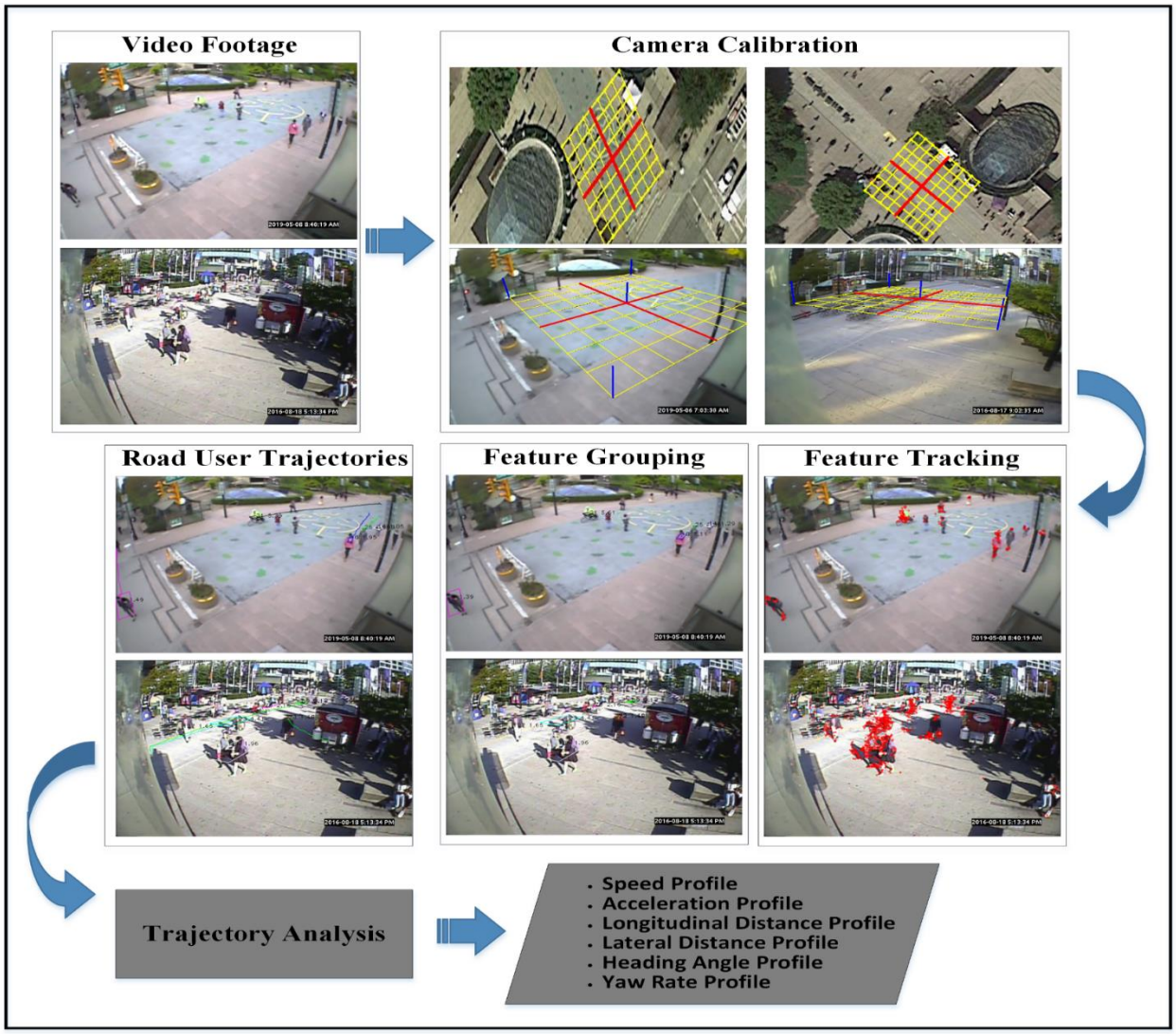


Figure 4.2 Demonstration of road user trajectories' extraction process

4.3 Behavioural Characteristics Extraction

The main set of variables that are used to describe the behaviour of cyclist and pedestrian interactions in shared spaces are based on previous studies of cyclist behaviour (Gavriilidou, et al., 2019; Ma & Luo, 2016), pedestrian behaviour (Hussein & Sayed, 2017; Wang, et al., 2014; Teknomo, 2006), cyclist-pedestrian interactions in shared spaces /paths (Beitel, et al., 2018;

Alsaleh, et al., 2020; Hussein, et al., 2016) and mix traffic interactions (Dias, et al., 2018; Anvari, et al., 2015; Luo, et al., 2015; Gorrini, et al., 2018). The previous studies used several variables to describe the behaviour of cyclists and pedestrians, including longitudinal distance, lateral distance, road user speed, the speed difference between interacting road users, interaction angle, road user acceleration, and yaw rate or changes in steering angle.

Road user trajectories capture the movement of each pedestrian and cyclist in the form of a sequence of spatial coordinates and instantaneous speed at each video frame (1/30 second), as described in chapter 3 of this thesis. The extracted road user trajectories and the derived variables were smoothed for noise using Savitzky-Golay filter (Savitzky & Golay, 1964). The methodology for calculating road user speed and acceleration profiles, longitudinal and lateral distance profiles, and interaction angle are described in detail in chapter 3.

The cyclists usually perform several swerving maneuvers while cycling in shared spaces to overtake slower road users (e.g., pedestrians) or to avoid collision with other road users. The yaw motion around the yaw axis describes the rotation of the cyclist that changes direction to the left or right of its direction of motion. The yaw rate of the cyclist is the angular velocity of this rotation or the rate of change of the heading angle. The yaw rate signal profile is useful in quantifying the swerving maneuvers of the cyclists. Equation (4.1) shows the calculation of the yaw rate (r) as the change of the heading angle Ψ of the cyclist (Ayres, et al., 2004). The road user speed difference (ΔS) is computed as the cyclist and pedestrian speeds difference, as given by Equation (4.2).

$$r(t) = \frac{d\Psi}{dt} \tag{4.1}$$

$$\Delta S = S_{Cyclist} - S_{Pedestrian} \tag{4.2}$$

where t is time, $S_{Cyclist}$ is the speed of the cyclist, and $S_{Pedestrian}$ is the speed of the pedestrian.

4.4 Modeling Methodology

4.4.1 Inverse Reinforcement Learning (IRL)

In this study, the road user decision is modeled as a finite-state Markov Decision Process (MDP). A Markov Decision Process (MDP) consists of a tuple $\left(S, A, P_{ss'}^{a}, \mathcal{R}, \gamma, D\right)$, where S is a finite set of states; $A = \{a_1, \dots, a_k\}$ is a set of k actions; $P_{ss'}^{a}$ is a set of the state transition probabilities; $\mathcal{R} : S \rightarrow A$ is the reward function, $\gamma \in [0,1)$ is a discount factor, which describes how much a given reward is worth on step in the future compared with getting the same reward at current state; and D is the distribution of the initial state. In this study, a discount factor of $\gamma = 0.975$ is used (Alsaleh & Sayed, 2020). Let $MDP \backslash \mathcal{R}$ refers to a Markov Decision Process without a reward function.

In forward Reinforcement Learning (RL), the reward function is known and is used to estimate the optimal agent policy, which maximizes the value of the agent action. However, the problem here is recovering the utility (reward) function that the agent was optimizing given some expert demonstrations (i.e., road user trajectories). Road users act to optimize their reward function, and thus the problem is to find the reward weights $\boldsymbol{\theta}$ that make their demonstrated behavior appear optimal or near optimal. The trajectories of the road users (expert demonstrations) ζ are assumed

to represent the optimal or near-optimal behavior. A trajectory ζ of a road user is a sequence of states and actions and is defined according to Equation (4.3). Figure 4.3 shows the structure of the IRL Problem applied in this paper.

$$\zeta = \{s_0, a_0, \dots, s_i, a_i, \dots, s_T\} \tag{4.3}$$

where s_i and a_i are the road user state and the observed action at a time step $i \in (0, T)$ of the trajectories lifetime (T frames).

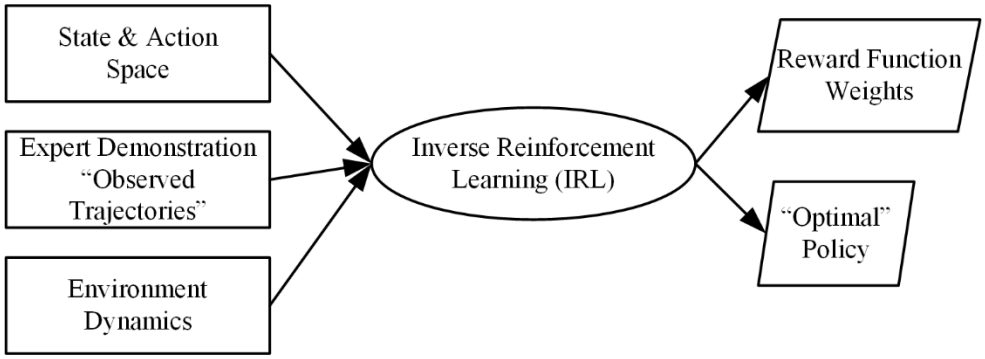


Figure 4.3 Structure of the Inverse Reinforcement Learning (IRL)

It's assumed that there is a true reward function ($\mathcal{R}$) that linearly maps the features of each state f_s to a state reward value which represents the utility of visiting that state. The reward function $\mathcal{R}_\omega{}^*(s) = \omega^T * f(s)$ is parameterized by the reward weights ω, which represent the weights on features over states $f(s)$. The reward is simply the sum of the state rewards (i.e., the reward weighs applied to the path features). A policy $\pi(.; s)$ is a mapping from states to a probability distribution over the action space A. The value of the policy π (V^π) is the sum of the discounted reward and given by Equation (4.4) (Abbeel & Ng, 2004).

$$E_{S_0 \sim \eta}[V^{\pi}(s_0)] = E[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t)|\pi] \quad (4.4)$$

$$E_{S_0 \sim \eta}[V^{\pi}(s_0)] = \omega . E[\sum_{t=0}^{\infty} \gamma^t f(s_t)|\pi] \quad (4.5)$$

The expectation here is taken with a random state sequence draw by starting from state $S_0 \sim D$, and picking and action according to the policy π. The feature expectations value vector $\mu(\pi)$, i.e., the expected discounted accumulated feature value vector, is defined by Equation (4.6). Thus, the expectation of the value of the policy can be redefined as in Equation (4.7) (Abbeel & Ng, 2004).

$$\mu(\pi) = [\sum_{t=0}^{\mathrm{T}} \gamma^t f(s_t) \big| \pi] \quad (4.6)$$

$$E_{S_0 \sim D}[V^{\pi}(s_0)] = \omega . \mu(\pi) \quad (4.7)$$

The expert feature expectation $\mu_E = \mu(\pi_E)$ can be empirically estimated given a set of m trajectories $\{s_0^i, s_1^i, \ldots\}_{i=1}^{m}$ generated by the expert as presented in Equation (4.8) (Abbeel & Ng, 2004).

$$\hat{\mu}_E = \frac{1}{m} \sum_{i=1}^{m} \sum_{t=0}^{T} \gamma^t f(s_t{}^{(i)}) \quad (4.8)$$

In this study, two algorithms were used to recover cyclists' utility (reward) functions during their interactions with pedestrians in shared spaces. The first algorithm is the Feature Matching (FM) IRL algorithm (Abbeel & Ng, 2004). This algorithm assumes a linear reward function that maps the features in each state and optimal expert demonstrations. The algorithm estimates the reward function for which the feature expectation of the policy with respect to the reward function matches the feature counts of the expert trajectories. The second algorithm is the Maximum

Entropy (ME) IRL algorithm (Ziebart, et al., 2008). This algorithm assumes sub-optimal behaviour of expert demonstrations and accounts for the inherent noise and imperfect trajectories. The algorithm uses a probabilistic approach based on the principle of maximum entropy to account for the sub-optimal and imperfect expert trajectories.

4.4.1.1 Feature Matching (FM) IRL Algorithm

In this algorithm, the reward function is estimated for $MDP\backslash\mathcal{R}$ for which the feature expectation of the policy with respect to the linear reward function $\mathcal{R}_\omega{}^*(s) = \omega^T * f(s)$ matches the feature counts of the expert trajectories. The algorithm finds a policy whose performance is close to that of the expert under the unknown linear reward function $\mathcal{R}_\omega{}^*(s)$, i.e., $||\mu(\pi) - \mu_E||_2 < \varepsilon$. The steps of the algorithms to find the policy π is as follow (Abbeel & Ng, 2004):

1. Pick a random policy $\pi^{(0)}$, compute the feature expectation of the policy $\mu^{(0)} = \mu\left(\pi^{(0)}\right)$, and set $i = 1$.
2. Compute $t^{(i)} = max_\omega\ min_{j\in\{0,\dots,i-1\}}\ \omega^T(\mu_E - \mu^{(j)})$, then set $\omega^{(i)}$ to the value of ω that attains this maximum.
3. Terminate if $t^{(i)} < \varepsilon$.
4. Compute the optimal policy $\mu^{(i)}$ for the MDP using rewards $\mathcal{R}_\omega(s) = \omega^T * f(s)$ through reinforcement learning (RL).
5. Compute $\mu^{(i)} = \mu\left(\pi^{(i)}\right)$
6. Set $i = i + 1$, and go back to step 2.

In this study, the value iteration algorithm in RL is used to estimate cyclist policies during the interactions with pedestrians in shared space facilities. The value iteration algorithm can be given

by the simple backup operation of policy improvement as given by Equation (4.9) (Sutton & Barto, 1998). Equation (4.9) can be rewritten as given by Equation (4.10).

$$V_{k+1}(s) = \max_a E[r_{t+1} + \gamma V_k(s_{t+1}) | s_t = s, a_t = a] \quad (4.9)$$

$$V_{k+1}(s) = \max_a \sum_{s'} P_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V_k(s')] \quad (4.10)$$

where $s \in S$, $a \in A$, $\max_a$ is the value of a at which the expression that follows is maximized, and $\mathcal{R}_{ss'}^a$ is the expected reward and equal to $E[r_{t+1} | s_t = s, a_t = a, s_{t+1} = s']$. The optimal policy π^* can b computed as $\pi^* = \arg V^*(s)$, where the superscript * denotes the optimal condition.

4.4.1.2 Maximum Entropy (ME) IRL Algorithm

In the Maximum Entropy (ME) IRL approach, the problem of modeling road user behavior is formulated as a problem of recovering the reward function that makes the behavior induced by a near-optimal or suboptimal policy of road users closely mimic the demonstrated expert behavior. In the ME algorithm (Ziebart, et al., 2008), the optimality of decisions is defined by a binary vector $O_{1:T}$ over a sequence of T decisions. Figure 4.4 shows an illustration of the MDP process containing the optimality vector. The probability of optimality for each state and action is proportional to the reward associated with that state and action as given by Equation (4.11) (Ziebart, et al., 2008).

$$P(O_t | s_t, a_t) \propto e^{\mathcal{R}(s_t, a_t)} \quad (4.11)$$

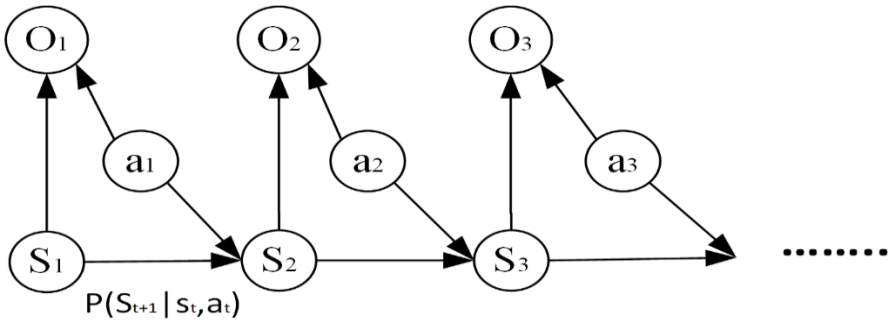


Figure 4.4 Illustration of the optimality vector of the decision-making in MDP (Levine, 2019)

Thus, the probability of a trajectory to be observed in the demonstration dataset given the optimality vector $P(\zeta|O_{1:T})$ is proportional to the probability of the trajectory to occurring times the exponential reward of that trajectory, as given in Equation (4.12). This means that trajectories of equal reward all have the same probability of being executed by the expert, and trajectories with lower reward are exponentially less likely (i.e., the expert can have some noise and not always outputting perfect optimal trajectories). Handling the uncertainty or noise in expert demonstrations in the ME algorithm can potentially lead to obtaining more robust and clean reward functions. Equation (4.12) can be reformulated as given in Equation (4.13) (Ziebart, et al., 2008).

$$P(\zeta|O_{1:T}) \propto P(\zeta) \prod_t e^{\mathcal{R}(s_t,a_t)} = P(\zeta)\, e^{\sum_t \mathcal{R}(s_t,a_t)} \tag{4.12}$$

$$P(\zeta|\omega,\tau) \approx \frac{e^{\omega^T * f(\zeta)}}{Z(\omega,\tau)} \prod_{s_{t+1},a_t,s_{t\in\zeta}} P_\tau(s_{t+1}|a_t,s_t) \tag{4.13}$$

where τ is the transition distribution, and $Z(\omega,\tau)$ is the partition function, which is a normalization constant over all trajectories, and it is defined by Equation (4.14) (Ziebart, et al., 2008).

$$Z(\omega, \tau) = \sum_{i=1}^{m} e^{\mathcal{R}(\zeta_i)} \tag{4.14}$$

where $i \in \{1, \dots, m\}$ is a discrete index for the number of trajectories.

The distribution of the actions of each state over the paths provides a stochastic policy, where the probability of selecting an action is proportional to the sum of all probabilities of taking paths begin with that action, as given by Equation (4.15) (Ziebart, et al., 2008).

$$P(a|\omega, \tau) \propto \sum_{\zeta: a \in \zeta_t} P(\zeta|\omega, \tau) \tag{4.15}$$

In this algorithm, the objective of the expert, instead of being maximizing the reward, is to maximize the difference between the expectation of the reward under the policy and the entropy of that policy. The entropy part in the objective function accounts for the uncertainty and noise in expert demonstrations. Estimating the reward function parameters ω is obtained by maximizing the likelihood of the expert demonstrations under the maximum entropy distribution as presented in Equations (4.16)-(4.17) (Ziebart, et al., 2008).

$$\omega^* = argmax_{\omega}\big(L(\omega)\big) = argmax_{\omega} \frac{1}{m} \sum_{i=1}^{m} log\, P(\zeta_i|\omega, \tau) \tag{4.16}$$

$$\omega^* = argmax_{\omega} \frac{1}{m} \sum_{i=1}^{m} \mathcal{R}(\zeta_i) - \log Z(\omega, \tau) \tag{4.17}$$

where $L(.)$ is the likelihood function, and m is the number of observed trajectories.

In this study, the reward weights were estimated using a reward function in the linear form with intercept, taking the first level as a reference level of each feature, as shown in Equation (4.18).

$$\mathcal{R} = Intercept + \omega^T . f_{level} \quad (4.18)$$

where $\mathcal{R}$ is the reward function, $Intercept$ is the intercept assuming the first level is the reference level for each feature, ω^T is the weight vector estimated for the features relative to the reference level for each feature, and f_{level} is a dummy variable specifying the levels of each feature.

4.4.2 Evaluation Metrics

The basis of the trajectory error computation is the distance between the predicted (simulated) trajectories and the actual trajectories in the validation dataset. Two performance metrics are used to compare the simulated and actual trajectories as follow:

1. *Mean Absolute Error (MAE).* The MAE measures the average magnitude of error in the simulated trajectory over the trajectory lifetime (n frames). Equation (4.19) shows the calculation of the MAE for a simulated trajectory.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | u_{sim,i} - u_{actual,i} | \quad (4.19)$$

where u_{sim} and u_{actual} are the simulated and the actual trajectories, respectively, and $i = \{1, \dots, n\}$ is a discrete temporal index of the trajectory length.

2. *Hausdorff Distance (HausD).* The Hausdorff distance measures the degree of mismatch (deviation) between the simulated and the actual trajectories. This measure computes the largest distance between the simulated and the actual trajectories while ignoring the time step alignment. This measure relaxes the penalty for errors caused by the time step offset but

emphasizes the overall parameter displacement between the simulated and the actual trajectories. Equation (4.20) shows the calculation of the Hausdorff distance between a finite point set of a simulated trajectory $A = \{a_1, \dots, a_n\}$ and an actual trajectory $B = \{b_1, \dots, b_n\}$ (Huttenlocher, et al., 1993; Rockafellar & Wets, 2009).

$$H(A, B) = \max\{h(A, B), h(B, A)\} \tag{4.20}$$

where

$$h(A, B) = \underset{a \in A}{max} \left(\underset{b \in B}{min} ||a - b|| \right) \tag{4.21}$$

$$h(B, A) = \underset{b \in B}{max} \left(\underset{a \in A}{min} ||b - a|| \right) \tag{4.22}$$

and || . || is the Euclidean norm (distance) between the points of A and B. The function $h(A, B)$ is the directed Hausdorff distance from A to B, which identifies the largest distance of $a \in A$ from any point of $b \in B$, and measures the distance from $a \in A$ to its nearest neighbour in $b \in B$ using the Euclidean norm. Thus, $h(A, B)$ ranks each point of $a \in A$ based on its distance to the nearest point in B, and then use the largest ranked point distance, as it represents the most mismatched point of A. Similarly, $h(B, A)$ is the directed Hausdorff distance from B to A. The Hausdorff distance $H(A, B)$ is the maximum of $h(A, B)$ and $h(B, A)$.

4.5 IRL Algorithm Implementations: Analysis and Results

In this study, the training dataset was used to estimate the reward function weights for each type of interaction. The estimated optimal policies were used to simulate road user trajectories, which were validated using the validation dataset.

4.5.1 Expert Demonstrations and State and Action Discretization

Cyclist and pedestrian trajectories involved in the following and overtaking interactions were analyzed separately and were used to estimate the variables (features) that describe the cyclist state and action at each time step. Most of the interactions analyzed in this study were between a single pedestrian/ cyclist pair and took place in low shared space densities. Five features were used to describe the state of the cyclist at each time step, including the longitudinal distance, the lateral distance, the angle difference between the cyclist and the pedestrian, cyclist speed, and speed difference between the cyclist and the pedestrian. The action of the cyclist at each time step is defined by two variables, including the cyclist acceleration and yaw rate (i.e., the change in cyclist steering angle between the next time step and the current time step). Analysis of these profiles revealed multiple characteristics of cyclists and pedestrians in the following and overtaking interactions, as shown in Figure 4.5. Descriptive statistics for these characteristics are presented in Table 4.1. It can be seen that the average cyclist speed and the cyclist-pedestrian speed difference are higher for the overtaking interaction compared to the following interaction. The overtaking interaction is associated with a larger average lateral distance and a smaller average longitudinal distance than the following interaction. Regarding the cyclist action, the average acceleration and yaw rate are higher for the overtaking interaction compared to the following interaction.

The space of the state features was discretized for each interaction type by dividing each state feature into six levels based on equal frequency observation in each level. The space of the action was discretized for each interaction type by dividing the acceleration into five levels based

on equal frequency observation in each level, and dividing the yaw rate into five equal intervals length. The number of levels of the state features was selected based on a trade-off between the policy prediction accuracy and the computational cost, i.e., CPU computational time and required RAM. This discretization results in a number of states of $6^5 = 7776$ states, and a number of actions of $5^2 = 25$ actions. The discretization intervals of the state and action for the following and overtaking interactions are represented in Table 4.2, and Table 4.3, respectively. Negative values for the angle difference (state feature) or change in cyclist direction (action feature) indicate a counter-clockwise angle. Negative speed difference values indicate higher pedestrian speed values. Negative longitudinal distance values indicate that the cyclist has overtaken the pedestrian.

Variable	Following Interaction mean [SD*]	Overtaking Interaction mean [SD]
Cyclist Speed [m/s]	2.099 [0.993]	3.697 [1.852]
Pedestrian Speed [m/s]	1.541 [0.647]	1.392 [0.547]
Speed Difference [m/s]	0.558 [1.130]	2.305 [1.973]
Angle Difference [rad]	0.013 [0.507]	0.0789 [0.688]
Lateral Distance [m]	1.452 [1.369]	2.587 [1.880]
Longitudinal Distance [m]	5.132 [3.219]	3.611 [5.223]
Cyclist Acceleration [m/s^2]	0.057 [1.533]	0.1772 [2.216]
Cyclist Yaw Rate [rad/s]	-0.012 [3.167]	0.111 [3.438]

* standard deviation

Table 4.1 Data descriptive statistics for the following and overtaking interactions (Alsaleh & Sayed, 2020)

State Level	Longitudinal Distance [m]	Lateral Distance [m]	Angle Difference [rad]	Cyclist Speed [m/s]	Speed Difference [m/s]
1	$(-\infty^*, 2.27)$	$(-\infty^*, 0.35)$	$(-\infty^*, -0.396)$	$(-\infty^*, 1.32)$	$(-\infty^*, -0.394)$
2	[2.27, 3.59)	[0.35, 0.657)	[-0.396, -0.164)	[1.32, 1.64)	[-0.394, 0.0399)
3	[3.59, 4.54)	[0.657, 1.06)	[-0.164, 0.00565)	[1.64, 1.87)	[0.0399, 0.381)
4	[4.54, 5.71)	[1.06, 1.67)	[0.00565, 0.169)	[1.87, 2.24)	[0.381, 0.767)
5	[5.71, 7.48)	[1.67, 2.58)	[0.169, 0.397)	[2.24, 2.77)	[0.767, 1.44)
6	[7.48, ∞^*)	[2.58, ∞^*)	[0.397, ∞^*)	[2.77, ∞^*)	[1.44, ∞^*)

Action Level	Acceleration [m/s^2]	Yaw Rate [rad/s]
1	$(-\infty^*$,-0.876)	$(-\infty^*$,-23.6)
2	[-0.876,-0.153)	[-23.6,-10.3)
3	[-0.153,0.281)	[-10.3,2.95)
4	[0.281,0.902)	[2.95,16.2)
5	[0.902, ∞^*)	[16.2, ∞^*)

* $(-\infty^*)$ and (∞^*) represents the possible smallest and largest rational values.

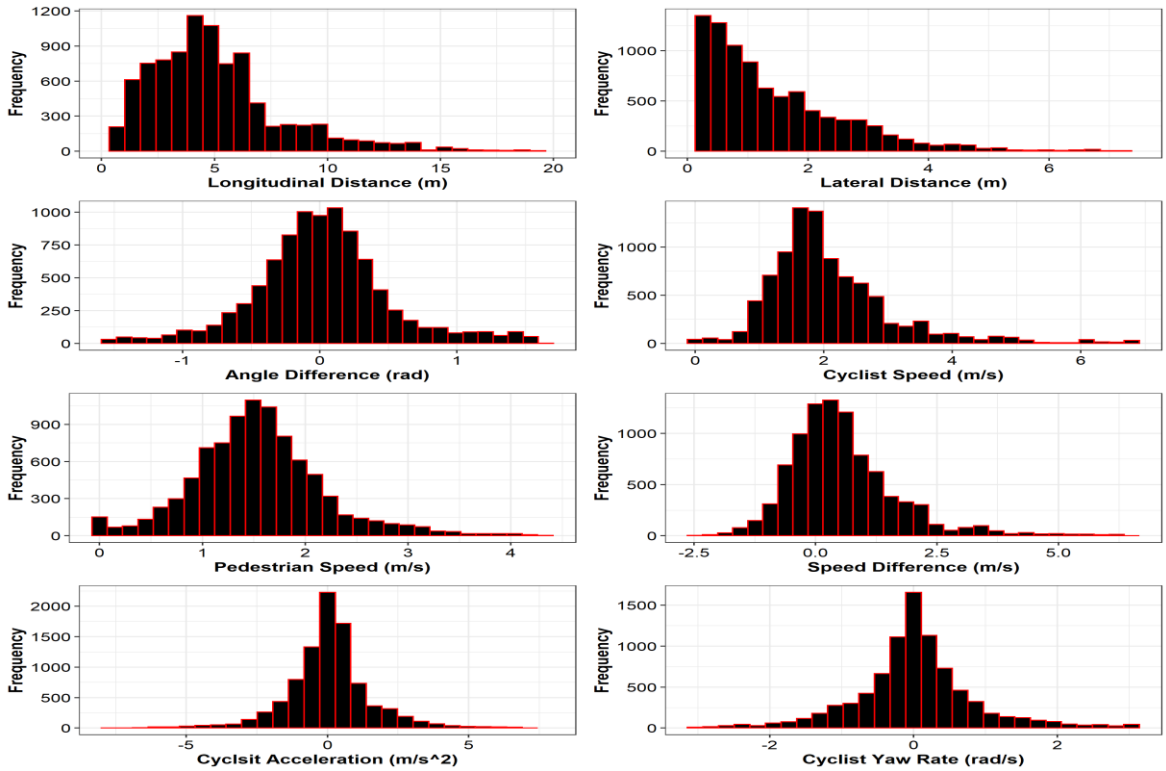
Table 4.2 States and actions discretization intervals for the following interaction (Alsaleh & Sayed, 2020)

State Level	Longitudinal Distance [m]	Lateral Distance [m]	Angle Difference [rad]	Cyclist Speed [m/s]	Speed Difference [m/s]
1	$(-\infty^*$, -1.44)	$(-\infty^*$, 0.963)	$(-\infty^*$, -0.47)	$(-\infty^*$, 2)	$(-\infty^*$, 0.483)
2	[-1.44, 1.36)	[0.963,1.78)	[-0.47,-0.132)	[2, 2.66)	[0.483,1.38)
3	[1.36, 3.47)	[1.78,2.42)	[-0.132,0.0546)	[2.66, 3.4)	[1.38,2.02)
4	[3.47, 5.59)	[2.42,2.9)	[0.0546,0.287)	[3.4, 4.23)	[2.02,2.78)
5	[5.59, 8.18)	[2.9,3.91)	[0.287,0.678)	[4.23, 5.47)	[2.78,3.98)
6	[8.18, ∞^*)	[3.91, ∞^*)	[0.678, ∞^*)	[5.47, ∞^*)	[3.98, ∞^*)

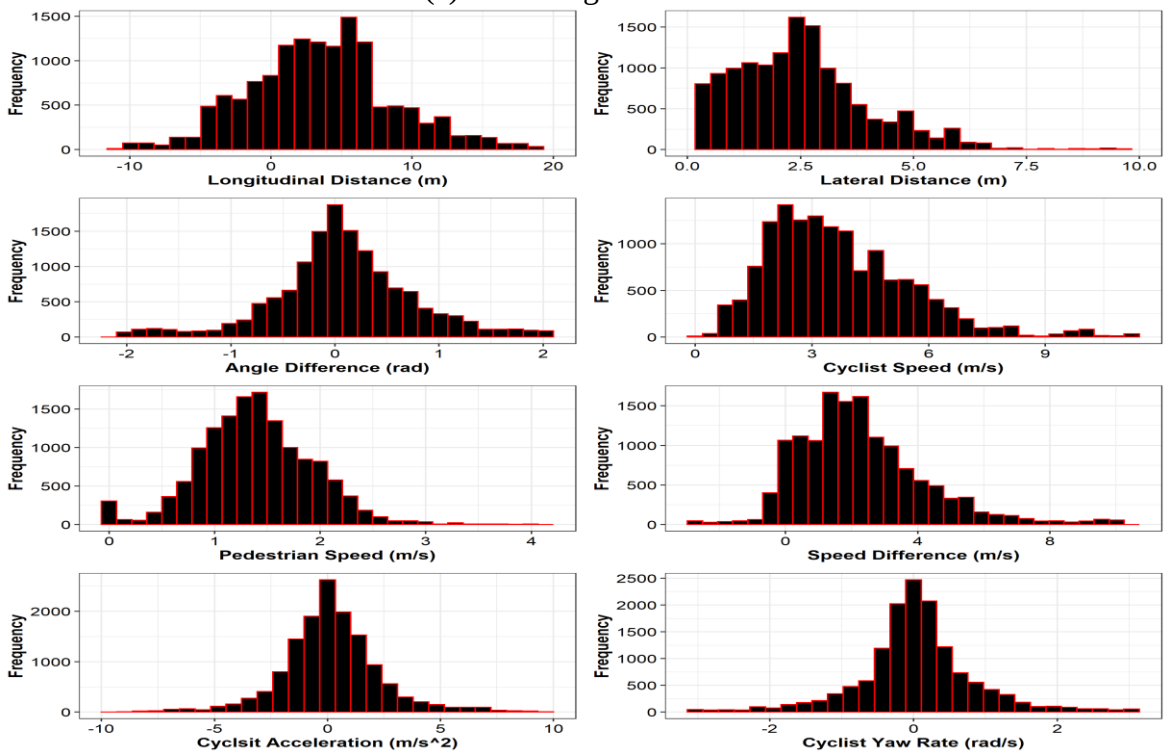
Action Level	Acceleration [m/s^2]	Yaw Rate [rad/s]
1	$(-\infty^*$, -1.29)	$(-\infty^*$, -20.5)
2	[-1.29, -0.276)	[-20.5,-6.71)
3	[-0.276, 0.467)	[-6.71,7.11)
4	[0.467, 1.59)	[7.11,20.9)
5	[1.59, ∞^*)	[20.9, ∞^*)

* $(-\infty^*)$ and (∞^*) represents the possible smallest and largest rational values.

Table 4.3 States and actions discretization intervals for the overtaking interaction (Alsaleh & Sayed, 2020)



(a) Following Interaction



(b) Overtaking Interaction

Figure 4.5 Characteristics of the cyclist and pedestrian behaviour in the following and overtaking interactions (Alsaleh & Sayed, 2020)

4.5.2 Reward Function Weights Estimation

4.5.2.1 Cyclist-Pedestrian Following Interaction

The estimated reward function weights for the following interaction using the Maximum Entropy (ME) and Feature Matching (FM) algorithms (Levine, et al., 2011) are presented in Figure 4.6. The estimated reward function weights give insights about cyclists' preferences during their interaction with pedestrians in shared spaces. The estimated reward function weights for each state feature are estimated relative to the first level (level 1) of that feature. Reward function estimates suggest that states containing lateral distances in the largest-lateral distance level 6 have lower reward than the states containing lateral distances in the reference level 1, indicating a low value of being in a large-lateral distance level. Lateral distances in level 2 have the highest reward value, while reward weights for lateral distances levels 3 and 4 are slightly lower, suggesting a preferred lateral distance in level 2 [0.35, 0.657 m) according to the ME algorithm. However, the FM algorithm suggests that cyclists highly prefer intermediate lateral distances (level 4) which has a range of [1.06, 1.67 m). Both FM and ME algorithms predict similar behavior except for lateral distances levels 2 and 3.

For the longitudinal distance between cyclists and pedestrians, the estimated reward function weights from both the FM and ME algorithms suggest that cyclists have two preferences which are either to be in states with low-longitudinal distance (level 2) or intermediate/high-longitudinal distance (levels of 4 and 5 for ME and level 6 for FM). The first preference is associated with the cases of following at relatively higher shared space density conditions where the average following distance was measured to be 2.36 m according to a shared space behavioral study (Alsaleh, et al., 2020). This case is suggested to be associated with moderate cyclists' preferences to be in states with angle difference

level 3, which has a range around zero. However, the other preference is suggested to be associated with a low density shared space condition where cyclists prefer to keep moderate longitudinal distance with pedestrians and prefer to change the steering angle slightly to avoid the pedestrians, which is shown in cyclists preferences to have an angle difference with pedestrians deviated from the zero, i.e., angle difference level 5 (9-22 degrees (deg)) for ME and level 1 (<-22 deg) for FM algorithms. Both FM and ME algorithms generally predict similar behavior except for levels 4 and 5 for longitudinal distances and levels 5 and 6 for angle difference.

For the cyclist speeds, both the FM and ME algorithms suggest that cyclists do not prefer to have very low speed; instead, they prefer to be in states with high-speed levels. The cyclists highly prefer to be in speed levels of 4 and 5 [1.87 to 2.77 m/s) according to ME algorithm, and speed level 6 (>2.77 m/s) according to FM algorithms. For the speed difference, both FM and ME algorithms generally suggest that cyclists do not prefer to be very slow compared with pedestrians. The ME algorithm suggests that cyclists prefer to be within the speed of the hindered pedestrians as the preference is to be in speed difference levels 2 and 3, which have a speed difference range from -0.394 to 0.381 m/s. The FM algorithm is less consistent across levels as it suggests that cyclists prefer to have intermediate speed difference levels 3 and 4; however, their highest preference to be in high-speed difference level 6. The results of the ME algorithm are generally consistent with the observational behavioral study (Alsaleh, et al., 2020) as the cyclists in the following interactions try to maintain following distance with the hindered pedestrians with small fluctuation in their speeds. Overall, the ME reward function estimates are more stable compared with the FM estimates. The estimated reward function using the Feature Matching (FM) algorithm has a higher intercept to

parameter weight ratio compared to the estimated weights using the Maximum Entropy (ME) algorithm. This can indicate that the estimated reward function using FM algorithm did not adequately learn the following interaction behavior compared with the estimates of the ME algorithm, which can limit the prediction accuracy and the transferability of the FM model.

A visualization of the reward function value estimates over states from applying both the ME and the FM algorithms are presented in Figure 4.7. The figures show the differences in the reward value across the different states based on the feature values of each state. The ME reward function figure shows that having low lateral distance (level 2) combined with a low or intermediate longitudinal distance (level 2, levels 4, or 5) provides the highest reward for the ME algorithm compared to other combinations of lateral and longitudinal distances. However, the FM function figure shows that having an intermediate lateral distance (level 4) combined with a low or large longitudinal distance (level 1 or 6) provides the highest reward. Moreover, the figures show that having intermediate/high cyclist speed (levels 4 and 5), e.g., cyclist speed range from 1.87 to 2.77 m/s, combined with an intermediate speed difference (levels 2 and 3) provides the highest reward for the estimates of ME algorithm comparing with having very low or very high cyclist speed with low or high speed difference. However, having high cyclist speed (level 6) combined with high speed difference (level 6) provides the highest reward for the estimates of FM algorithm. The states associated with an angle difference around zero (levels 3) and angles deviated from zero (level 5) or (level 1) have higher rewards compared with the rewards associated with other states for the estimates of ME and FM, respectively.

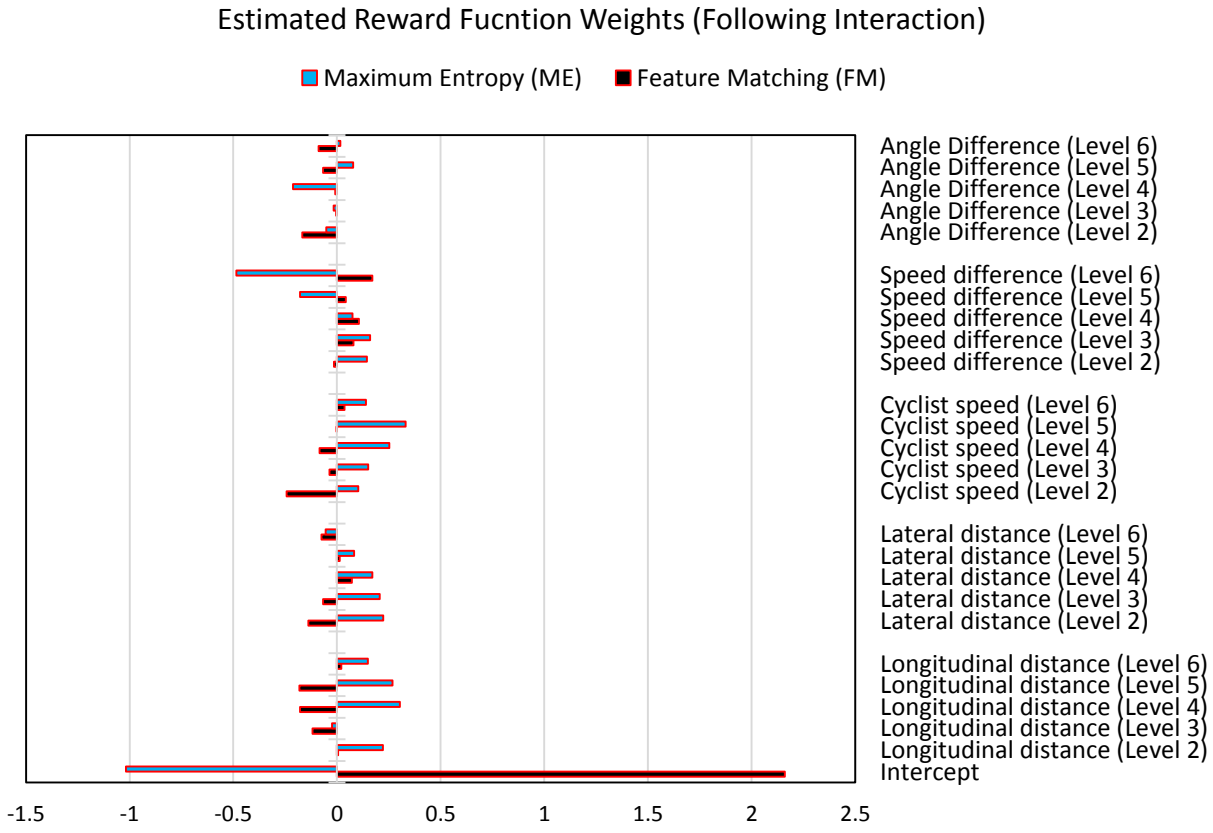


Figure 4.6 Estimated reward function weights for the following interaction (Alsaleh & Sayed, 2020)

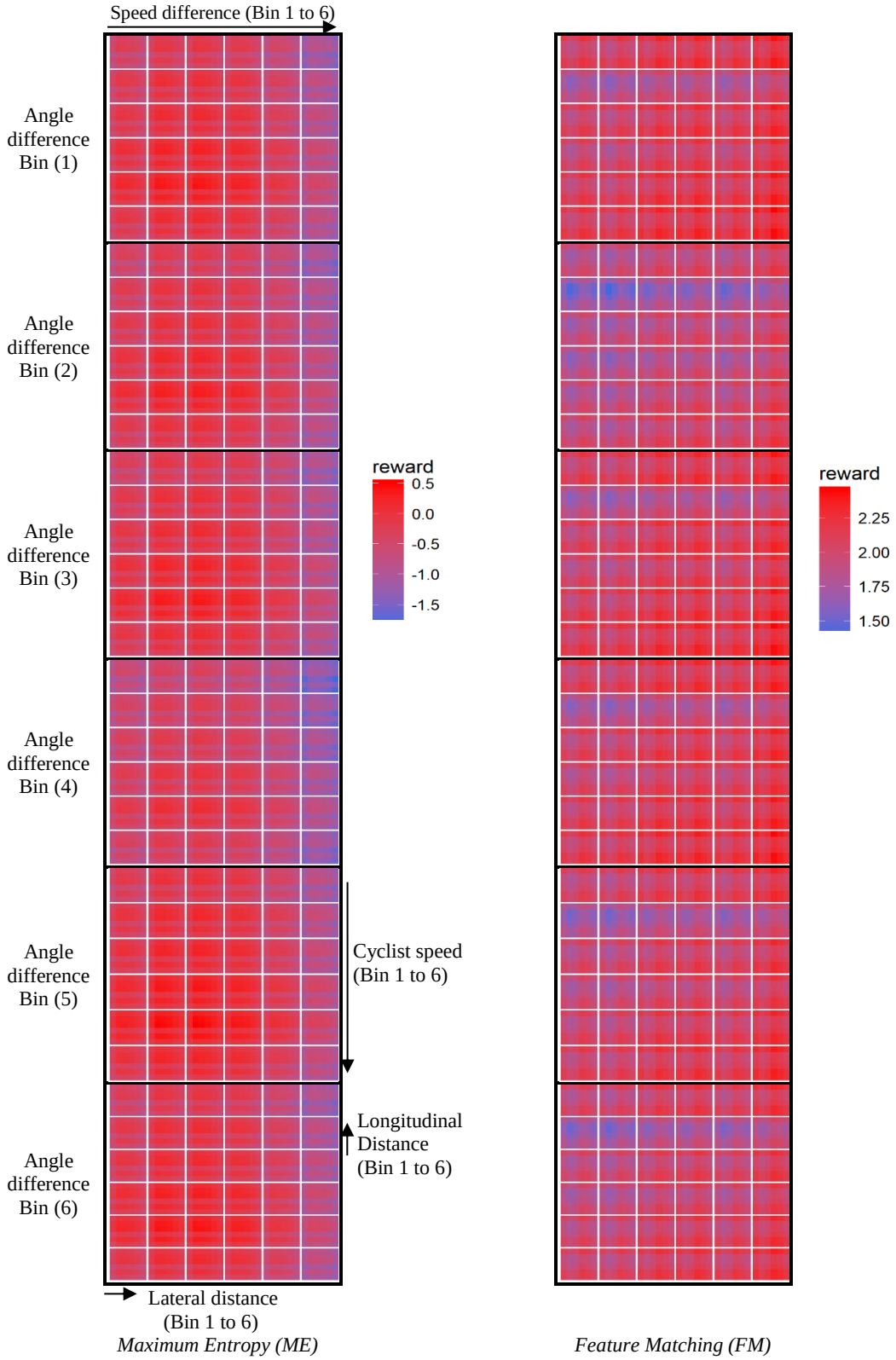


Figure 4.7 Visualization of the estimated reward function for the following interaction (Alsaleh & Sayed, 2020)

4.5.2.2 Cyclist-Pedestrian Overtaking Interaction

The estimated reward function weights for the overtaking interaction using the Maximum Entropy (ME) and Feature Matching (FM) algorithms are presented in Figure 4.8. Similar to the previous analysis, the estimated reward function weights for each state feature are estimated relative to the first level (level 1) of that feature. The ME reward function estimates suggest that states containing intermediate lateral distances have higher rewards than the states containing lateral distances in the reference level 1, indicating a low value of being in the lowest-lateral distance level. Lateral distances in the intermediate levels of 2 to 5 [0.963, 3.91 m) have higher reward values than the extreme largest lateral distance of level 6, with the highest reward value for lateral distance level 5, suggesting a preferred intermediate lateral distance, especially in level 5 [2.9, 3.91 m). This result is expected as the overtaking maneuver involves swerving and increase of lateral distance to overtake slower pedestrians. However, the FM algorithm suggests that cyclists prefer lateral distances in the low lateral distance level 1 [0, 0.963 m) with a lower preference for lateral distance in the intermediate level 5. The algorithms predict different lateral distance preferences.

For the cyclist speeds, the ME algorithm reward weights indicate that states that have speed in the intermediate and high speed levels have higher reward weights than the states containing the speed reference level 1. This suggests that cyclists prefer to be in states with intermediate to high speed levels of 4 to 6, which have a speed range of (>3.4 m/s). However, the FM estimates suggest that cyclists prefer to be in states with very high cycling speed (level 6) compared to the intermediate speed level 4, while their highest preference is to be in the low speed reference level 1. For the speed difference, both the FM and ME algorithms generally suggest that cyclists do not prefer to be much

faster than pedestrians as they assign the lowest weight of being in the highest speed difference level 6. The ME and FM algorithms assign high weights of being in speed difference level 2 compared with the lowest speed difference reference level 1, suggesting that cyclists prefer to be faster than pedestrians within the speed difference level 2, which has a speed difference range from 0.483 to 1.38 m/s. However, the FM algorithm is less consistent across states and assigns a relatively higher weight of being in speed difference level 5 [2.78, 3.98 m/s) compared to the speed difference level 2, suggesting a slightly higher preference in being in speed difference level 5 compared with level 2 according to FM algorithm.

For the longitudinal distance between cyclists and pedestrians, the estimated reward function weights from the ME algorithms show that states containing longitudinal distances in the large longitudinal distance levels have lower weights than longitudinal distances in the reference level 1, which have a range of (< -1.44 m), indicating a low reward value of being in large-longitudinal distance states. This suggests that cyclists prefer to overtake pedestrians with their highest preference to be in states with longitudinal distance level 1 (< -1.44 m), where the negative value of the longitudinal distance indicates that cyclists have overtaken pedestrians. Similarly, the FM algorithm assigns lower weights of being in high longitudinal distance levels 4, 5, and 6 compared to the low longitudinal distance level 1. However, the FM algorithm is inconsistent across levels and assigns a higher weight of being in longitudinal distance level 3 [1.36, 3.47 m) compared to longitudinal reference level 1.

For the direction (angle) difference between the cyclists and pedestrians, the reward function weights of the ME algorithm assigns higher reward weights for states that contain angle difference levels of 2,

3, and 5, which have ranges of [-27, -7 deg), [-7, 3 deg), and [16, 38 deg), respectively. This suggests that cyclists prefer to be within these angle difference levels. The angle difference levels of 2 and 5 are associated with cyclists during the overtaking maneuvers where the overtaking takes place from the left (counter-clockwise) or the right (clockwise), respectively. The angle difference of level 3 is suggested to be associated with the end of the overtaking maneuvers as the cyclist return to the original heading direction. Similarly, the reward functions estimated using the FM algorithm assign higher reward weights for states containing angle difference levels of 3 and 5, suggesting that cyclists prefer to be within these angle difference levels. However, the FM algorithm agrees with ME preferences except for the angle difference level 2. Similar to the modeling of the following interaction, the ME algorithm provides more stable estimates of the reward function weights, while the FM algorithm provides less consistent reward weights. Moreover, the estimated reward function using the Feature Matching (FM) algorithm has a higher intercept to parameter weight ratio compared with the estimated weights using the Maximum Entropy (ME) algorithm, which can limit the prediction accuracy and the transferability of the FM model.

A visualization of the reward function value estimates over states from applying both the ME and FM algorithms is presented in Figure 4.9. The ME reward function figure shows that having intermediate to high lateral distances (level 4, 5, and 6) combined with a low longitudinal distance (level 1) provides the highest reward for the ME algorithm when compared to states of having very low or very large lateral distances and large longitudinal distances. In contrast, the FM reward function figure shows that having very low lateral distances (level 1) with intermediate longitudinal distances (level 3) provides the highest reward. Moreover, the figures show that having high cyclist speed

(levels 5 and 6 for ME) combined with speed difference (levels 2 for ME) provides the highest reward value compared to other combinations of speed and speed difference. The FM Figure shows that having low cyclist speed (level 1) combined with low or high speed difference (level 2, 3, or 5) provides the high reward value. The states associated with an angle difference in levels 2, 3, and 5 for ME and FM (except for level 2 for FM) have higher rewards compared with the states associated with other levels.

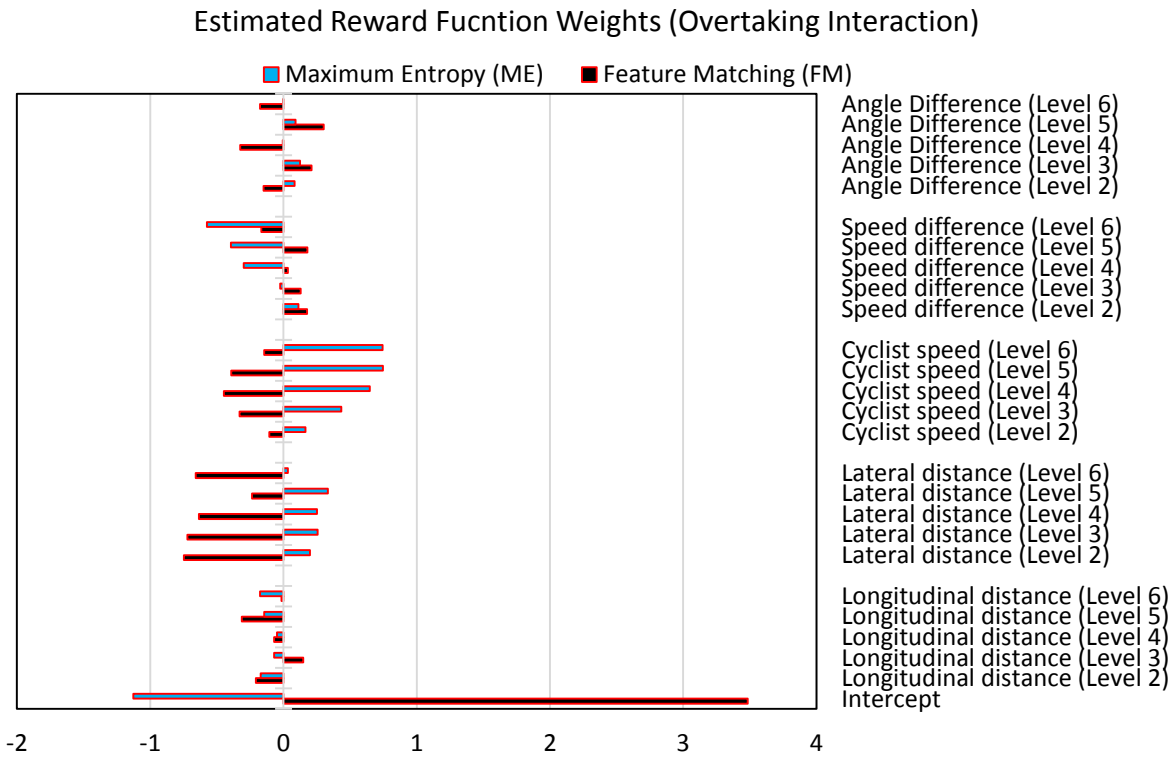


Figure 4.8 Estimated reward function weights for the overtaking interaction (Alsaleh & Sayed, 2020)

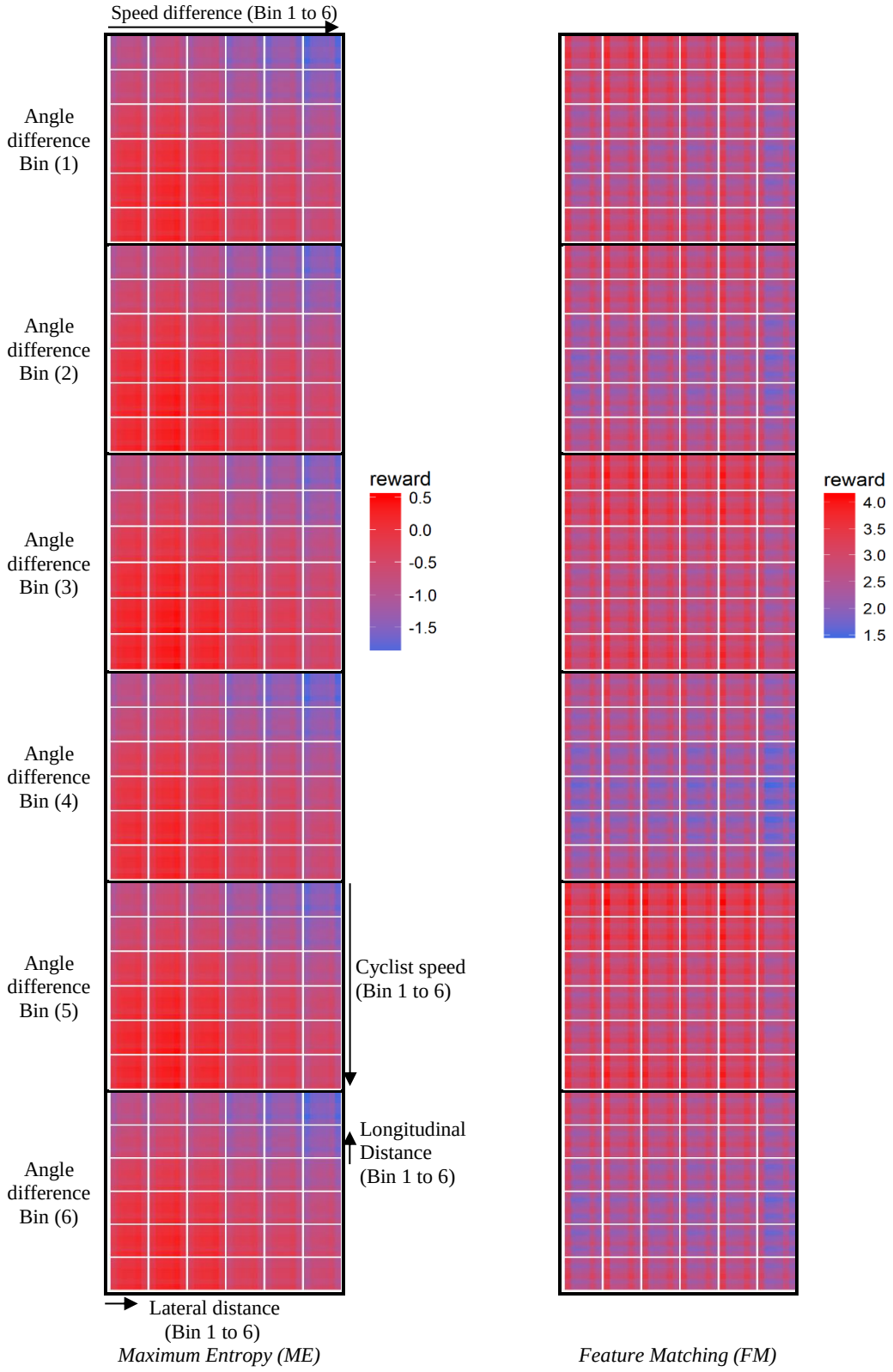


Figure 4.9 Visualiation of the estimated reward function for the overtaking interaction (Alsaleh & Sayed, 2020)

4.6 Trajectories Prediction

In this study, a simulation tool was developed using R-software (R Core Team, 2018) to simulate the cyclist trajectories given the surrounding environment of the validation data set, which includes unconstrained pedestrian flow. The simulation was run at a resolution of 30 HZ (1/30 sec). Figure 4.10 shows the workflow of the simulation tool/code. The simulation tool first initializes the simulation environment that includes the pedestrian flow and their behaviour over time. The stimulating agent (e.g., cyclist) is then initialized with information about its initial position, speed, and yaw angle. Then, the simulation tool calculates the agent state variables based on the surrounding environment and assesses the state of the cyclist. The cyclist then takes an appropriate action based on the estimated optimal policy each time step. The average action values (acceleration and yaw rate) in each action interval (Table 4.2 and Table 4.3) were used to represent the action for that interval for each type of interaction. The cyclist position, speed, and yaw angle are then updated each time step based on the motion equations.

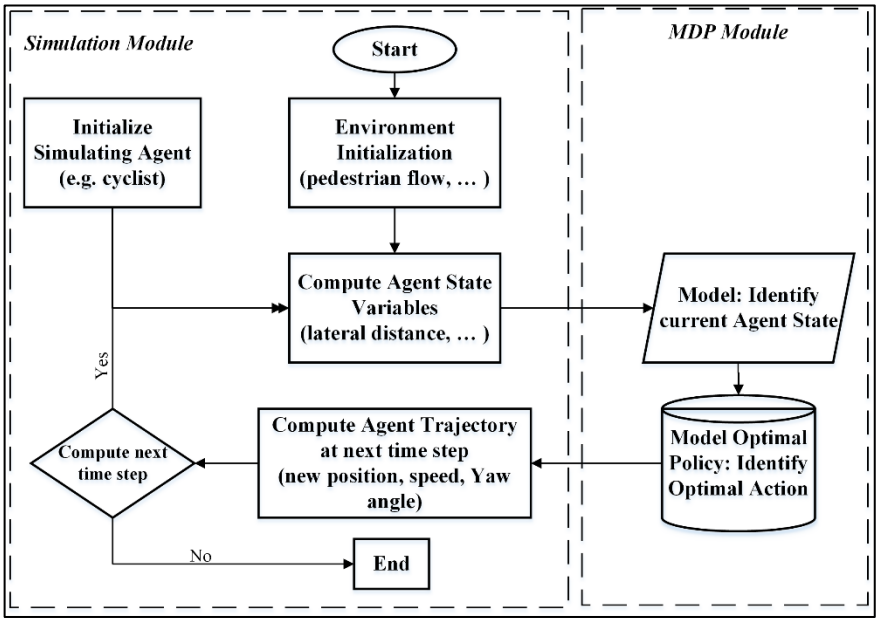
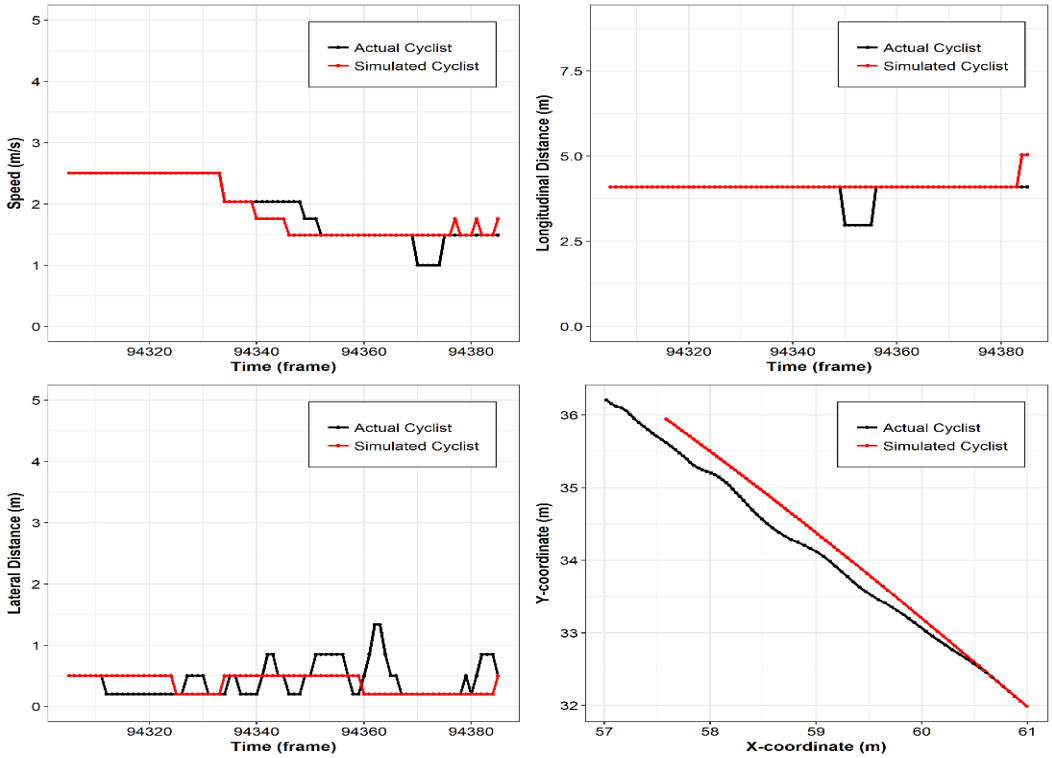


Figure 4.10 Discrete simulation tool workflow for cyclist-pedestrian interactions (Alsaleh & Sayed, 2020)

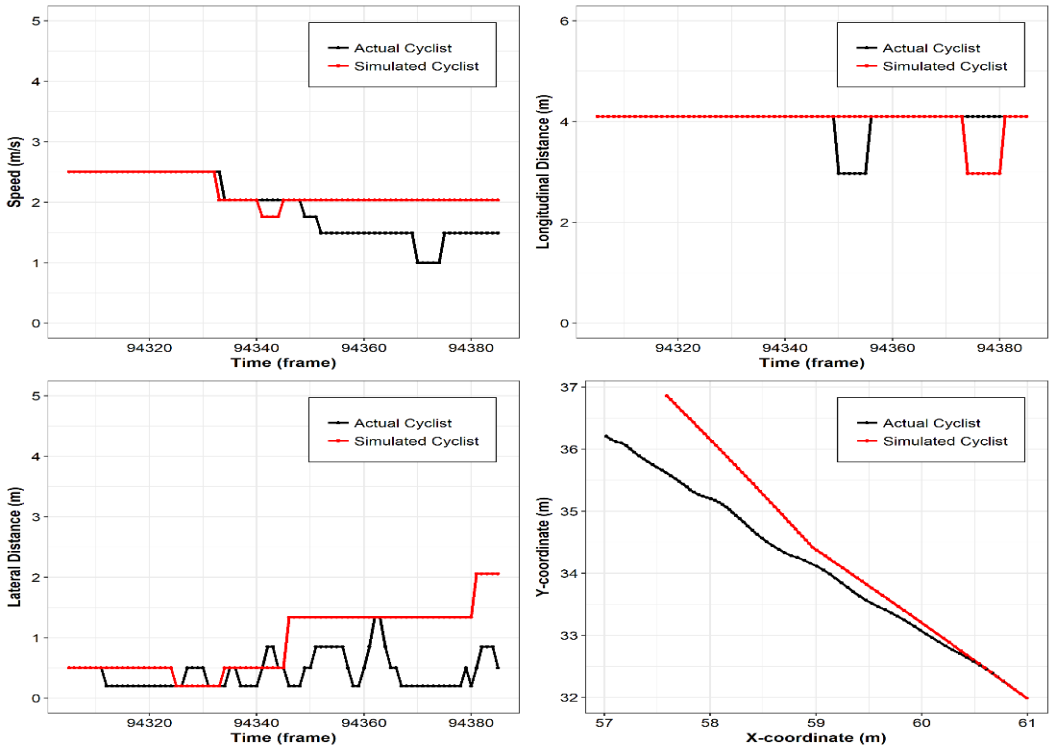
The cyclist trajectories in the validation datasets were simulated using the developed models from applying the two FM and ME IRL algorithms. The differences between the actual and simulated trajectories were evaluated using the mean absolute error and the Hausdorff distance (*HausD),* as shown in Table 4.4. Overall, the ME algorithm shows better accuracy and less dissimilarity between the actual and simulated trajectories for both the following and overtaking interactions. The ME algorithm also produces more accurate estimates of cyclist speed and longitudinal and lateral distances compared to the FM algorithm. Both algorithms predict the cyclist speed more accurately than the cyclist position. For the following interaction, the ME algorithm achieved an average improvement in predicting cyclist speed by about 12.3% (10.6% using *HausD*) compared to the FM algorithm. The corresponding average improvements in the prediction of the longitudinal and lateral distances are more pronounced and equal to 19.5% (28.7% using HausD) and 20.7% (22.9% % using *HausD*). For the overtaking interaction, the ME algorithm shows an average improvement in predicting cyclist speed by about 20.5% (27.9% using *HausD*) compared to the FM algorithm. The corresponding average improvements in the prediction of the longitudinal and lateral distances are 17.3% (21.6% using *HausD*) and 20.6% (17.1%), respectively. Examples of actual and simulated trajectories and their corresponding speed profiles and longitudinal and lateral distances' profiles from applying both the FM and ME IRL algorithms for the following and overtaking interactions are presented in Figure 4.11 and Figure 4.12 (as discrete intervals), respectively. As shown in the figures, the ME model is capable of reproducing more accurate following and overtaking behaviour compared to the FM models.

Interaction Type	Variable	Maximum Entropy (ME)		Feature Matching (FM)	
		Avg. MAE	Avg. Hausdorff distance	Avg. MAE	Avg. Hausdorff distance
Following Interaction	Speed (m/s)	0.33	0.51	0.37	0.58
	Longitudinal distance (m)	0.64	1.10	0.80	1.54
	Lateral distance (m)	0.86	1.00	1.08	1.30
Overtaking Interaction	Speed (m/s)	0.66	1.11	0.83	1.54
	Longitudinal distance (m)	1.65	2.78	2.00	3.54
	Lateral distance (m)	1.47	1.74	1.86	2.10

Table 4.4 Prediction errors of the FM and ME IRL models for the following and overtaking interactions (Alsaleh & Sayed, 2020)

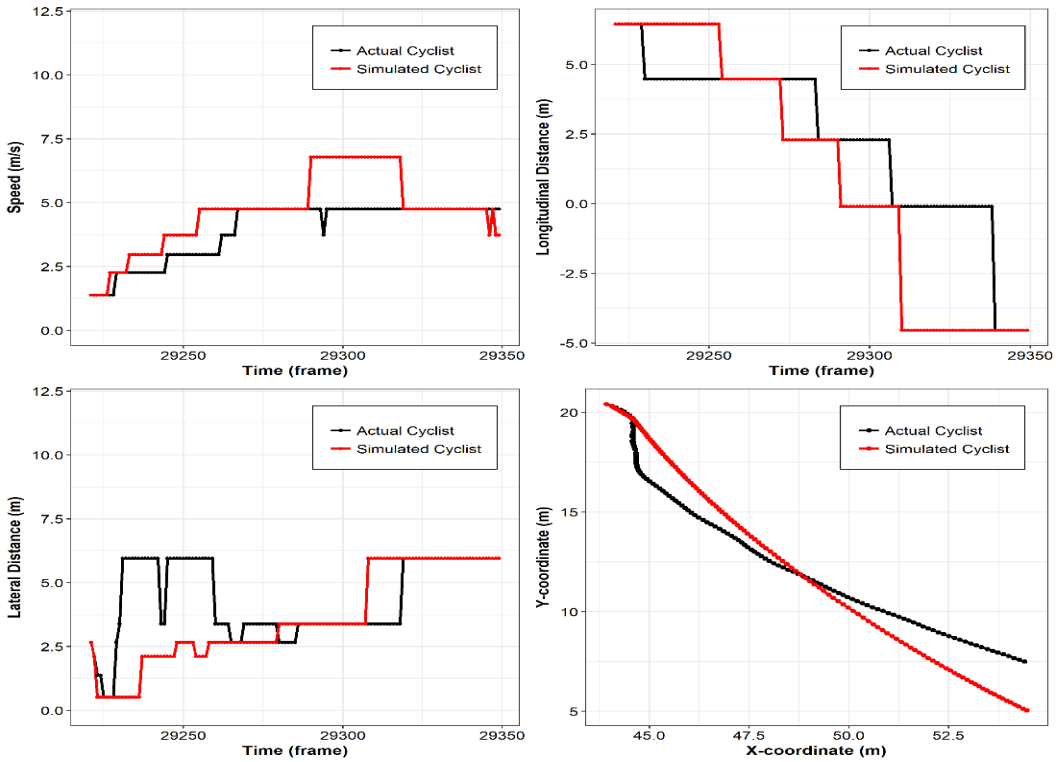


(a) Maximum Entropy (ME)

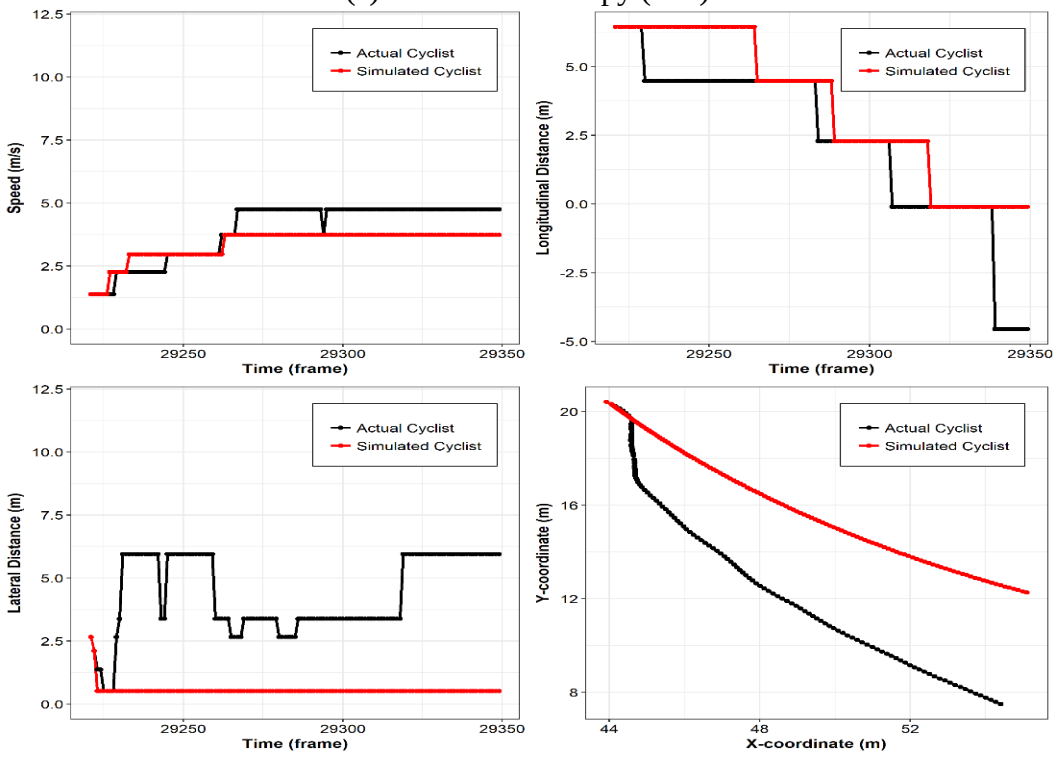


(b) Feature Matching (FM)

Figure 4.11 Following interaction discrete simulation using the ME and FM IRL algorithms (Alsaleh & Sayed, 2020)



(a) Maximum Entropy (ME)



(b) Feature Matching (FM)

Figure 4.12 Overtaking interaction discrete simulation using ME and FM IRL algorithms (Alsaleh & Sayed, 2020)

4.7 Summary and Conclusion

This chapter presented the details of the cyclist-pedestrian interaction modeling in shared space facilities. Two cyclist-pedestrian interactions were considered in the modeling, the following and overtaking interactions. The road users were modeled as utility-based intelligent and rational agents using the MDP modeling framework. Such a modeling approach accounts for road users' intelligence and their ability to logically assess the surrounding environment and take optimal actions that maximize their utilities in order to achieve their goals. This is considered an important step in modeling road users' intelligence in microsimulation platforms, as most of the previous modeling frameworks ignored the intelligence of road users. Considering the intelligence of road users in microsimulation models is important, especially in shared space modeling, as they can have different degrees of freedom in accomplishing specific tasks (e.g., overtaking). The utility (reward) function is the key component that represents how road users logically assess their surrounding environment.

The recovered reward functions from applying the IRL algorithms are important for estimating the road users' policies and developing an agent-based microsimulation model. Such a simulation tool can benefit urban designers and traffic engineering in visualizing the road user trajectories and evaluating the safety and efficiency of shared space facilities. In this study, two IRL algorithms, the Maximum Entropy (ME) and the Feature Matching (FM), were proposed to recover road users' reward functions and estimate their optimal policies. Generally, the results showed that the Maximum Entropy (ME) algorithm outperformed the Feature Matching (FM) algorithm in developing MDP models for cyclist-pedestrian interactions in shared spaces. The ME reward function estimates were more consistent across levels and in line with expectations than the FM reward function estimates.

Moreover, the estimated ME reward function yielded a higher prediction accuracy of road user trajectories than the FM algorithm. The ME IRL algorithm solves the ambiguity issue in reward function estimation and accounts for imperfect (non-optimal) observed behaviour (Ziebart, et al., 2008).

A microsimulation platform was developed to simulate cyclist trajectories in cyclist-pedestrian interactions in shared space facilities. The results were compared to actual cyclist trajectories, relying on their microscopic behaviour parameters. The difference between the actual and simulated trajectories was evaluated using the mean absolute error and the Hausdorff distance. Generally, the ME algorithm outperformed the FM algorithm, and both algorithms predict the cyclist speed more accurately than the cyclist position. The abrupt change in the cyclist behavior (e.g., speed, distance, and yaw rate) under the discrete MDP modeling framework shown in Figure 4.11 and Figure 4.12 can affect the safety assessment of these interactions as it may lead to more severe traffic conflicts.

Overall, the results presented in this chapter confirm the validity of using the MDP framework for modeling road user interactions in shared space facilities. This paves the way for developing more advanced models that consider the continuous nature of road user movements. The continuous modeling approach can be useful for developing a simulation tool for several applications (e.g., traffic safety assessment) as it can improve the trajectory prediction accuracy and avoid abrupt changes in the predicted road users' behaviour. These models can rely on the continuous state MDP modeling framework, which models road user states and actions in continuous spaces. Moreover, the implementation of advanced IRL techniques that consider the expected nonlinearity in the data,

including estimating the road user reward function and policies using the Gaussian Process (GP) and deep neural networks, can be considered. In such a model, the nonlinearity in road user behaviour, preferences, and decisions can be captured. The MDP framework can also be extended to model other types of interaction in shared spaces, including the crossing and head-on interactions. The development of the continuous MDP for cyclists crossing and head-on interactions with pedestrians in shared spaces facilities, including the GP and deep neural network modeling of their reward functions and policies, will be discussed in the fifth chapter of this thesis.

Chapter 5: Multi-directional Models for Cyclist-Pedestrian Interactions: A Continuous Nonlinear Inverse Reinforcement Learning Approach

5.1 Background

The Markov Decision Process (MDP) modeling framework was proposed, in chapter four of this thesis, as an alternative approach for developing agent-based models. This framework can be used to model intelligent agents that can learn from their interaction experience (e.g., demonstrations). The study presented in the previous chapter of this thesis modeled cyclist-pedestrian interactions in shared spaces using the discrete MDP modeling framework (Alsaleh & Sayed, 2020). The study implemented IRL using the maximum entropy and feature matching algorithms to recover reward functions of cyclists involved in following and overtaking interactions with pedestrians. The recovered reward functions were used to infer cyclists' preferences and estimate their optimal policies during their interactions in shared spaces. However, the implemented approaches have several major shortcomings. First, the road user movements (i.e., state variables) and decisions (i.e., action variables) are modeled discretely. However, the road user movement behaviour (e.g., speed, distances) and decisions are of continuous natures. The discretization of the state and action spaces leads to low fidelity predictions of road user trajectories. Moreover, it leads to a significant increase in the problem's dimensionality and the computational cost, referred to as the "curse of dimensionality" issue in IRL (Bellman, 1957; Sutton & Barto, 1998). Second, road user reward functions were assumed linear. However, this assumption can be inaccurate as the road users' reward function can be complex and nonlinear.

This chapter presents the details of developing a novel model for cyclist-pedestrian interactions in shared space facilities. The objectives of this study are to infer cyclists' preferences and model their interactions with pedestrians in shared space facilities. Road user trajectories from three shared space locations in Vancouver, Canada, and New York, USA, are extracted by means of computer vision algorithms, presented in chapter three in this thesis. In this study, two types of cyclists' interactions with pedestrians are considered in the analysis: head-on and crossing interactions. Advanced Artificial Intelligent (AI) techniques that consider road user intelligence and rationality are adopted in this study. This work bridges the gap in modeling road user interactions in shared spaces by accounting for their rationality, intelligence, and continuous decision-making through implementing the continuous state and action MDP modeling approach. The main contributions of this study are: (1) The use of continuous IRL approaches to estimate road users' linear and nonlinear reward functions in shared spaces. This study proposes a novel Gaussian Process Inverse Reinforcement Learning (GPIRL) algorithm (Levine and Koltun 2012) to estimate road users' continuous nonlinear reward functions. (2) The use of a continuous Deep Reinforcement Learning (DRL) approach, referred to as synchronous advantage actor-critic (A2C) algorithm (Mnih et al. 2016; Wu et al. 2017), to estimate road users' optimal policies. The proposed algorithms have several advantages. First, the GPIRL algorithm models the reward functions as Gaussian Process (GP), which can handle continuous, nonlinear, and stochastic rewards of complex road user behaviour. Second, the A2C algorithm models road users' optimal policies using deep neural networks, which can handle continuous, complex, and highly nonlinear policies. Third, the employed algorithms are scalable for high-dimensional state and action spaces.

The following sections in this chapter provide a detailed description of the study locations and the data used for models development and validation, the continuous MDP methodology applied to model cyclist-pedestrian interactions, the Deep Reinforcement Learning (DRL) approach utilized to estimate optimal cyclist policies, the various performance metrics used to assess the accuracy of the developed models, the development of the continuous DRL shared space simulation tool, and the results of estimating cyclists' reward functions and policies during their interactions with pedestrians in shared space facilities.

5.2 Study Locations and Dataset Description

Three non-motorized shared space facilities located in the United States and Canada are considered in this study, as shown in Figure 5.1. The first shared space is located at the intersection of Park Avenue South and East 29th Street, during the summer street program in New York City, New York. New York City considered Park Avenue's temporal closure from Brooklyn Bridge to Central Park (11km) for motorized vehicles to create a safe area for active road users during the summer street program. The other two shared space locations are at Robson Square in downtown Vancouver, BC, as described in detail in chapter 3 and chapter 4 of this thesis. Video data were obtained from the three locations using video cameras during mild weather conditions. Video data were obtained for the first, second, and third study locations for 2 hours in August 2014, 39 hours in May and August 2019, and 10 hours in August 2016, respectively.

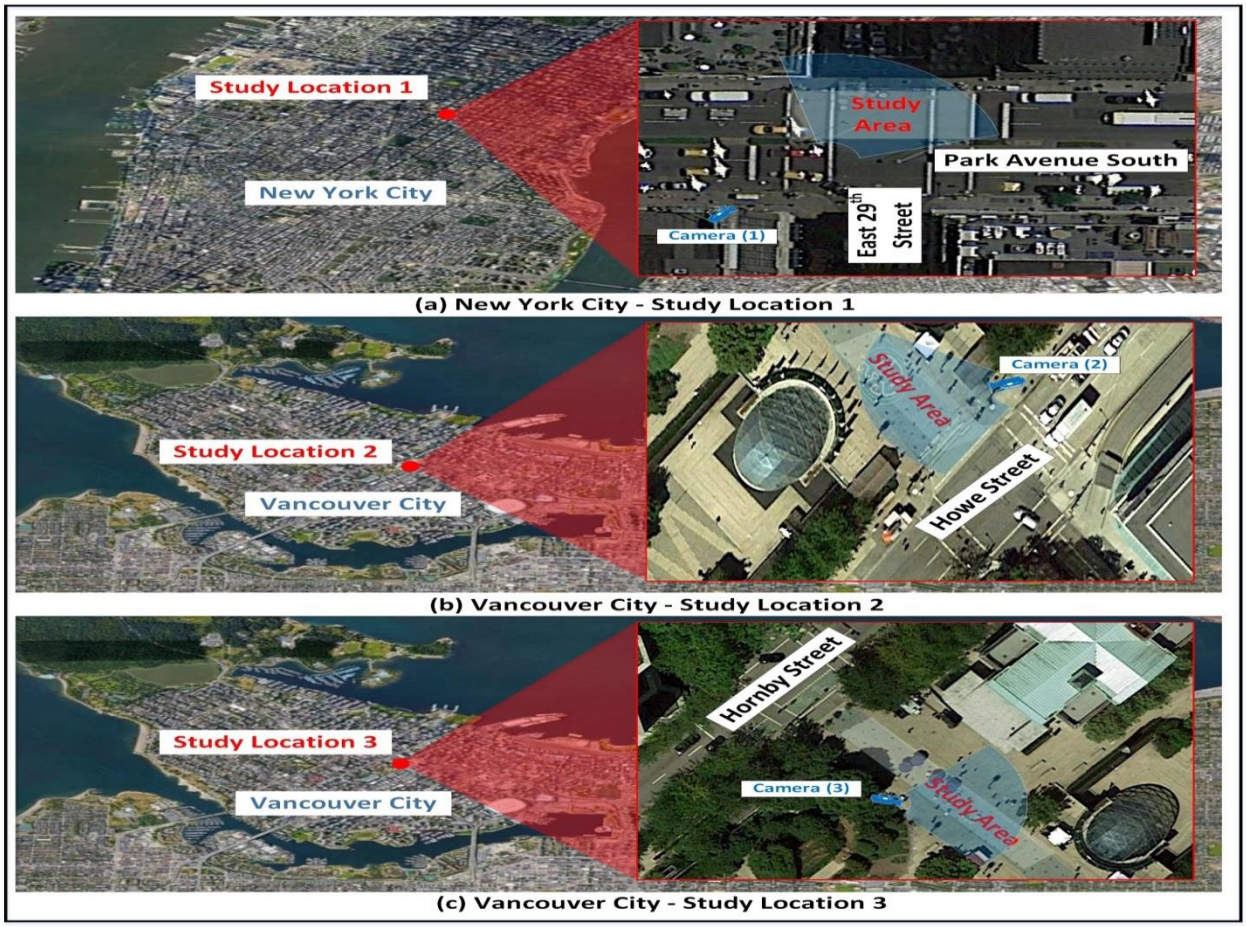


Figure 5.1 Study locations

5.2.1 Data for Training and Testing

Trajectories of cyclists and pedestrians that are involved in head-on and crossing interactions were extracted from the 51 hours of the collected video footage using the computer vision methodology described in chapter 3. Road user trajectories were extracted for the 51 hours of the collected video footage at a frame rate of 30 Hz. The demonstration of road user tracking is shown in Figure 5.2. In total, 458 and 814 cyclist and pedestrian trajectories, which consist of 44,472 and 72,966 data points, were extracted for the head-on and crossing interactions, respectively. The data were split into two datasets, training and testing datasets. The training dataset forms about 80% of the data, which

contains 356 and 636 trajectories that consist of 35,634 and 58,518 data points for the head-on and crossing interactions, respectively. The testing dataset consists of the remaining data.

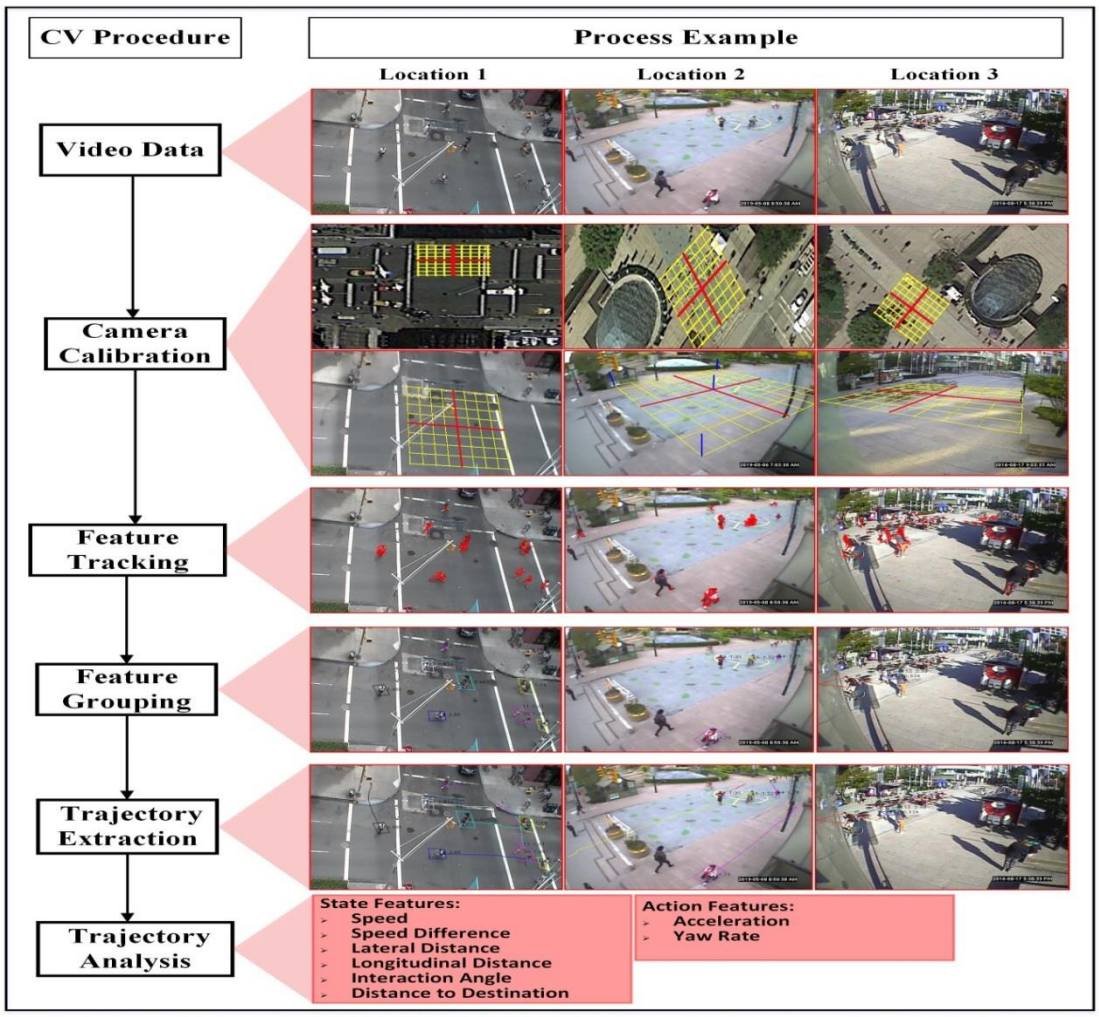


Figure 5.2 Demonstration of road user trajectories' extraction process

5.3 Behavioural Characteristics Extraction

The main set of variables used to describe cyclist and pedestrian interactions in shared spaces was based on previous studies, as discussed earlier in chapter 4 of this thesis. The variables that are used to describe each cyclist's state include its speed ($Speed_{Cyclist}$), speed difference ($Speed_{Difference}$), longitudinal distance ($d_{Longitudinal}$), lateral distance ($d_{Lateral}$), interaction angle ($\theta_{Inteaction\ Angle}$), and distance to the destination ($d_{Destination}$). The cyclist's yaw rate ($r_{Cyclist\ Yaw\ Rate}$) and acceleration ($Acceleration_{Cyclist}$) profiles are used to describe its action. More details about the methodology for calculating road user behaviour variables are provided in Chapters 3 and 4 of this thesis.

5.4 Modeling Methodology

5.4.1 Conceptual Framework

A Road user decision (e.g., cyclist) can be classified into strategic, tactical, and operational levels (Michon, 1985). The strategic level is a high-level decision that deals with the cyclist route planning decision. The tactical level is a mid-level decision that specifies the cyclist's short-term decision, such as yielding or turning decision. The last level, the operational level, is a detailed-level decision that describes the cyclist's microscopic control decision of accelerating and steering. This study focuses on modeling cyclists' operational-level decisions during their interactions with pedestrians in shared space facilities. The study assumes that cyclists observe their surrounding environment (i.e., pedestrians) and perform operational decisions to execute tactical decisions.

In this study, cyclists' operational decisions are modeled using the continuous MDP modeling framework. In this framework, cyclists' states and actions are modeled in continuous spaces. This framework presents a more realistic and accurate modeling technique of cyclists' movement behaviour than the discrete MDP framework (Alsaleh & Sayed, 2020). In this framework, cyclists are assumed to maximize underlying RF, which forms succinct representations of their preferences, intentions, and ways of reacting in different situations. Thus, cyclists' reward function (RF) captures and could infer their operational decisions under different circumstances. This study implemented two IRL algorithms (Levine & Koltun, 2012) to estimate cyclists' RF as linear and GP nonlinear formulations of their state features. The linear modeling approach assumes linear cyclists' preferences and behaviour. However, the GP modeling approach is more flexible and can handle nonlinear cyclists' preferences and behaviour. The GP function is not restricted to a specific functional form (i.e., flexible in modeling any functional form). In the GP, the reward function is assumed to follow a GP (i.e., infinite multivariate Gaussian distributions). The estimated cyclists' RF is used to compute their optimal policies using Deep Reinforcement Learning (DRL). In this approach, the policy function, which presents a map between cyclists' states to their operational-level decisions, is formulated as a deep neural network. Thus, it can handle complex and highly nonlinear cyclists' operational and evasive action decisions of acceleration and yaw rate.

5.4.2 Continuous Inverse Reinforcement Learning Approach

The MDP is described by a tuple of $(S, A, P_{ss'}^{a}, \mathcal{R}, \gamma, D)$ as discussed earlier in chapter 4 of this thesis. In the IRL, the aim is to recover the unknown reward function (RF) from expert demonstrations. The recovered RF can generalize expert demonstrations and infer the expert's goals and preferences. In

the context of cyclist behaviour, cyclist trajectories represent the expert demonstrations. Cyclists behave to optimize their unknown RF. Thus, the problem can be formulated to recover the RF parameters that make their demonstrations appear nearly optimal. A road user trajectory ζ is defined as a tuple of continuous states and actions according to Equation (5.1). The dynamic function *F* (environment dynamics) that specifies the agent's new state given the previous state after taking action is given by Equation (5.2).

$$\boldsymbol{\zeta} = \{s_0, a_0, \dots, s_i, a_i, \dots, s_T\} \tag{5.1}$$

$$\boldsymbol{F}(s_{t-1}, a_t) = s_t \tag{5.2}$$

where s_i is a state, and a_i is an observed action at time $i \in (0, T)$. The following subsections describe the IRL algorithms' implementation.

5.4.2.1 Learning Linear Reward Function

The continuous IRL approach is used to model road user behaviour in cyclist-pedestrian interactions in shared space. The IRL algorithm implemented in this study (Levine & Koltun, 2012) employed the maximum entropy principle to account for the inherent noise and uncertainty in road user demonstrations (Ziebart, et al., 2008), which can lead to computing a more robust RF. Under this model, the expert follows a stochastic policy, and the probability of selecting actions $\boldsymbol{a}$ given by $\boldsymbol{P}(\boldsymbol{a}|s_0)$ is proportional to the exponential reward along the trajectory (e.g., starting from the initial state s_0 and follow the actions $\boldsymbol{a}$) as given by Equation (5.3).

$$\boldsymbol{P}(\boldsymbol{a}|s_0) = \frac{1}{Z}\exp\left(\int_t \boldsymbol{r}(s_t, a_t)\right) \tag{5.3}$$

where Z is the partition function (i.e., trajectories normalization), $\boldsymbol{r}(s_t, a_t)$ is the instantaneous reward for the state-action pairs s_t and a_t. Let $\boldsymbol{r}(\boldsymbol{a})$ donates the sum of the discounted rewards starting from the initial state s_0 and then following the actions $\boldsymbol{a}$ (i.e., along the path (s_0, $\boldsymbol{a}$)), as given by Equation (5.4). The optimal actions, starting at the initial state s_0, are defined as given in Equation (5.5).

$$\boldsymbol{r}(\boldsymbol{a}) = \sum_{\boldsymbol{t}} \gamma^t \boldsymbol{r}(\boldsymbol{s_t}, \boldsymbol{a_t}) \tag{5.4}$$

$$\boldsymbol{a} = arg\ max_a \sum_{\boldsymbol{t}} \gamma^t \boldsymbol{r}(s_t, a_t) \tag{5.5}$$

Computing the partition function is intractable in high-dimensional spaces. Instead, the Laplace approximation, which locally models the distribution as Gaussian (Tierney & Kadane, 1986), can be used to evaluate Equation (5.3) without needing to compute the partition function Z, as given by Equation (5.6). The sum of the discounted reward along the path $\boldsymbol{r}(\boldsymbol{a})$ is approximated using the second-order Taylor expansion of $\boldsymbol{r}$ around $\boldsymbol{a}$ as given in Equation (5.7) (Levine & Koltun, 2012).

$$\boldsymbol{P}(\boldsymbol{a}|\boldsymbol{s_0}) = e^{\boldsymbol{r}(\boldsymbol{a})} \left[\int e^{\boldsymbol{r}(\tilde{\boldsymbol{a}})} d\tilde{\boldsymbol{a}} \right]^{-1} \tag{5.6}$$

$$\boldsymbol{r}(\tilde{\boldsymbol{a}}) \approx \boldsymbol{r}(\boldsymbol{a}) + (\tilde{\boldsymbol{a}} - \boldsymbol{a})^T \frac{\boldsymbol{\partial r}}{\boldsymbol{\partial a}} + \frac{\mathbf{1}}{\mathbf{2}} (\tilde{\boldsymbol{a}} - \boldsymbol{a})^T \frac{\boldsymbol{\partial^2 r}}{\boldsymbol{\partial a^2}} (\tilde{\boldsymbol{a}} - \boldsymbol{a}) \tag{5.7}$$

where $\boldsymbol{r}(\tilde{\boldsymbol{a}})$ is a second Taylor expansion of $\boldsymbol{r}$ around $\boldsymbol{a}$, $\tilde{\boldsymbol{a}}$ donates all possible actions, including the optimal actions, $\frac{\mathbf{\partial r}}{\mathbf{\partial} \boldsymbol{a}}$ is the RF gradient (**g**), $\frac{\mathbf{\partial^2 r}}{\mathbf{\partial} \boldsymbol{a^2}}$ is the RF Hessian (**H**). Thus, the approximate estimate of Equation (5.6) can be represented as given by Equation (5.8) (Levine & Koltun, 2012). The RF parameters can be learned by maximizing the log-likelihood of the expert demonstrations under the

parameterized RF. The IRL log-likelihood of the RF is given by Equation (5.9) (Levine & Koltun, 2012).

$$\boldsymbol{P}(\boldsymbol{a}|\boldsymbol{s_0}) \approx e^{\frac{1}{2}\mathbf{g}^T\mathbf{H}^{-1}\mathbf{g}} |-\mathbf{H}|^{\frac{1}{2}} (2\pi)^{-\frac{d_a}{2}} \tag{5.8}$$

$$\boldsymbol{L} = \frac{1}{2}\mathbf{g}^T\mathbf{H}^{-1}\mathbf{g} + \frac{1}{2}\log|-\mathbf{H}| - \frac{\boldsymbol{d_a}}{2}\log 2\pi \tag{5.9}$$

where $\boldsymbol{L}$ is the IRL log-likelihood of the expert demonstrations under the parameterized reward function $\boldsymbol{r}$, and d_{a} is the number of action features. The linear RF takes the form presented in Equation (5.10), and it is parameterized by the feature weights $\boldsymbol{\theta}^T$.

$$\boldsymbol{r}(\boldsymbol{s_t}, \boldsymbol{a_t}) = \boldsymbol{\theta}^T f(\boldsymbol{s_t}) \tag{5.10}$$

where $f(\boldsymbol{s_t})$ is the features that depend on the agent states. The Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Liu & Nocedal, 1989) is used to maximize the log-likelihood function.

5.4.2.2 Learning the Nonlinear Gaussian Process (GP) Reward Function

The RF's linearity assumption can be relaxed by modeling it nonlinearly as a Gaussian Process (GP) using the Gaussian Process Inver Reinforcement Learning (GPIRL) algorithm (Levine, et al., 2011; Levine & Koltun, 2012). The GP RF can capture the nonlinearity in expert demonstrations and account for the heterogeneity in road user behavior. In this model, road users follow a stochastic policy, and the probability of observing an action a in the state s is proportional to the exponential of the expected total reward after taking action a Equation (5.11). Thus, the log-likelihood of observing

trajectories *D* given the inferred RF (i.e., the demonstrations' log-likelihood under the RF) is given by Equation (5.12) (Ziebart, et al., 2008; Levine, et al., 2011).

$$P(a|s) = \exp(Q^r_{sa} - V^r_s) \tag{5.11}$$

$$logP(D|r) = \sum_i \sum_t logP(a_{i,t}|s_{i,t}) = \sum_i \sum_t (Q^r_{s_{i,t}a_{i,t}} - V^r_{s_{i,t}}) \tag{5.12}$$

where $V^r_s = \log \sum_a \exp(Q^r_{sa})$ is the value function, and it is presented as a soft version of the Bellman equation, Q^r_{sa} is the expected value of performing action a in the state s and following the policy. Like the RF's linear modeling approach, Laplace approximation is used to approximate the partition function's computation and the IRL log-likelihood.

The RF is formulated as a nonlinear combination of a set of state features as a Gaussian Process. The inputs for the Gaussian Process are a set of inducing state feature points $\mathbf{f} = (f^1 \dots f^n)^T$, and the noiseless outputs ($\boldsymbol{y}$) at these points are learned. The GP RF's structure is determined by the Kernel function, which presents the Gaussian Process's covariance. The Kernel function is a variant of the regularized radial basis function (RBF) given by Equation (5.13), and its hyperparameters $\boldsymbol{\theta}_{GP} = \{\beta, \lambda\}$ are leaned using the GP Bayesian framework (Levine & Koltun, 2012).

$$k(f^i, f^j) = \beta \exp\left(-\frac{1}{2}\sum_k \lambda_k [(f^i_k - f^j_k)^2 + 1_{i \neq j}\sigma^2]\right) \tag{5.13}$$

where f^i and f^j are the positions of two induced points in the feature space, β is a hyperparameter that presents the overall variance, λ is a hyperparameter that determines the weight on each feature. The GP log-likelihood of the noiseless outputs ($\boldsymbol{y}$) and the kernel hyperparameters $\boldsymbol{\theta}_{GP} = \{\beta, \lambda\}$

given the inducing feature points $\mathbf{f}$ is given by Equation (5.14) (Williams & Rasmussen, 2006; Levine & Koltun, 2012).

$$logP(\boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f}) = -\frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{K}^{-\mathbf{1}}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}| + logP(\boldsymbol{\Theta}_{GP}|\mathbf{f}) \tag{5.14}$$

where $\boldsymbol{K}$ is the GP covariance given by $\boldsymbol{K_{ij}} = k\left(f^i, f^j\right)$, and $\mathcal{P}(\boldsymbol{\Theta}_{GP}|\mathbf{f})$ is the hyperparameters prior.

The hyperparameters' prior encourages the feature weights λ to be sparse and prevent degeneracies (Levine & Koltun, 2012). The log of the hyperparameters prior $logP(\boldsymbol{\Theta}_{GP}|\mathbf{f})$ is given by Equation (5.15). The values of the hyperparameters $\boldsymbol{\Theta}_{GP}$ and the noiseless outputs ($\boldsymbol{y}$) can be found by maximizing their probability under the expert demonstrations D as given by Equation (5.16) (Levine, et al., 2011).

$$logP(\boldsymbol{\Theta}_{GP}|\mathbf{f}) = -\frac{1}{2}\sum_{ij}[\boldsymbol{K}^{-\mathbf{1}}]_{ij}^2 - \sum_i \log(\lambda_{\boldsymbol{ii}} + \mathbf{1}) \tag{5.15}$$

$$\mathrm{P}(D, \boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f}) = \left[\int_r \underbrace{P(D|\boldsymbol{r})}_{IRL\ term} \underbrace{P(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\Theta}_{GP}, \mathbf{f})}_{GP\ posterior} d\boldsymbol{r}\right] \underbrace{P(\boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f})}_{GP\ probability} \tag{5.16}$$

where $P(\mathrm{D}|\boldsymbol{r})$ is the IRL term, $P(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\Theta}_{GP}, \mathbf{f})$ is the GP posterior, which is the probability of the RF under the current values of $\boldsymbol{\Theta}_{GP}$ and $\boldsymbol{y}$, and $P(\boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f})$ is the GP prior probability of a particular assignment of $\boldsymbol{\Theta}_{GP}$ and $\boldsymbol{y}$. The integral in Equation (5.16) is intractable and cannot be computed in a closed form. Instead, it can be approximated under the mean value of the GP posterior ($\boldsymbol{r_m}$) as given by Equation (5.17). Thus, the GP RF can be obtained after optimizing the log-likelihood given in Equation (5.17) as the GP posterior's mean given by Equation (5.18) (Levine, et al., 2011).

$$logP(D, \boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f}) = logP(\mathrm{D}|\boldsymbol{r_m}) + logP(\boldsymbol{y}, \boldsymbol{\Theta}_{GP}|\mathbf{f}) \quad (5.17)$$

$$\boldsymbol{r_m}(\boldsymbol{s_t}, \boldsymbol{a_t}) = \boldsymbol{k_t}\boldsymbol{\varphi} \quad (5.18)$$

where $logP(\mathrm{D}|\boldsymbol{r_m})$ is the IRL log-likelihood, $\boldsymbol{r_m}$ is the mean of the GP RF after optimizing the log-likelihood, $\boldsymbol{\varphi} = \boldsymbol{K^{-1}y}$, $\boldsymbol{k_t}$ is a vector corresponding to the covariance between $f(s_t, a_t)$ and each inducing point f^i given by $\boldsymbol{k_{ti}} = k\big(f(s_t, a_t), f^i\big)$. The log-likelihood function is optimized using the L-BFGS algorithm (Liu & Nocedal, 1989).

5.5 Road User Policy Estimation: An Actor-Critic Deep Reinforcement Learning

Road users' policies (e.g., sequences of decisions and actions) are learned after obtaining the RFs using the Deep Reinforcement Learning (DRL) through the synchronous temporal difference (TD) advantage actor-critic (A2C) algorithm (Mnih, et al., 2016; Wu, et al., 2017). This approach takes advantage of both offline and online policy estimation. In this approach, two deep neural networks are used. The first deep neural network is used as a function approximator for the states' value function V^{π}. The second deep neural network, parameterized by a vector ϑ, is used as a function approximator for road users' policy π. The policy neural network forms as an actor taking various actions in the environment, while the value function neural network forms as a critic, which judges the taken actions and therefore encourages or discourages them. The training objective for the policy approximation neural network is to maximize the expected discounted cumulative return from the initial state, which is given by $J(\vartheta) = E_{\pi}[R_{\pi}] = E_{\pi}\big[\sum_{i\geq 0}^{\infty} \gamma^i r(s_{t+i}, a_{t+i})\big]$. The gradient of the policy objective with respect to the policy parameter vector ϑ is given by Equation (5.19) (Mnih, et al., 2016; Wu, et al., 2017).

$$\nabla_{\vartheta} J(\vartheta) = E_{\pi}[\sum_{t=0}^{\infty} \nabla_{\vartheta} log \pi_{\vartheta}(a_t|s_t).A^{\pi}(s_t, a_t)] \tag{5.19}$$

where $A^{\pi}(s_t, a_t)$ is the advantage function, which provides a relative measure of actions' values at each state. The advantage function, which is estimated as the temporal difference (TD) error $TD = Q(s_t, a_t) - V(s_t)$, is given by Equation (5.20) (Mnih, et al., 2016; Wu, et al., 2017).

$$A^{\pi}(s_t, a_t) = \sum_{i=0}^{k-1} \gamma^i r(s_{t+i}, a_{t+i}) + \gamma^k V_{\emptyset}^{\pi}(s_{t+k}) - V_{\emptyset}^{\pi}(s_t) \tag{5.20}$$

where $V_{\emptyset}^{\pi}(s_t)$ is the critic neural network estimator for the RL value function (i.e., deep neural network estimator of the expected discounted cumulative sum of rewards from a given state following the policy π), which is given by $V_{\emptyset}^{\pi}(s_t) = E_{\pi}[R_{\pi}]$. The critic neural network is optimized to minimize the temporal difference error (TD error). The neural networks are optimized using the stochastic gradient descent algorithm. A detailed illustration of the method is provided in Figure 5.3.

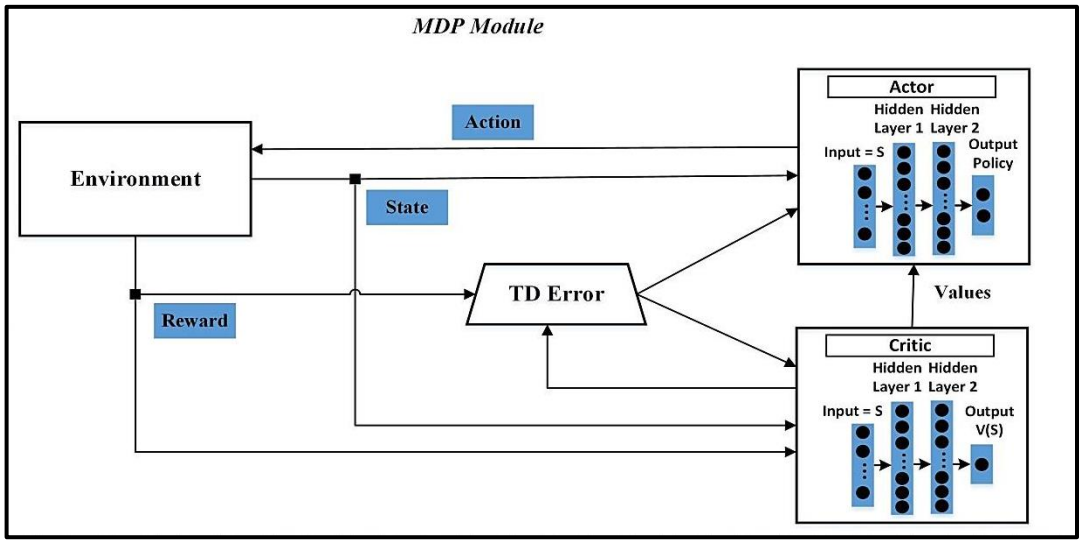


Figure 5.3 Synchronous TD Advantage Actor-Critic (A2C) algorithm flowchart

5.6 Performance Metrics

The accuracy of the simulated trajectories is assessed based on the relative distance from the actual trajectories. In this study, two evaluation metrics are used to assess the accuracy of the simulated trajectories, including the *Mean Absolute Error (MAE)* and the *Hausdorff Distance (HausD)*. The methodology for calculating these metrics is presented in detail in chapter 4 of this thesis.

5.7 Results and Discussion

In this study, cyclists' linear and GP nonlinear continuous reward functions are estimated for each type of interaction using the training dataset. The recovered reward functions are used to estimate cyclists' optimal policies and simulate their validation dataset trajectories. The following subsections present the estimated cyclists' reward functions and the trajectory validation results.

5.7.1 Road User Behavioural Characteristics

The analysis and molding of road user behaviour are carried out separately for the head-on and crossing interactions. Most of the interactions occurred at low shared space density. Road user trajectories form the basis for calculating cyclist and pedestrian variables/ features that describe their behaviour at each time step. Analysis of cyclist and pedestrian state and action profiles reveals multiple characteristics of their behaviour in the head-on and crossing interactions, as illustrated in Figure 5.4. Negative values of $Speed_{Difference}$, $r_{Cyclist\ Yaw\ Rate}$, or $d_{Longitudinal}$ indicate higher pedestrian speeds, counterclockwise angles, and pedestrians located behind the cyclists, respectively. A descriptive statistic of their behavioural profiles is presented in Table 5.1. These behaviour parameter values present road users' observed (actual) behaviour from the collected shared space

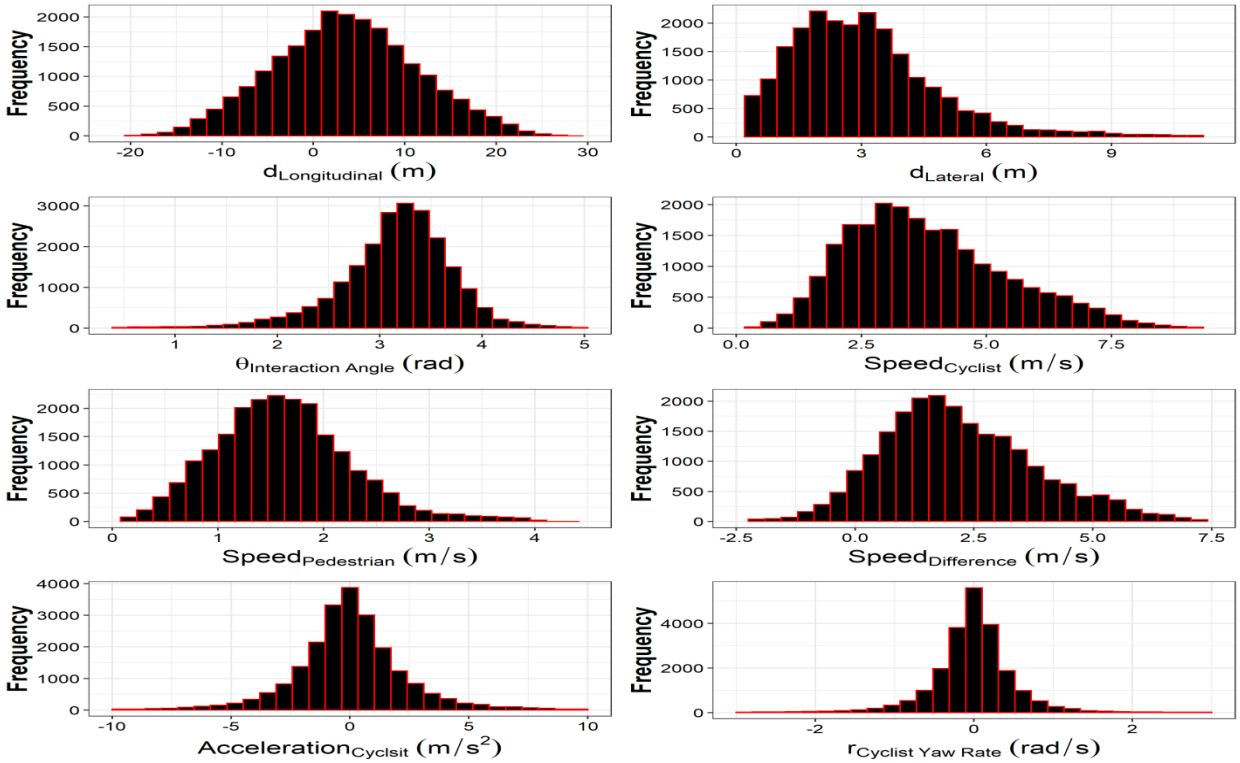
video footages. The analysis shows that the average $Speed_{Cyclist}$ and $Speed_{Difference}$ are higher for the head-on interaction compared to the crossing interaction. The head-on interaction is associated with a slightly larger average $d_{Longitudinal}$ and $d_{Lateral}$ compared to the crossing interaction. The average interaction angle ($\theta_{Inteaction\ Angle}$) for the head-on and crossing interactions are 3.131 rad (179.4^o) and 1.619 rad (92.7^o), respectively. For the cyclist actions, cyclists involved in head-on interaction are associated with higher average $Acceleration_{Cyclist}$ and lower magnitude of average $r_{Cyclist\ Yaw\ Rate}$ in comparison with those involved in the crossing interaction. The obtained cyclist and pedestrian behaviour parameter distributions are expected and agree with previous studies (Figliozzi, et al., 2013; Montufar, et al., 2007). In this study, the use of the computer vision system for tracking road users, instead of the manual approach, enabled the detailed and accurate extraction of their behavioural profiles at each time frame.

Variable	Head-on Interaction mean [SD*]	Crossing Interaction mean [SD]
$Speed_{Cyclist}\ [m/s]$	3.735 [1.590]	3.034 [1.566]
$Speed_{\text{Pedestrian}}\ [m/s]$	1.558 [0.663]	1.184 [0.576]
$Speed_{Difference}\ [m/s]$	2.177 [1.712]	1.851 [1.532]
$\theta_{Inteaction\ Angle}{}^{*}\ [rad]$	3.131 [0.649]	1.619 [0.535]
$d_{Lateral}\ [m]$	3.197 [2.343]	2.983 [2.956]
$d_{Longitudinal}\ [m]$	3.641[7.982]	3.404 [6.839]
$Acceleration_{Cyclist}\ [m/s^2]$	0.160 [1.915]	0.123 [1.798]
$r_{Cyclist\ Yaw\ Rate}\ [rad/s]$	-0.019 [0.790]	-0.073 [0.853]

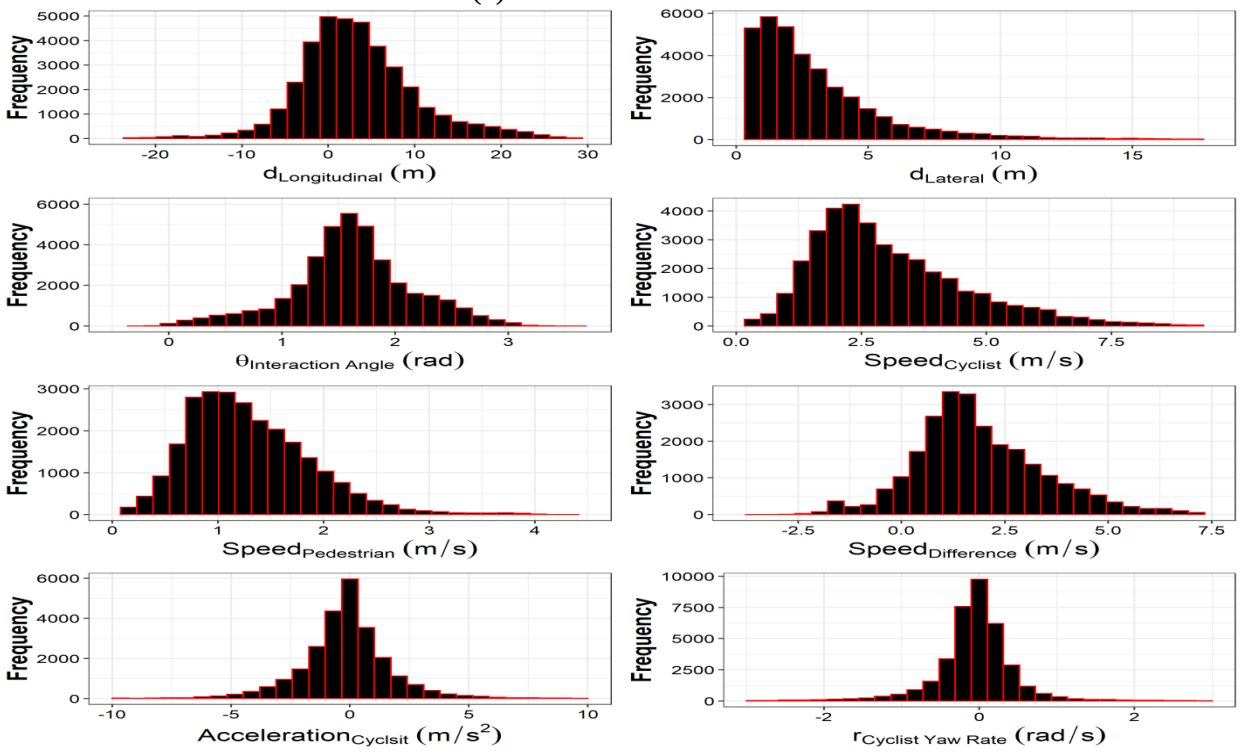
* Standard deviation

* Absolute angle difference for the crossing interaction.

Table 5.1 Data descriptive statistics for the head-on and crossing interactions (Alsaleh & Sayed, 2021a)



(a) Head-on Interaction



(b) Crossing Interaction

Figure 5.4 Road users behavioural characteristics for the head-on and crossing interactions (Alsaleh & Sayed, 2021a)

5.7.2 Linear and Gaussian Process Reward Functions Recovery

5.7.2.1 Cyclist-Pedestrian Head-on Interaction

The estimated continuous RFs for the head-on interaction using the linear and GP formulations are presented in Figure 5.5 and Figure 5.6, respectively. The figures present the RFs over bivariate state feature space while holding the other state features constant at their mean values. The RFs infer cyclists' preferences during their interactions with pedestrians in shared spaces. A summary of the inferred cyclist preferences is presented in Table 5.2.

The linear RF presented in Figure 5.5(a)-Figure 5.5(b) suggests that cyclists have insignificant preferences (comparing with other state features) for keeping larger $d_{Lateral}$ and lower $d_{Longitudinal}$ values (i.e., large negative $d_{Longitudinal}$) with respect to pedestrians. Negative values of $d_{Longitudinal}$ indicate that cyclists have crossed pedestrians. The GP RF presented in Figure 5.6(a)-Figure 5.6(b) suggests that cyclists prefer to keep intermediate $d_{Lateral}$ (e.g., Gaussian distribution (GD) with a mean of about 3.5 m) and lower $d_{Longitudinal}$ values. These inferred cyclists' preferences using the GP RF are similar to their overtaking interaction preferences (Alsaleh, et al., 2020). Moreover, the GP RF suggests nonlinear influences on cyclists lateral distances where cyclists prefer to increase their $d_{Lateral}$ at higher $Speed_{Cyclist}$ or $Speed_{Difference}$ as presented in Figure 5.6(f)-Figure 5.6(h). For example, at a low $Speed_{Cyclist}$ of 2 m/s or $Speed_{Difference}$ of 1 m/s, cyclists prefer to keep a small $d_{Lateral}$ (e.g., GD with a mean of about 2.5 m). However, at a higher $Speed_{Cyclist}$ of 4.5 m/s or $Speed_{Difference}$ of 2.5 m/s, cyclists prefer to keep a larger $d_{Lateral}$ (e.g., GD with a mean of about 3.5 m). The inferred cyclist preferences using the GP RF are expected and more consistent with previous studies

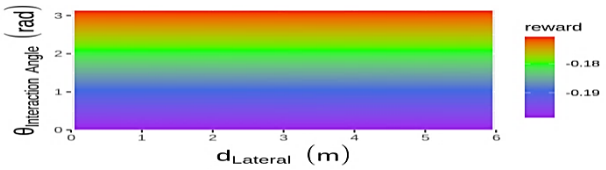
(Alsaleh, et al., 2020; Luo, et al., 2015) than the linear RF. Previous studies show that the head-on interaction typically involves swerving maneuvers and increasing of $d_{Lateral}$ to avoid collisions between road users. Moreover, previous studies show that road users prefer to keep larger lateral distances at higher speeds during their interactions in mixed traffic conditions (Luo, et al., 2015). Capturing the nonlinearity in the lateral distance preferences by the GP model is likely to provide more accurate lateral distance predictions than the linear RF model. The cyclists' preferences for keeping low $d_{Longitudinal}$ values indicate their higher preferences to resolve the conflicts with pedestrians.

Furthermore, the linear RF presented in Figure 5.5(d)-Figure 5.5(e) suggests higher cyclists' preferences for being in high $Speed_{Cyclist}$ and low $Speed_{Difference}$ states. However, the GP RF presented in Figure 5.6(d)-Figure 5.6(e) suggests that cyclists prefer to have intermediate $Speed_{Cyclist}$ and $Speed_{Difference}$ (e.g., GDs with means of about 4.5 m/s and 3m/s, respectively). The inferred cyclists' preferences using the GP RF are expected and more consistent with the behavioral study than the linear RF (Alsaleh, et al., 2020). Moreover, these cyclists' preferences are similar to their overtaking interaction preferences (Alsaleh & Sayed, 2020). For the $\theta_{Interaction\ Angle}$, both the linear and GP RFs assign higher reward values (i.e., higher preference) for the states with larger $\theta_{Interaction\ Angle}$ than being aligned with pedestrians in the same direction as presented in Figure 5.5(g) and Figure 5.6(g), respectively. This result is expected, as the cyclists in the head-on interaction are associated with large angle difference with respect to pedestrians (e.g., angle difference of about 3.14 rad). Moreover, the GP RF captures important nonlinearity in $\theta_{Interaction\ Angle}$ preference as it suggests higher ranges of angle

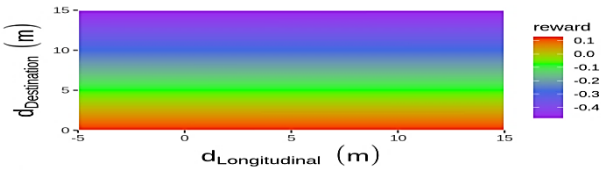
difference preferences at higher cycling speeds as presented in Figure 5.6(d). This case is associated with cyclists swerving maneuvers/ evasive actions in the head-on interaction to avoid collisions with pedestrians. For the $d_{Destination}$, both the linear and GP RFs presented in Figure 5.5(c) and Figure 5.6(c) suggest that cyclists prefer to move toward their destinations, which agrees with the previous studies (Anvari, et al., 2015; Dias, et al., 2018). These results are consistent with the shared space behavioral study (Alsaleh, et al., 2020).

State Feature	Cyclists' Preferences in Interactions with Pedestrians	
	Linear Reward Function	GP Reward Function
$d_{Lateral}$	Insignificant preference for large lateral distance	Preference for intermediate lateral distance
$d_{Longitudinal}$	Insignificant preference for low longitudinal distance value (i.e., large negative longitudinal distance)	Preference for low longitudinal distance value (i.e., large negative longitudinal distance)
$Speed_{Cyclist}$	Preference for high cycling speed	Preference for intermediate cycling speed
$Speed_{Differenc}$	Preference for low speed-difference with respect to pedestrians	Preference for intermediate speed difference with respect to pedestrians
$\theta_{Interaction\ Angl}$	Preference for large interaction angle with pedestrians	Preference for large interaction angle with pedestrians
$d_{Destination}$	Preference for small distance to destination	Preference for small distance to destination

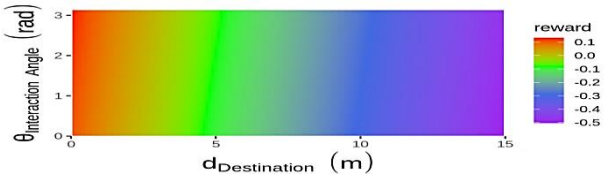
Table 5.2 Reward function inferences of cyclist preferences in the head-on interaction (Alsaleh & Sayed, 2021a)



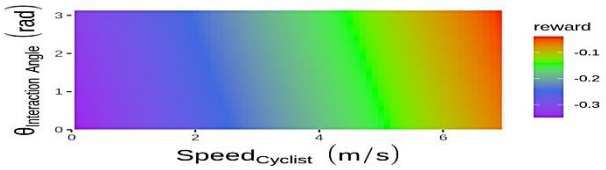
(a) Lateral Distance Vs. Interaction Angle



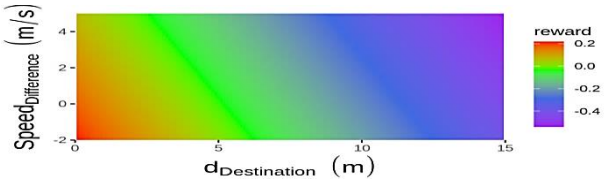
(b) Longitudinal Distance Vs. Distance to Destination



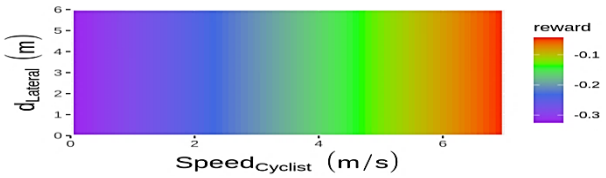
(c) Distance to Destination Vs. Interaction Angle



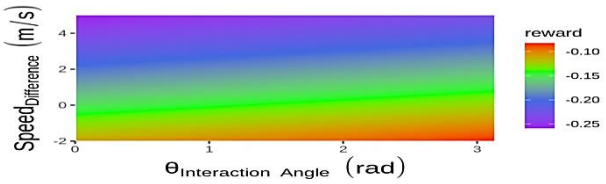
(d) Speed Vs. Interaction Angle



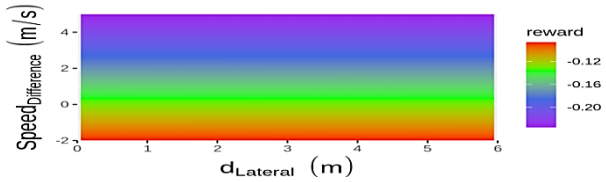
(e) Distance to Destination Vs. Speed Difference



(f) Speed Vs. Lateral Distance

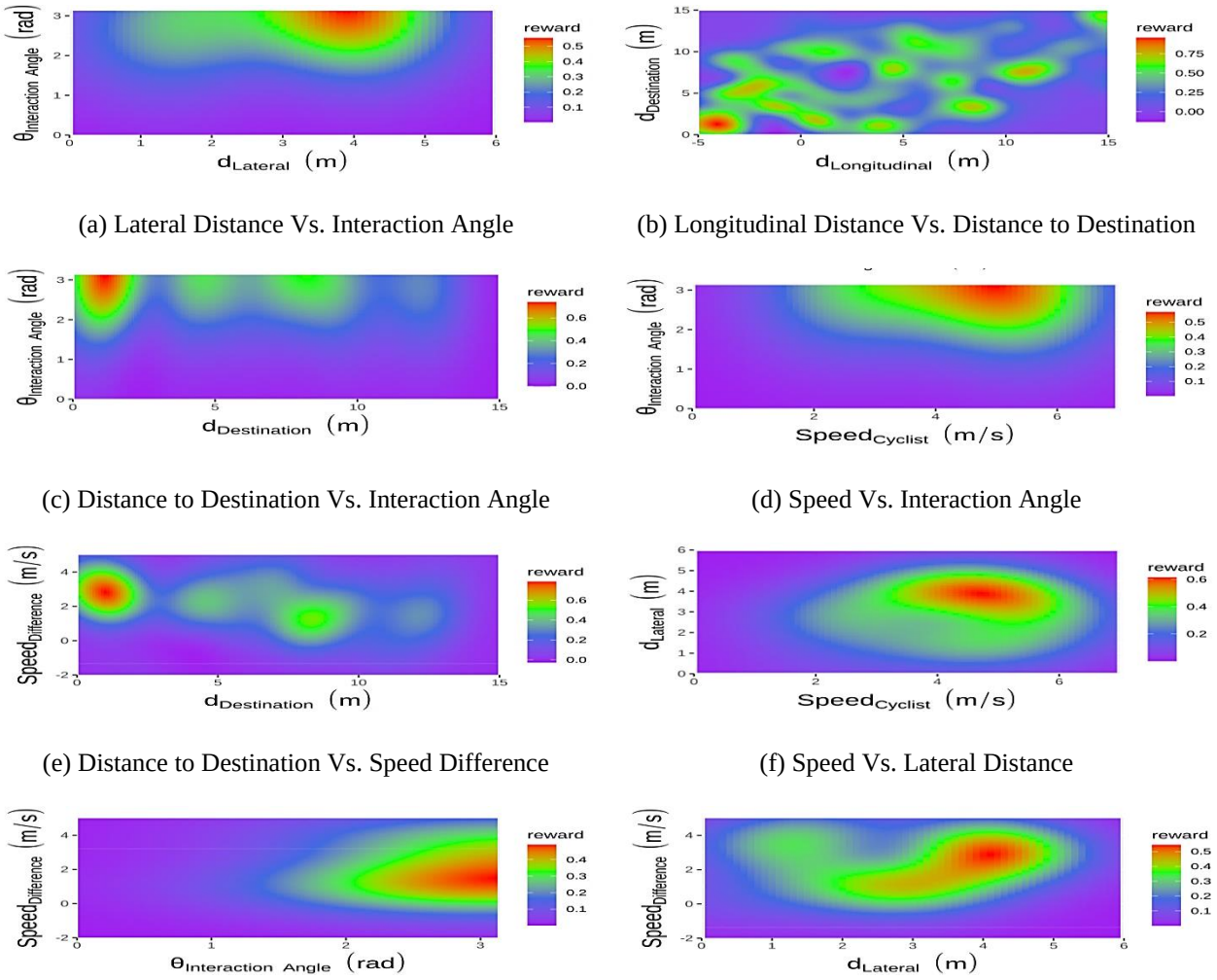


(g) Interaction Angle Vs. Speed Difference



(h) Lateral Distance Vs. Speed Difference

Figure 5.5 Linear reward function visualization (at mean feature values) for the head-on interaction (Alsaleh & Sayed, 2021a)



(a) Lateral Distance Vs. Interaction Angle

(b) Longitudinal Distance Vs. Distance to Destination

(c) Distance to Destination Vs. Interaction Angle

(d) Speed Vs. Interaction Angle

(e) Distance to Destination Vs. Speed Difference

(f) Speed Vs. Lateral Distance

(g) Interaction Angle Vs. Speed Difference

(h) Lateral Distance Vs. Speed Difference

Figure 5.6 Gaussian Process (GP) reward function visualization (at mean feature values) for the head-on interaction (Alsaleh & Sayed, 2021a)

5.7.2.2 Cyclist-Pedestrian Crossing Interaction

The estimated linear and GP continuous RFs for the crossing interaction are presented in Figure 5.7 and Figure 5.8, respectively. The estimated RFs are presented over bivariate state feature space while holding the other state features constant at their mean values. A summary of the RFs inferences of cyclist preferences is presented in Table 5.3.

The linear RF, presented in Figure 5.7(a)-Figure 5.7(b), suggests that cyclists have insignificant preferences for keeping larger $d_{Lateral}$ and lower $d_{Longitudinal}$ values with respect to pedestrians. The GP RF accounts for the expected nonlinearity in cyclists' behaviour as it suggests that cyclists prefer to keep intermediate $d_{Lateral}$ (e.g., GD with a mean of about 3.5 m) as presented in Figure 5.8(a). Similar to the head-on interaction, the GP RF presented in Figure 5.8(f)-Figure 5.8(h) suggests nonlinear influences of $Speed_{Cyclist}$, and $Speed_{Difference}$ on cyclists' preference for $d_{Lateral}$. The GP RF suggests that cyclists prefer to keep a larger $d_{Lateral}$ at a higher $Speed_{Cyclist}$ or $Speed_{Difference}$. For example, a cyclist lateral distance of about 2.5 m is associated with a low $Speed_{Cyclist}$ of 2 m/s or $Speed_{Difference}$ of 1 m/s. However, at a higher $Speed_{Cyclist}$ of 3.5 m/s or $Speed_{Difference}$ of 1.5 m/s, cyclists prefer to keep a larger $d_{Lateral}$ (e.g., GD with a mean of about 3.5 m). Moreover, the GP RF captures expected nonlinearity in the $d_{Longitudinal}$ feature as it accounts for the different preferences in the crossing interaction. The GP RF presented in Figure 5.8(b) suggests that cyclists have two preferences for the $d_{Longitudinal}$. These preferences are to yield for pedestrians and keep adequate $d_{Longitudinal}$ (e.g., small positive $d_{Longitudinal}$), which is the highest preference, or to cross first (e.g., negative $d_{Longitudinal}$). Such different cyclists' preferences are captured by the

different reward values assigned at each particular cyclist state. Similar to the head-on interaction, the inferred cyclist preferences using the GP RF are expected and more consistent with the previous studies (Alsaleh, et al., 2020; Alsaleh & Sayed, 2020; Luo, et al., 2015) than the linear RF.

Furthermore, the linear RF presented in Figure 5.7(e)-Figure 5.7(f) suggests higher cyclists' preferences for being in high $Speed_{Cyclist}$ and low $Speed_{Difference}$ states. However, the GP RF presented in Figure 5.8(e)-Figure 5.8(f) suggests that cyclists prefer to have intermediate $Speed_{Cyclist}$ and $Speed_{Difference}$ (e.g., GDs with means of about 3.5 m/s and 1.5 m/s, respectively). The inferred cyclist preferences using the GP RF are expected and more consistent with the previous studies than the linear RF (Alsaleh, et al., 2020; Alsaleh & Sayed, 2020). For the $\theta_{Interaction\ Angle}$, the linear RF presented in Figure 5.7(g) suggests higher cyclists' preference for being in the states with low $\theta_{Interaction\ Angle}$ than approaching pedestrians against their movement direction. However, the GP RF suggests that cyclists prefer to keep intermediate $\theta_{Interaction\ Angle}$ with pedestrians, e.g., a GD with a mean of about 1.6 rad (about 92^{o}) as presented in Figure 5.8(g). This result is expected and more consistent with the previous studies (Alsaleh, et al., 2020) than the linear RF, as the crossing interaction is associated with intermediate $\theta_{Interaction\ Angle}$ distribution. Moreover, the GP RF presented in Figure 5.8(d) suggests that at high $Speed_{Cyclist}$, cyclists prefer to keep their movement direction (i.e., the highest preference for $\theta_{Interaction\ Angle}$ is about 1.6 rad). This result indicates that cyclists do not prefer to swerve or yield to pedestrians but prefer to cross first (i.e., pedestrians take evasive actions and yield for cyclists). However, at low $Speed_{Cyclist}$, cyclists may prefer to yield for

pedestrians or swerve around them, as the GP RF suggests a broader range of preferences for the crossing angle difference. These results are consistent with the behavioral study (Alsaleh, et al., 2020), as it suggests that cyclist speed is an influential factor in specifying road users' evasive actions in cyclist and pedestrian interactions. The study found that pedestrians prefer to yield to cyclists at high cyclists' speeds. For the $d_{Destination}$ feature, both the linear and GP RFs presented in Figure 5.7(c) and Figure 5.8(c) suggest that cyclists prefer to move toward their destinations. However, the GP RF accounts for the different cyclists' preferences in the crossing interaction (i.e., yielding to pedestrians or cross first).

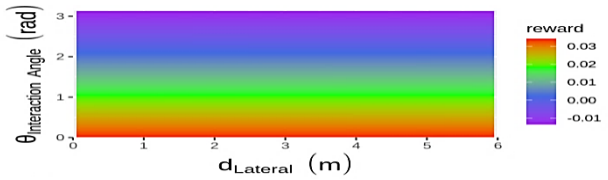
State Feature	Cyclists' Preferences while Interacting with pedestrians	
	Linear Reward Function	GP Reward Function
$d_{Lateral}$	Insignificant preference for large lateral distance	Preference for intermediate lateral distance
$d_{Longitudinal}$	Insignificant preference for low longitudinal distance value	Highest preference for small positive longitudinal distance (yield for pedestrian)
$Speed_{Cyclist}$	Preference for high cycling speed	Preference for intermediate cycling speed
$Speed_{Differenc}$	Preference for low speed-difference with respect to pedestrians	Preference for intermediate speed difference with respect to pedestrians
$\theta_{Interaction\ Angl}$	Preference for low interaction angle with pedestrians	Preference for intermediate interaction angle with pedestrians
$d_{Destination}$	Preference for small distance to destination	Preference for small distance to destination

Table 5.3 Reward function inferences of cyclist preferences in the crossing interaction (Alsaleh & Sayed, 2021a)

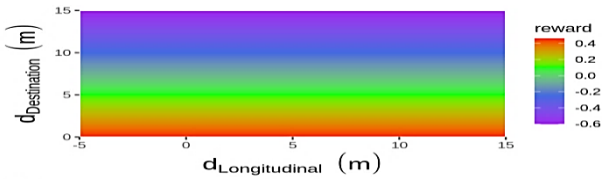
Overall, the GP RFs' estimates, for both the head-on and crossing interactions, are more consistent with the shared space and mixed traffic studies (Alsaleh, et al., 2020; Alsaleh & Sayed, 2020; Luo, et al., 2015) and in agreement with expectations than the linear RFs. The GP

modeling approach, which models the reward function as infinite multivariate Gaussian distributions, captures complex, nonlinear and heterogeneous road users' behavioural interactions and preferences. However, the estimated linear RFs did not adequately learn road users' behaviour in the head-on and crossing interactions, which may limit the generality and transferability of the linear RFs. Capturing the nonlinearity and accounting for the heterogeneity of road users' behaviour in the GP RFs can significantly improve the model's prediction accuracy compared with the linear model.

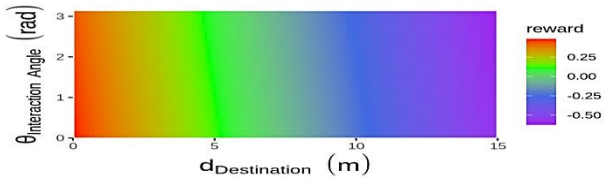
Furthermore, unlike the traditional modeling approaches for road user interactions (e.g., SFM and CA), the current approach recovers road users' RFs, which form the basis for modeling their rationality and intelligence. Moreover, this approach explicitly infers road user preferences in different situations and rationalizes their actions, which are challenging to be accurately inferred and rationalized using these traditional modeling approaches. Accurate inferences of road user preferences in shared spaces help in better designing such facilities and their regulatory laws (e.g., speed limits). Moreover, the estimated RFs could infer road user evasive actions and operational-level decisions in different situations. The reward given for each particular cyclist state signals its appropriate transition over state space, which therefore implicitly infers its operational-level decisions of acceleration and yaw rate at various situations. For example, Figure 5.8(b) shows that the highest cyclists' preference is to yield to pedestrians and maintain their longitudinal distance values (small positive longitudinal distances). This indicates that cyclists would likely decelerate and maintain their movement direction (avoid swerving) to resolve the conflict situation.



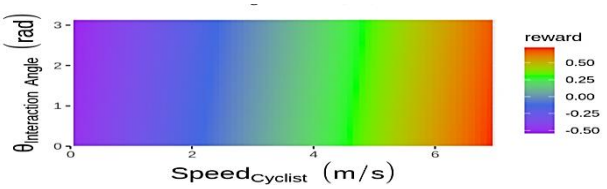
(a) Lateral Distance Vs. Interaction Angle



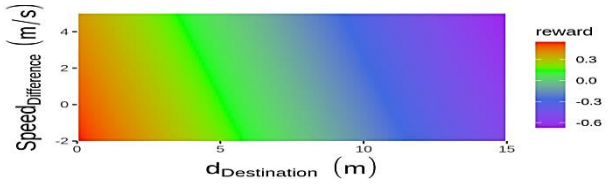
(b) Longitudinal Distance Vs. Distance to Destination



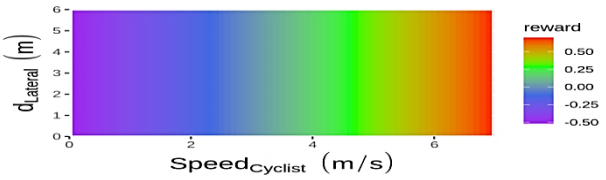
(c) Distance to Destination Vs. Interaction Angle



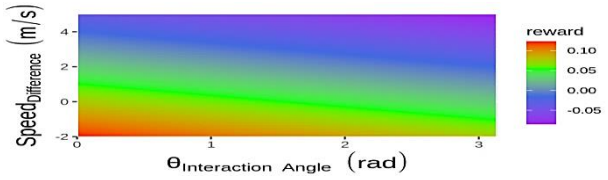
(d) Speed Vs. Interaction Angle



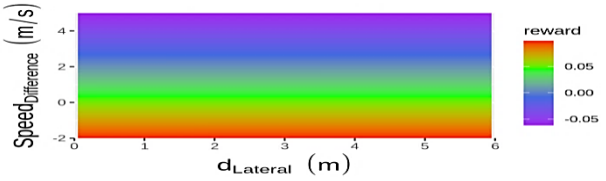
(e) Distance to Destination Vs. Speed Difference



(f) Speed Vs. Lateral Distance

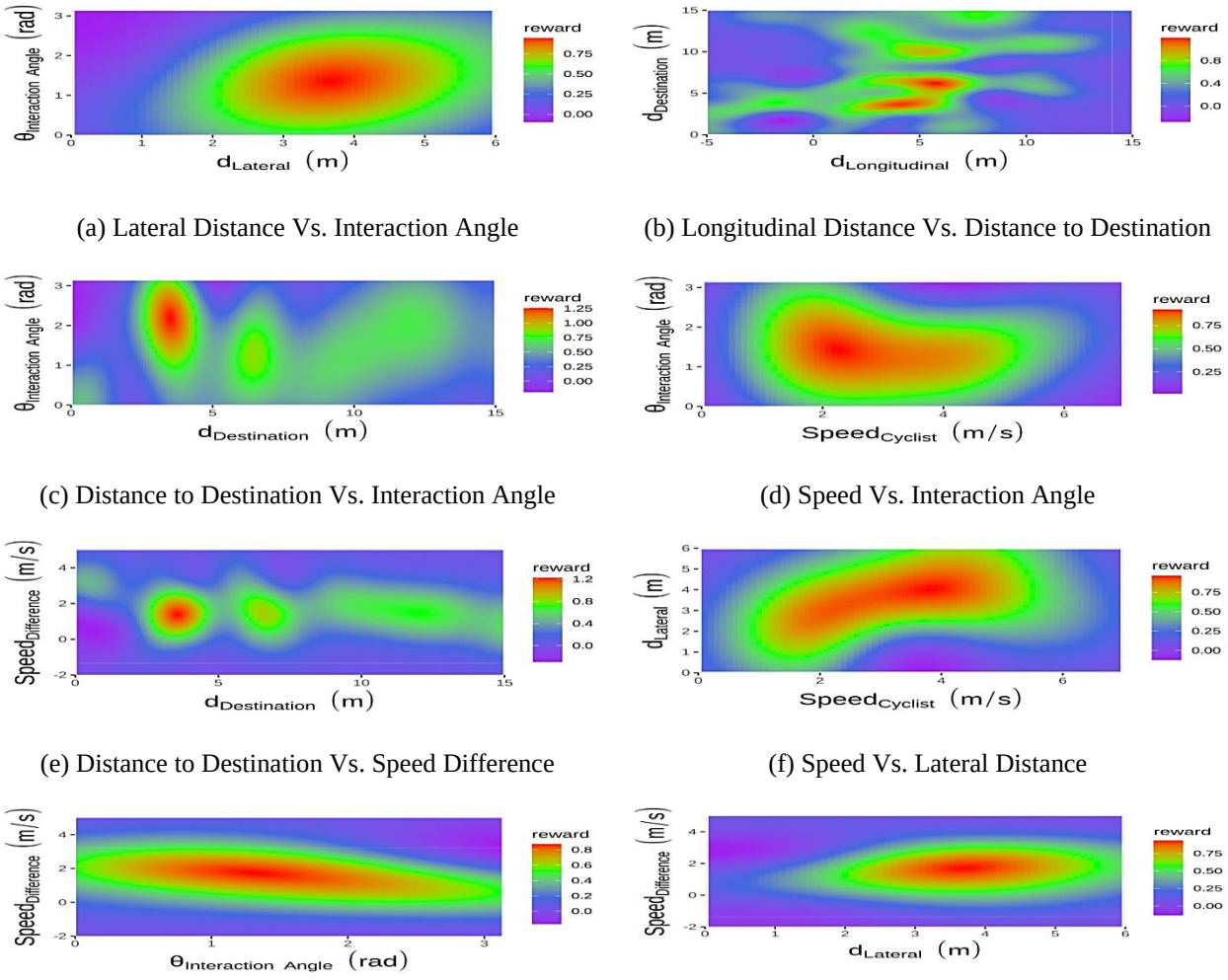


(g) Interaction Angle Vs. Speed Difference



(h) Lateral Distance Vs. Speed Difference

Figure 5.7 Linear reward function visualization (at mean feature values) for the crossing interaction (Alsaleh & Sayed, 2021a)



(a) Lateral Distance Vs. Interaction Angle

(b) Longitudinal Distance Vs. Distance to Destination

(c) Distance to Destination Vs. Interaction Angle

(d) Speed Vs. Interaction Angle

(e) Distance to Destination Vs. Speed Difference

(f) Speed Vs. Lateral Distance

(g) Interaction Angle Vs. Speed Difference

(h) Lateral Distance Vs. Speed Difference

Figure 5.8 Gaussian Process (GP) reward function visualization (at mean feature values) for the crossing interaction (Alsaleh & Sayed, 2021a)

5.7.3 Road User Policy Estimations and Trajectory Predictions

In this study, a customized shared space simulation environment is developed using OpenAI Gym (Brockman, et al., 2016). The simulation environment emulates cyclists' interactions with pedestrians in shared spaces. The workflow for the simulation tool is illustrated in Figure 5.9. First, the customized environment is reset to initialize the initial state distribution of the road users. Pedestrian flow for the validation dataset is then initialized with information about their behaviour over time. The customized environment contains simulation physics that controls the cyclist's movement after taking action (e.g., acceleration and yaw rate). Thus, it determines the cyclist's next state/observation after taking a specific action. Moreover, it determines the agent gained reward after taking action based on the recovered linear and GP RFs. The deep reinforcement learning through the synchronous TD advantage actor-critic algorithm is used to estimate cyclist policies based on the linear and GP nonlinear RF models. Cyclists perform suitable actions at each state by sampling for the computed policies. The simulation tool was run at a rate of 30 HZ.

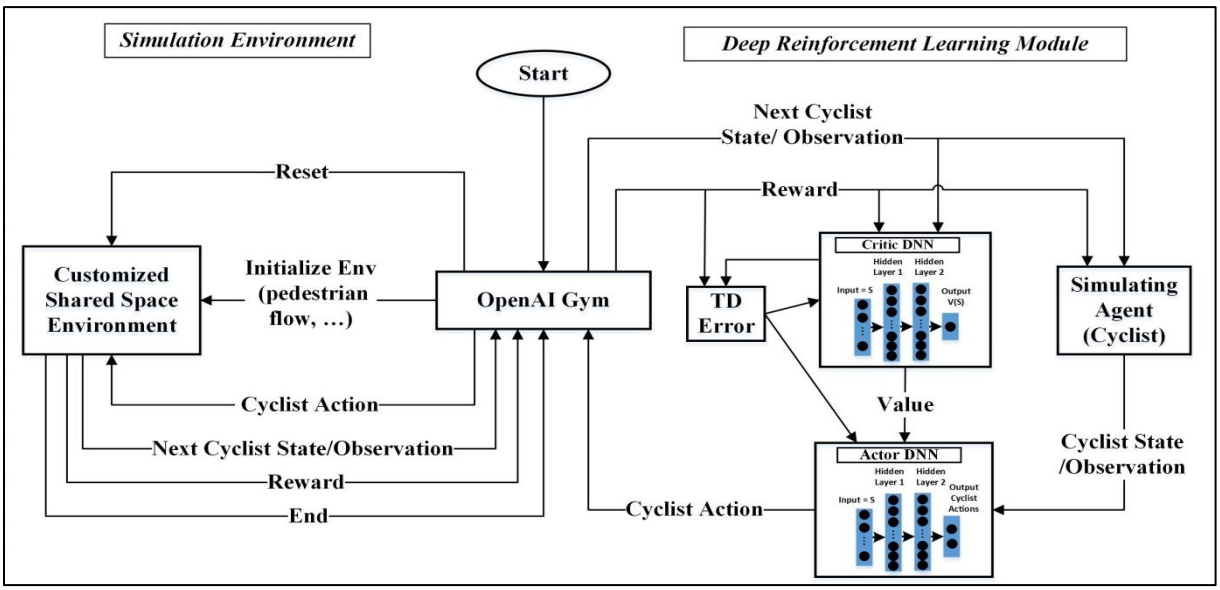


Figure 5.9 Shared space DRL simulation tool (Alsaleh & Sayed, 2021a)

The developed shared space simulation tool was used to simulate cyclists' interactions in the validation dataset. The predicted trajectories' accuracy was assessed using the MAE and the *HausD* error metrics, as presented in Table 5.4. Overall, the GP nonlinear RF predicted cyclists' trajectories more accurately than the linear RF for both the head-on and crossing interactions. The GP nonlinear RF model led to significantly more accurate predictions of $Speed_{cyclist}$, and relative distances (e.g., $d_{Longitudinal}$, and $d_{Lateral}$) compared to the linear RF model. The prediction of $Speed_{cyclist}$ was more accurate compared to the cyclist position for both the linear and GP RFs. For the head-on interaction, the GP nonlinear RF improves the prediction accuracy of $Speed_{cyclist}$ by about 54.0% (50.7% in terms of Hausdorff distance) compared to the linear RF. As well, the GP RF improves the predictions of cyclist positions with respect to pedestrians (e.g., $d_{Longitudinal}$, and $d_{Lateral}$). The GP nonlinear RF improves the prediction accuracy of both $d_{Longitudinal}$, and $d_{Lateral}$ by about 47.8% and 50.0% (57.4% and 42.4% in terms of Hausdorff distance), respectively.

For the crossing interaction, the GP nonlinear RF improves the prediction accuracy of $Speed_{cyclist}$ by about 54.8% (48.4% in terms of Hausdorff distance) comparing with the linear RF. For the $d_{Longitudinal}$ and $d_{Lateral}$, the GP RF achieves average improvement in the prediction accuracy by about 52.5% and 48.2% (57.5% and 42.0% in terms of Hausdorff distance), respectively. Figure 5.10 and Figure 5.11 show examples of actual and simulated trajectories in terms of their $Speed_{cyclist}$, $d_{Longitudinal}$ and $d_{Lateral}$ from applying both the linear and GP nonlinear RFs for the head-on and crossing interactions, respectively. As shown in the figures, the GP nonlinear RF model reproduces more accurate trajectories for the head-on

and crossing interactions compared to the linear reward model. Moreover, the figures show that both the linear and the GP models predicted the head-on interaction behaviour (i.e., cyclists' speeds and relative distances with respect to pedestrians) more accurately than the crossing interaction. This result is expected as the crossing interaction is associated with more complex road user behaviour than the head-on interaction.

Interaction Type	Variable	Nonlinear GP Reward		Linear Reward	
		$Avg.\ MAE$	$Avg.\ HausD$	$Avg.\ MAE$	$Avg.\ HausD$
Head-on Interaction	Speed (m/s)	0.52	0.75	1.13	1.52
	Longitudinal distance (m)	0.83	1.06	1.59	2.49
	Lateral distance (m)	0.66	0.95	1.32	1.65
Crossing Interaction	Speed (m/s)	0.56	0.83	1.24	1.61
	Longitudinal distance (m)	0.87	1.14	1.83	2.68
	Lateral distance (m)	0.71	0.98	1.37	1.69

Table 5.4 Prediction errors of the linear and GP nonlinear reward function models for the head-on and crossing interactions (Alsaleh & Sayed, 2021a)

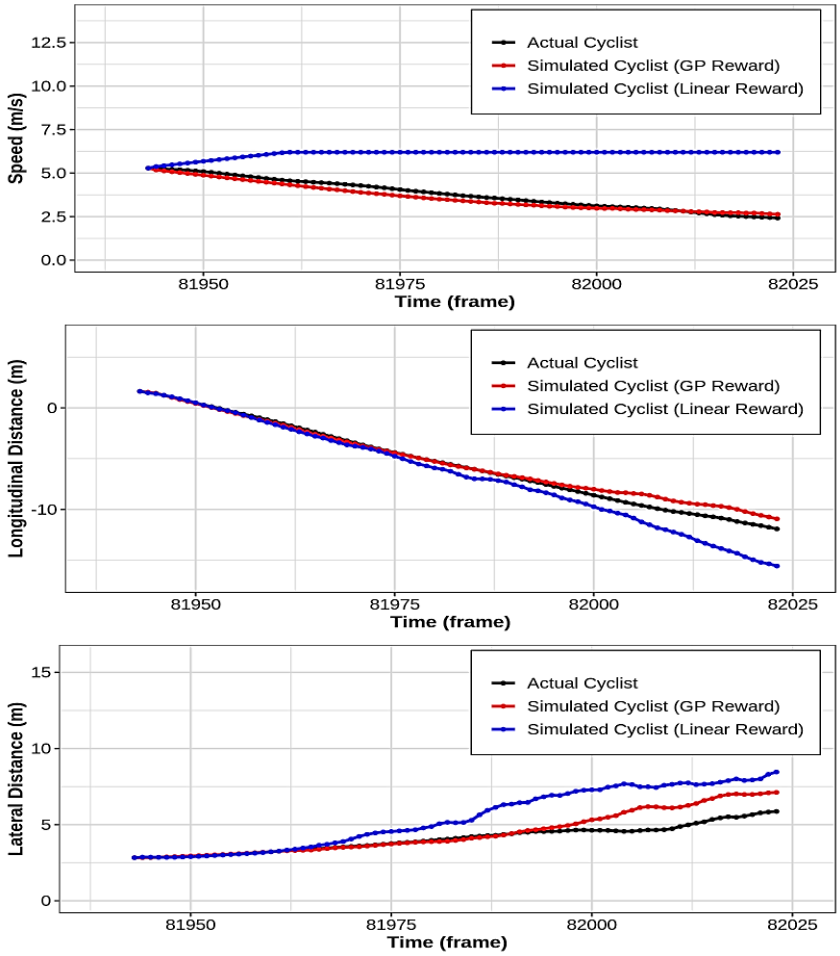


Figure 5.10 Head-on interaction simulation using the linear and GP nonlinear reward functions (Alsaleh & Sayed, 2021a)

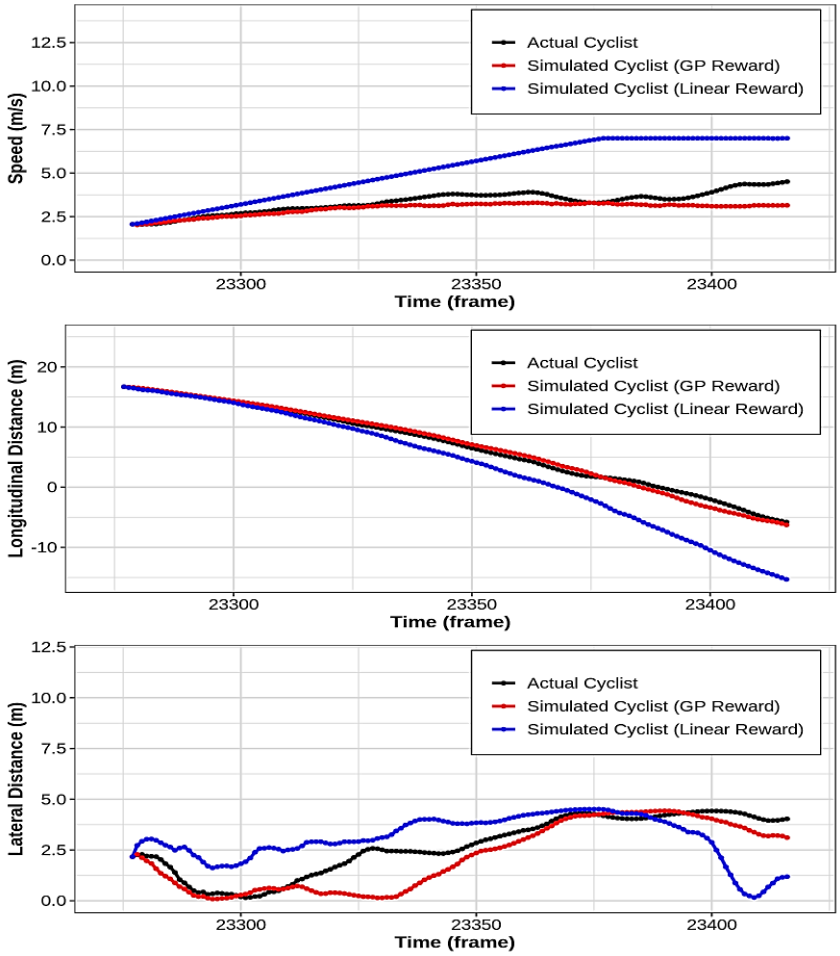


Figure 5.11 Crossing interaction simulation using the linear and GP nonlinear reward functions (Alsaleh & Sayed, 2021a)

5.8 Summary and Conclusion

This chapter presented the details of developing a microsimulation-oriented framework for modeling cyclist movement and interaction with pedestrians in shared space facilities. The modeling framework is based on implementing Artificial Intelligent (AI) techniques to recover road user RFs. Road user RFs form the basis for modeling their behaviour as reward-based rational agents. This study bridges that gap in modeling road user interactions in shared spaces

by accounting for their rationality and intelligence. Accounting for such road user characteristics in simulation frameworks is crucial in shared spaces due to their less restricted movement and interaction patterns. The modeling framework learns the intended goals and preferences of road users and the ways to accomplish such goals. The utility/ reward function is a representation of agents' goals and preferences, forming the key to understanding how road users rationally assess and react in different situations.

This study proposes a continuous IRL approach aiming at recovering continuous road users' utility/ reward functions. Unlike the discretization approach for modeling utility/ reward function, the continuous approach models road users' states and actions in continuous spaces. The continuous modeling approach leads to smoother predicted trajectories and avoids the abrupt changes resulted from states/ utility function discretization. The continuous IRL algorithms are implemented to recover the linear and Gaussian Process (GP) stochastic nonlinear utility/ reward functions for cyclists in head-on and crossing interactions. The GP nonlinear reward function provides a way for learning heterogeneity in road users' behaviour by modeling the distribution of their behaviour. The recovered utility/ reward functions are used to develop an agent-based microsimulation platform to emulate road users' interactions in shared spaces. The study utilized DRL using the synchronous TD advantage Actor-Critic algorithm to estimate road users' optimal policies. The accuracy of the simulated trajectories was validated against actual trajectories in the validation dataset. The results show that the recovered GP nonlinear utility/ reward functions led to more accurate predictions of cyclists' trajectories than the linear reward functions. The

developed simulation tool has a wide range of practical applications, including analyzing shared space facilities' operation and various designs, cyclists' power and energy, and road user safety.

Overall, the results presented in this chapter confirm the validity of using the continuous MDP framework for modeling road user interactions in shared space facilities. This paves the way for developing more advanced models that use different distributional assumptions than Gaussian for modeling road user interactions in shared space facilities. The use of advanced modeling approaches as the deep Neural Network can be beneficial in considering the high nonlinearity in road users' utility/ reward functions, which could improve the trajectory prediction accuracy. Generative adversarial neural networks that can learn the complex and highly nonlinear road users' behaviour can be implemented for modeling road user interactions. Moreover, the implementation of advanced techniques that consider the joint interaction behaviour between road users, as the multi-agent inverse reinforcement learning (MAIRL) modeling framework, can be considered. This approach represents a more realistic way of modeling road user interactions in shared space facilities. The development of the continuous multi-agent models for cyclist and pedestrian interactions in shared spaces facilities, including the generative adversarial neural network modeling of their reward functions, will be discussed in the sixth chapter of this thesis.

Chapter 6: Markov-Game Modeling of Cyclist-Pedestrian Interactions: A Multi-Agent Adversarial Inverse Reinforcement Learning Approach

6.1 Background

Recently, advanced AI techniques have been used to model road users' interaction behaviour. These modeling techniques showed superior performance in emulating road user interactions compared with the classical modeling approaches (e.g., the social force (SF) and cellular automata (CA) models). For example, Zamboni et al. (Zamboni, et al., 2020) compared the performance of an AI model (2D convolutional neural network) with the SFM in emulating active road user interactions. The study found that the AI-based model significantly outperformed the SFM in predicting road user trajectories. Zhang et al. (Zhang, et al., 2020) compared the performance of a deep reinforcement learning (DRL) model with the SFM for emulating complex pedestrian behaviour. The study found that the generated DRL policy outperformed the SFM policy. Mohammed et al. (Mohammed, et al., 2021) compared a single-agent imitation learning model's performance with CA models in emulating cyclists' behaviour on unidirectional off-street paths. The study found that the AI model significantly outperformed the CA models in predicting cyclists' behaviour.

Moreover, Alsaleh and Sayed (Alsaleh & Sayed, 2020; Alsaleh & Sayed, 2021a) modeled road user interactions in shared spaces as reward-based agents using the Markov Decision Process (MDP) modeling framework. This modeling framework considers road users' intelligence and rationality and has been proposed to overcome the shortcoming of the classical modeling

approaches of active road user behaviour. These studies utilized the inverse reinforcement learning (IRL) approach to recover the reward functions of road users using their interaction demonstrations. However, such an approach suffers from several shortcomings. Mainly, it models road users' reward functions discretely and linearly (Alsaleh & Sayed, 2020) or assumes they fall within a particular distributional class (e.g., Gaussian) (Alsaleh & Sayed, 2021a). Moreover, the developed models are significantly affected by the environment dynamics, making them challenging to handle non-stationary environments such as shared space facilities.

Furthermore, despite the multi-agent nature of road user interactions, most developed road users' interaction models (Dias, et al., 2018; Hussein & Sayed, 2017; Anvari, et al., 2015; Luo, et al., 2015; Alsaleh & Sayed, 2020; Alsaleh & Sayed, 2021a) are based on the single-agent modeling framework. In this framework, a single road user's operational level actions are considered for modeling while assuming the other interacting road users' actions are known over time (i.e., part of the passive environment). However, this assumption is unrealistic and could limit the model's accuracy and transferability in non-stationary road user environments. Road user interactions are better represented as multi-agent interactions. In the multi-agent modeling framework, the action sequences of all interacting agents are modeled. Moreover, agents can be cooperative or competitive and can coordinate their actions to achieve specific goals (i.e., avoid collision). The multi-agent framework's main advantages lay in handling non-stationary environments and optimal mixed-strategy policies and capturing the equilibrium information compared to the single-agent framework. For example, (Lin, et al., 2014) compared the performance of single-agent IRL and multi-agent inverse reinforcement learning (MAIRL) in a simulated soccer game.

The study found that the multi-agent IRL recovers agents' rewards that are substantially closer to the ground truth rewards and yields better policies than the single-agent IRL. (Georgila, et al., 2014) compared the performance of the single-agent reinforcement learning (RL) and multi-agent reinforcement learning (MARL) for learning dialogue policies in a resource allocation negotiation scenario. The study found that the RL might fail to converge to the optimal policy under concurrent agents' learning since it is designed for stationary environments. Moreover, the performance of the RL deteriorates comparing to the MARL as the number of agents increases.

The work presented in this chapter aims to develop a novel microscopic multi-agent simulation model for cyclist and pedestrian interactions in shared space facilities. The contribution of this study is in the use of multi-agent modeling of road user interactions using the Markov Game (MG) modeling framework. Markov Game is a generalization of the MDP to the case of multiple interacting agents. In MG, multiple interacting road users learn simultaneously, and their sequences of actions are modeled concurrently. MG has been proposed as a framework for multi-agent reinforcement learning (MARL) (Littman, 1994). Recent advances in MG overcome serious shortcomings in the classical game theory, where it fails to provide any usable solution concepts for some theoretical and empirical classes of games, including (1) games with sequential structure (i.e., two or more successive stages) and (2) games with bounded rationality agents (i.e., limited information access) (Harsanyi & Selten, 1988). In MG, unlike the MDP, each agent's optimal policy is influenced by the other interacting agents' policies. Thus, the optimality notion is defined by an equilibrium solution concept, such as the Nash Equilibrium

(NE) (Hu & Wellman, 1998; Nash, 1951) or the Logistic Stochastic Best Response Equilibrium (LSBRE) (Yu, et al., 2019; McKelvey & Palfrey, 1998; McKelvey & Palfrey, 1995).

The multi-agent modeling framework has been widely used to analyze and model multi-agent systems in several fields, including economics (Yeung, et al., 1999), robotics (Mataric, et al., 2003), and demand management (Lu, et al., 2020). The MARL approach shows promising performance in addressing challenging sequential decision-making problems in several multi-agent systems like multi-player games (Peng, et al., 2017), energy demand response management (Lu, et al., 2020), and negotiation problems (Georgila, et al., 2014). In such a multi-agent approach, designing agents' reward functions become more challenging as agents can have different and conflicting goals and different state-action representations (Song, et al., 2018; Yu, et al., 2019). Moreover, agents can be prone to undesired behaviour due to reward miss-specification (Amodei, et al., 2016). Thus, the imitation learning framework can be used to tackle this issue via expert demonstrations, where agents directly learn desired behaviours by imitating experts.

This study proposes a novel Multi-Agent Adversarial Inverse Reinforcement Learning approach (MA-AIRL) for modeling and simulating road user interactions at shared space facilities. Unlike the traditional game-theoretic framework that models multi-agent systems as a single time-step payoff, the proposed approach is based on Markov Games (MG), which models road users' sequential decisions concurrently. The proposed approach recovers road users' multi-agent reward functions using adversarial deep neural network discriminators and estimates their

optimal policies using Multi-agent Actor-Critic with Kronecker factors (MACK) deep reinforcement learning. The proposed MA-AIRL approach (Yu, et al., 2019) has several advantages. First, it is based on MG, which considers the stochastic nature of shared space environments and models road user sequential decisions (i.e., concurrent road users' actions). Second, the algorithm models road users' reward functions using adversarial deep neural networks, which can handle complex road user behaviour. Third, the algorithm is scalable for high dimensional state and action spaces and unknown dynamics, and recovers robust reward functions that are less affected by the environment dynamics. Forth, the algorithm is capable of recovering road users' multi-agent reward functions with no prior assumption about the agents' interaction type considered for the modeling (e.g., cooperative or competitive). Fifth, the algorithm can handle bounded rationality agents (e.g., limited information access), which are challenging to be handled in the traditional game-theoretic framework. Lastly, to the best of the authors' knowledge, this is the first work to model active road user interactions in shared spaces using the MA-AIRL modeling framework.

The following sections in this chapter provide a detailed description of the study locations and the data used for models development and validation, the MG modeling framework and the equilibrium solution concept in MG, the multi-agent adversarial inverse reinforcement learning (MA-AIRL) methodology applied to model cyclist-pedestrian interactions, the Multi-Agent Deep Reinforcement Learning (MA-DRL) utilized to estimate road user optimal policies, the various performance metrics used to assess the accuracy of the developed models, the development of

the MA-DRL shared space simulation tool, and the results of estimating cyclist and pedestrian reward functions and policies during their interactions in shared space facilities.

6.2 Study Locations and Dataset Description

The video data described in detail in chapter 5 of this thesis were used in the analysis. The data were recorded from three shared space locations in the United States and Canada. The first shared space is located in New York City, New York. The other two shared space locations are at Robson Square in downtown Vancouver, BC. More details about the selected study locations and the video data are provided in Figure 5.1 and chapter 5 of this thesis.

6.2.1 Data for Training and Testing

Trajectories of cyclists and pedestrians involved in head-on and crossing interactions were extracted from 51 hours of the collected video footage using the computer vision methodology described in chapter 3. The demonstration of road user tracking is shown in Figure 5.2. More details about the data used for training and testing the models presented in this study are provided in chapter 5 of this thesis.

6.3 Cyclist and Pedestrian Behavioural Profiles Extraction

The variables that are used to describe cyclist and pedestrian behaviours in shared space are based on previous studies of pedestrian behaviour, cyclist behaviour, cyclist and pedestrian interaction at shared space, and mix traffic interaction, as discussed in chapter 4 of this thesis. In this study, the behaviours of both cyclists and pedestrians are considered for modeling. Thus, the

behaviour variables that describe road user state and action were calculated for each road user (i.e., cyclist and pedestrian). The variables that are used to describe each road user's state include its speed, speed difference, longitudinal and lateral distances, interaction angle, and distance to the destination. The road user's yaw rate and acceleration profiles are used to describe its action. More details about the methodology for calculating road user behaviour variables are provided in chapters 3 and 4 of this thesis.

6.4 Modeling Approach

This section provides details about the methodology of developing the proposed multi-agent-based simulation model for road user interactions. First, an overview of Markov Game (MG) theory, on which the model is based, is described. Then, detailed descriptions of both MG types and solution concepts for MG are provided. The formulation of the multi-agent adversarial inverse reinforcement learning algorithm (MA-AIRL), which is used to recover road users' multi-agent reward functions, is then presented. Lastly, the formulation of the multi-agent deep reinforcement learning (MA-DRL), which is used to estimate road users' policies, is then presented. The single-agent Gaussian process IRL (GPIRL) algorithm, which is used as a baseline to compare the multi-agent model's accuracy, is described in chapter 5 of this thesis. Details about the evaluation metrics, which are used to assess the accuracy of the predicted trajectories, are described in chapter 4 of this thesis.

The modeling procedure is illustrated in Figure 6.1. Road user trajectories (e.g., cyclists and pedestrians) are firstly extracted, and their decision process parameters (i.e., states and actions)

are then computed. Under the multi-agent modeling framework, the multi-agent adversarial neural network discriminators are used to recover multi-agent road users' reward functions given their demonstrations (Yu, et al., 2019), and the Multi-agent Actor-Critic with Kronecker factors (MACK) DRL algorithm is used to estimate their policies (Song, et al., 2018). In this approach, the learning algorithms converge when the induced multi-agent road user policies (i.e., learned policies or trajectories) by the recovered multi-agent reward functions are indistinguishable from their actual policies. In the single-agent Gaussian Process IRL, only the single-agent (e.g., cyclist) decision process parameters are computed. In this approach, the cyclists' reward function is recovered by maximizing the likelihood of their demonstrations over the GP reward function (Levine & Koltun, 2012; Levine, et al., 2011). The cyclists' policy is then learned using the single-agent synchronous actor-critic (A2C) DRL (Mnih, et al., 2016; Wu, et al., 2017), as discussed in chapter 5 of this thesis. Subsequently, a dual multi-agent and single-agent shared space simulation tool is developed, incorporating the learned multi-agent and single-agent policies to simulate road user interaction for the validation dataset.

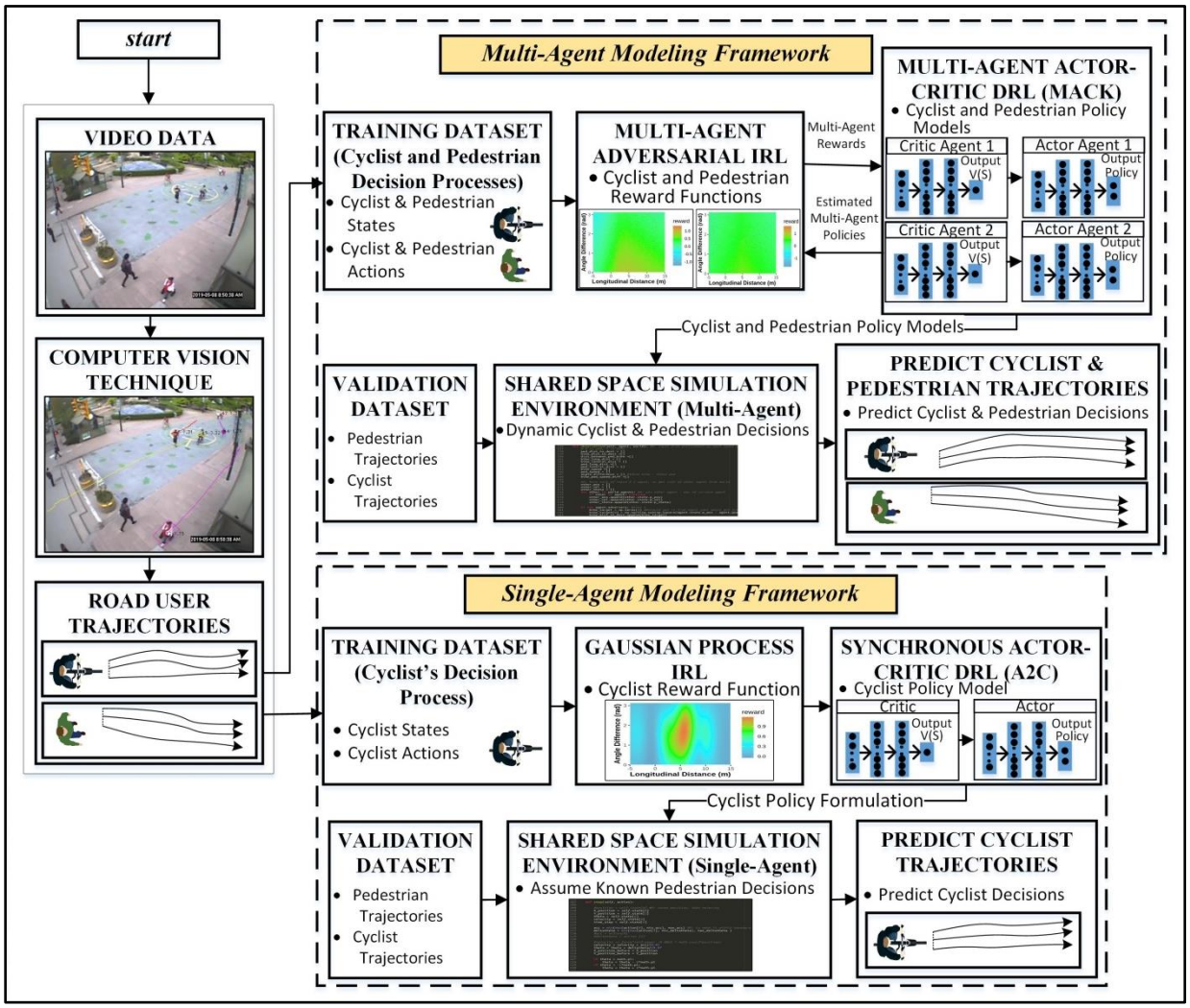


Figure 6.1 Flowchart for the single-agent and multi-agent modeling frameworks (Alsaleh & Sayed, 2021b)

6.4.1 Markov Game Concept

Markov Game framework extends the traditional game-theoretic framework and MDP to the case of modeling the decisions' sequences of N interacting agents simultaneously (Nowé, et al., 2012). In the traditional game-theoretic framework, the sequential decision-making process with the state transition concept cannot be handled. A Markov Game consists of a tuple $(N, \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \eta, \boldsymbol{r}, \gamma)$, where N is the number of interacting agents, $\boldsymbol{S}$ is a set of states, and $\boldsymbol{A} =$

$\{A_i\}_{i=1}^{N}$ is N sets of action spaces, one for each agent. The state transition function $P: S \times A_1 \times ... \times A_N \rightarrow P(S)$ describes the stochastic transition between states. In MG, the state transition is controlled by the agent's current state and the actions taken by each interacting agents. The function $\boldsymbol{P}(\boldsymbol{S})$ is a set of state transition probability distributions over the set of states $\boldsymbol{S}$. The initial states are determined by a distribution function $\eta \in \wp(S) \rightarrow [0, 1]$. The discount factor $\gamma \in [0,1)$ describes the effect of the reward one time the in the future, and it is set to 0.975 similar to (Alsaleh & Sayed, 2020; Alsaleh & Sayed, 2021a). Each agent i holds its policy $\pi_i(a_i|s)$ and receives a bounded reward by making decision defined as $r_i: S \times A_1 \times ... \times A_N \rightarrow \mathbb{R}$. Let any bold variable without a subscript i denotes the concentration of all variables for all agents. Let the subscript $-i$ denotes all agents except for the agent i. The objective of each agent is to maximize its expected return $E_{\boldsymbol{\pi}}\left[\sum_{t=1}^{T} \gamma^t r_{i,t}\right]$, where t is the time steps. The expected return for the agent i for a state-action pair can be defined as $ExpRet_i^{\pi_i, \boldsymbol{\pi}_{-i}}(s_t, \boldsymbol{a}_t) = E_{s^{t+1:T}, \boldsymbol{a}^{t+1:T}}[\sum_{l \geq t} \gamma^{l-t} r_i(s^l, \boldsymbol{a}^l)|s_t, \boldsymbol{a}_t, \boldsymbol{\pi}]$, where l is a time step $l \in \{t, T\}$.

6.4.2 Types of Markov Games

In the MDP modeling framework, agents follow solely their intended goals that maximize their rewards/ utilities with no anticipation of other agent reactions. However, in MG, each agent anticipates other agents' reactions and chooses actions that maximize its own/ or group utility. The behaviour of agents in multi-agent systems varies from being cooperative to being competitive. In the fully cooperative multi-agent setting, agents share the same goal and act to maximize their group utility (i.e., the sum of their rewards). On the other hand, in the

competitive system setting, agents can have non-aligned goals and solely act to maximize their own utility. Agent's goals are a reflection of their intentions, and thus it is challenging to describe their goals solely based on their apparent behaviour. In practice, agents in multi-agent system can show both cooperative and competitive behaviours at different circumstances. For example, the competitive agents may temporarily cooperate when a higher reward is achieved by cooperation. The cooperative agents may also show competitive behaviour if an agent gets a higher reward by acting in a way that detriment the other agents (Jan't Hoen, et al., 2005). In the context of shared space areas, interacting road users can show both cooperative and competitive behaviour (i.e., mixed behaviour). For example, in the crossing interaction, road users apparently cooperate to avoid collision, however, a competitive behaviour may be observed in their aggressiveness to take crossing priority and avoid severe decelerating.

6.4.3 Solution Concepts for Markov Games

In MDP, an agent's optimal policy depends only on its surrounding environment, and the sequences of decisions of other agents in the environment are treated as known. However, in the multi-agent systems as of MG, each agent's optimal policy depends on other agents' policies, which may directly affect their rewards and environments. In such multi-agents systems, agents' capability to cooperate or compete between them is vital in finding an appropriate solution to the problem. Nash Equilibrium (Hu & Wellman, 1998; Nash, 1951) is the most popular solution concept for MG, where each agent policy is the best response for other agents' decisions (i.e., no agent can gain by unilateral deviation). A Correlated Equilibrium (CE) solution concept for MG (Ziebart, et al., 2011; Aumann, 1974) is a generalization of Nash Equilibrium (NE), which

further requires agents' actions in each state to be independent. In CE, agents coordinate their actions to achieve higher expected rewards, i.e., each agent knows the conditional distribution of other agent s' actions given its own action $P(\boldsymbol{a}_{-i}^{t}|a_{i}^{t})$. Thus, no agent can achieve higher expected reward by unilateral deviation of its own policy. The Nash Equilibrium (NE) and Correlated Equilibrium (CE) solution concepts might lack the ability to handle non-optimal or bounded rationality agents' behaviour. Specifically, they are incompatible with the Maximum Entropy principle of IRL (MaxEnt IRL) (Ziebart, et al., 2008) as they assume that agents never take sub-optimal actions or behaviour. MaxEnt IRL accounts for sub-optimal behaviour and provides a probabilistic framework for solving the ambiguity issue in recovering agents' reward function from demonstrations (i.e., many reward functions can explain agents' behaviour given a sample of demonstrations). The MaxEnt IRL principle finds a trajectory distribution with maximum entropy that matches the agents' reward expectation.

To tackle the issue of bounded rationality, McKelvey and Palfrey (McKelvey & Palfrey, 1995; McKelvey & Palfrey, 1998) proposed the Logistic Quantal Response Equilibrium (LQRE), which is a stochastic generalization of the Nash Equilibrium (NE) and Correlated Equilibrium (CE). In LQRE, the best response function becomes probabilistic, and the better actions are more likely to be chosen than worse actions, but the best actions are not chosen with certainty. Thus, agents choose actions with higher expected return with higher probability. Yu et al. (Yu, et al., 2019) proposed the Logistic Stochastic Best Response Equilibrium (LSBRE) solution concept for MG, based on the LQRE solution concept in game theory (McKelvey & Palfrey, 1995; McKelvey & Palfrey, 1998) and Gibbs sampling (Hastings, 1970). The LSBRE solution concept

is compatible with MaxEnt IRL principle and can handle sub-optimal and bounded rationality agents' behaviour. The LSBRE solution concept allows the characterization of rational joint policies (i.e., trajectory distribution) induced by multi-agent reward functions.

6.4.3.1 The Logistic Stochastic Best Response Equilibrium (LSBRE) Solution Concept

The LSBRE solution for MG characterizes road users' trajectory distribution in line with the MaxEnt principle. Specifically, the trajectory distribution induced by the LSBRE policies can be presented as an energy-based formulation. Thus, the probability of observing a trajectory is proportional to the exponential sum of its rewards (i.e., road user trajectories are not always optimal but can be sub-optimal). The LSBRE solution is based on repeatedly applying entropy-regularization (i.e., a stochastic best response mechanism) for agents' policies (Yu, et al., 2019).

Let $\{\pi_i^t\}_{i=1}^T$ donates a set of time-dependent stochastic policies for each agent i. A policy $\pi_i(.;s)$ of each agent i is a probability distribution over the action space of the agent A_i given the agent's states. The state-action value function (Q) for each agent i is recursively defined as given by Equation 6.1 (Yu, et al., 2019).

$$Q^{\boldsymbol{\pi}^{t+1:T}}(s^t, a_i^t, \boldsymbol{a}_{-i}^t) = r_i(s^t, a_i^t, \boldsymbol{a}_{-i}^t) + E_{s^{t+1}\sim \mathrm{P}(.|s^t,\boldsymbol{a}^t)}[\mathcal{H}(\pi_i^{t+1}(.|s^{t+1})) + E_{\boldsymbol{a}^{t+1}\sim\boldsymbol{\pi}^{t+1}(.|s^{t+1})}[Q^{\boldsymbol{\pi}^{t+2:T}}(s^{t+1}, \boldsymbol{a}^{t+1})]] \quad (6.1)$$

where $\mathcal{H}(\pi) = E_\pi[-log\pi(a|s)]$ is the policy entropy.

The LSBRE solution for *T*-horizon MG is a sequence of *T* stochastic joint policies ($\boldsymbol{\pi}^t$) over $(A_1 \times ... \times A_N)^{|S|}$. Considering T Markov chains, the state of the *t-th* Markov chain at step k is given by $(z_i^{t,(k)}: S \to A_i)_{i=1}^N$, where $z_i^{t,(k)}(s)$ takes values in A_i. The state of the Markov chain is updated using Equation 6.2. Each LSBRE stochastic joint policy $\boldsymbol{\pi}^t: S \to \mathcal{P}(A_1 \times ... \times A_N)$ is given by Equation 6.3 (Yu, et al., 2019).

$$z_i^{t,(k+1)}(s^t) \sim \mathrm{P}_i^t\left(a_i^t \middle| \boldsymbol{a}_{-i}^t = \boldsymbol{z}_{-i}^{t,(k)}(s^t), s^t\right) = \frac{\exp\left(\lambda Q_i^{\boldsymbol{\pi}^{t+1:T}}\left(s^t, \boldsymbol{a}_i^t, \boldsymbol{z}_{-i}^{t,(k)}(s^t)\right)\right)}{\sum_{a_i'} \exp\left(\lambda Q_i^{\boldsymbol{\pi}^{t+1:T}}\left(s^t, a_i', \boldsymbol{z}_{-i}^{t,(k)}(s^t)\right)\right)} \tag{6.2}$$

$$\boldsymbol{\pi}^t(a_1, \dots, a_N | s^t) = P\left(\bigcap_i \{z_i^{t,(\infty)}(s^t) = a_i\}\right) \tag{6.3}$$

where $s^t \in S$, $\lambda \in \mathbb{R}^+$is the rationality level and it set to 1, and $\{\mathrm{P}_i^t\}_{i=1}^N$ is the conditional LSBRE stationary joint distribution.

6.4.4 Multi-Agent Adversarial Inverse Reinforcement Learning

In the Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL) modeling framework, the trajectories of *N* interacting road users ($\{\mathcal{T}\}_{i=1}^N$) are used as expert demonstrations. The expert demonstrations form the whole source of data required for the learning algorithm. In the context of IRL, road user trajectories are defined as sequences of state and actions according to Equation 6.4.

$$\mathcal{T}_i = \{(s_i^t, a_i^t)\}_{t=1}^T \tag{6.4}$$

where s_i^t and a_i^t are the state and the observed action of agent i at a time $t \in (0, T)$, T is the trajectory length.

In the MA-AIRL framework, experts' demonstrations can be rationalized by maximizing their likelihood with respect to the LSBRE stationary distribution, which is induced by the ω −parameterized reward functions $\{r_i(s, \boldsymbol{a}; \omega_i)\}_{i=1}^N$. The probability of a trajectory ($\mathcal{P}(\mathcal{T})$) to be generated by the LSBRE policies in the MG is given by Equation 6.5 (Yu, et al., 2019).

$$\mathcal{P}(\mathcal{T}) = \eta(s^1) \,.\, \prod_{t=1}^{T} \boldsymbol{\pi}^t(\boldsymbol{a}^t|s^t; \boldsymbol{\omega}) \,.\, \prod_{t=1}^{T} P(s^{t+1}|s^t, \boldsymbol{a}^t) \tag{6.5}$$

where $\eta(s^1)$ is the initial state distribution, $\boldsymbol{\pi}^t(\boldsymbol{a}^t|s^t; \boldsymbol{\omega})$ are the LSBRE unique stationary joint distributions (i.e., learned agents policies) induced by $r_i(s, \boldsymbol{a}; \omega_i)\}_{i=1}^N$, and $P(s^{t+1}|s^t, \boldsymbol{a}^t)$ is the transition dynamics.

Reward functions' parameters ($\boldsymbol{\omega}$) can be obtained using the MaxEnt principle by maximizing the likelihood of expert trajectories using the distribution defined in Equation 6.5 as given by in Equation 6.6. However, direct optimization of the joint maximum likelihood estimation objective given in Equation 6.6 is intractable as no closed-form solution can be obtained. The joint likelihood $\boldsymbol{\pi}^t(\boldsymbol{a}^t|s^t)$ can be approximated by pseudolikelihood, i.e., product of conditional policies, $\prod_{i=1} \pi_i^t(a_i^t|\boldsymbol{a}_{-i}^t, s^t)$. Optimizing the maximum pseudolikelihood objectives can be instead achieved by optimizing a surrogate loss, which in a variational approximation of the pseudolikelihood, as given by Equation 6.7 (Yu, et al., 2019).

$$max_{\omega} E_{\mathcal{T} \sim \boldsymbol{\pi}_E}[\sum_{t=1}^{T} log \boldsymbol{\pi}^t(\boldsymbol{a}^t|s^t; \omega)] \tag{6.6}$$

$$E_{\boldsymbol{\pi}_E}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\omega}} r_i(s^t, \boldsymbol{a}^t; \omega_i)\right] - \sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\omega}} \log Z_{\omega_i} \tag{6.7}$$

where $\boldsymbol{\pi}^t(\boldsymbol{a}^t|s^t)$ is the joint policy, and Z_{ω_i} is the partition function of the trajectory distribution. Computing the partition function is intractable for large or continuous state-action spaces or when the environment dynamics are unknown, which is the case in this study.

Instead, an efficient sampling-based approximation to the MaxEnt principle can be used to recover the multi-agent road users' reward functions and their corresponding policy. Thus, the partition function can be estimated using importance sampling with an adaptive sampler q_θ. Over the MA-AIRL framework, a ω −parameterized discriminator (D_{ω_i}) and θ −parameterized generator (q_{θ_i}) are defined for each agent i. Each discriminator takes a particular structure as given by Equation 6.8, and forms as a binary classifier that classifies expert and policy trajectories. Each generator (i.e., parameterized policy) is an adaptive importance sampler. The role of the generator is to generate trajectories that fool the discriminator (i.e., difficult to be distinguished from the expert ones). The discriminators and the generators are trained to maximize the objectives given by Equation 6.9 and Equation 6.10 (Yu, et al., 2019), respectively.

$$D_{\omega_i}(s, a) = \frac{\exp\left(f_{\omega_i}(s,\boldsymbol{a})\right)}{\exp\left(f_{\omega_i}(s,\boldsymbol{a})\right)+q_{\theta_i}(a_i|s)} \tag{6.8}$$

$$max_{\boldsymbol{\omega}} E_{\boldsymbol{\pi}_E}\left[\sum_{i=1}^{N}\log\left(D_{\omega_i}(s,\boldsymbol{a})\right)\right] + E_{\boldsymbol{q_\theta}}\left[\sum_{i=1}^{N}\log\left(1 - D_{\omega_i}(s,\boldsymbol{a})\right)\right] \tag{6.9}$$

$$max_{\boldsymbol{\theta}} E_{\boldsymbol{q_\theta}}\left[\sum_{i=1}^{N}\log\left(D_{\omega_i}(s,\boldsymbol{a})\right) - \log\left(1 - D_{\omega_i}(s,\boldsymbol{a})\right)\right] \tag{6.10}$$

where $q_\theta(a|s)$ is the probability of the adaptive sampler, and it is trained to minimize the Kullback-Leibler (KL) diverge between the trajectory distribution generated by the generator and that induced by the reward functions, f_ω is the learned function and it is trained to estimate the reward function. At optimality, q_θ and f_ω will approximate expert policy and expert policy's advantage function, respectively. The learned function f_{ω_i} of each agent i has the structure given in Equation 6.11 (Fu, et al., 2017).

$$f_{\omega_i,\emptyset_i}(s^t, a^t, s^{t+1}) = g_{\omega_i}(s^t, a^t) + \gamma h_{\emptyset_i}(s^{t+1}) - h_{\emptyset_i}(s^t) \tag{6.11}$$

where g_{ω_i} is the reward estimator, and $h_\emptyset$ is the potential function.

In this algorithm, the learned functions' parameters $(\omega_i, \emptyset_i)$ are trained to maximize the objective function of the discriminators presented in Equation 6.9. These parameters are optimized by training the gradient of the discriminators objective, as given by Equation 6.12 (Fu, et al., 2017).

$$\nabla_{\omega_i} L(\omega_i) = \sum_{t=0}^{T} E_{\boldsymbol{\pi}_E}\left[\nabla_{\omega_i} f_{\omega_i}(s_t, \boldsymbol{a}_t)\right] - E_{\boldsymbol{\mu}}\left[\frac{\hat{\rho}_{\omega_i}(s_t, \boldsymbol{a}_t)}{\hat{\boldsymbol{\mu}}_{\omega_i}(s_t, \boldsymbol{a}_t)} \nabla_{\omega_i} f_{\omega_i}(s_t, \boldsymbol{a}_t)\right] \tag{6.12}$$

where $\nabla_{\omega_i} L(\omega_i)$ is the gradient of the discriminator objective function of agent i, $\hat{\rho}_{\omega_i}(s, \boldsymbol{a}) = \exp(f_{\omega_i}(s, \boldsymbol{a}))\, q_{\theta_i}(a_i|s)$ is a rough density estimate trained on the demonstrations, and $\hat{\boldsymbol{\mu}} = 0.5\, q_{\theta_i}(a_i|s) + 0.5\hat{\rho}_{\omega_i}(s|\boldsymbol{a})$ is a mixture policy between $\hat{\rho}_{\omega_i}$ and policy samples q_{θ_i}.

An illustration of the MA-AIRL algorithm is presented in Figure 6.2. The steps of the algorithm training process are as follow:

1. Obtain expert demonstrations (i.e., road user trajectories) $\boldsymbol{\mathcal{D}}_E = \mathcal{T}_j^E$, and Markov Game parameters $(N, \boldsymbol{S}, \boldsymbol{A}, \eta, \boldsymbol{P}, \gamma)$
2. Initialize the parameters of road users' policies $\boldsymbol{q}$, reward function estimators $\boldsymbol{g}$, and potential functions $\boldsymbol{h}$ with $\boldsymbol{\Theta}, \boldsymbol{\omega}, \emptyset$.
3. Sample trajectories $\boldsymbol{\mathcal{D}}_\pi = \{\mathcal{T}_j\}$ from $\boldsymbol{\pi}$: sample state-action pairs $\mathcal{X}_{\boldsymbol{\pi}}$, $\mathcal{X}_E$ from $\boldsymbol{\mathcal{D}}_\pi$, $\boldsymbol{\mathcal{D}}_E$
4. For each agent $i \in [1, \dots, N]$, update the parameters $\omega_i, \emptyset_i$ to increase the objective function of the discriminators presented in Equation 6.9 using Equation 6.12.
5. For each agent $i \in [1, \dots, N]$, update the reward estimates $\hat{r}_i(s, a_i, s')$ with $g_{\omega_i}(s, a_i)$ using $\hat{r}_i = (\log D_{\omega_i}(s, \boldsymbol{a}) - \log[1 - D_{\omega_i}(s, \boldsymbol{a})])$.
6. For each agent $i \in [1, \dots, N]$, update the policy parameters θ_i with respect to $\hat{r}_i(s, a_i, s')$ using MA-DRL, as will be described in the next subsection of the paper.
7. Terminate if converge; otherwise, go back to step 3.
8. The outputs are the learned road users' policies $\boldsymbol{\pi}_{\boldsymbol{\Theta}}$ and reward functions $\boldsymbol{g}_{\boldsymbol{\omega}}$.

6.4.4.1 Multi-Agent Actor-Critic with Kronecker Factors Deep Reinforcement Learning

The optimal policies of road users are learned during the training process of the MA-AIRL algorithm using Multi-Agent Deep Reinforcement Learning (MA-DRL). The agents' reward functions that guide the adaptive samplers' training process (i.e., road user policies) are computed from agents' discriminators. The adaptive samplers $\{q_\theta\}_{i=1}^N$ are trained using the Multi-agent Actor-Critic with Kronecker factors (MACK) (Song, et al., 2018), which is a version of the Actor-Critic with Kronecker factors Trust Region (ACKTR) (Wu, et al., 2017). The MACK is a model-free natural policy gradient algorithm in DRL, which uses the framework of

centralized training with decentralized execution. Thus, during the training process, additional information of other agents' actions and states is used to reduce each agent's learning variance; however, at the execution time, only each agent's state is used to predict its own action. In this algorithm, two deep neural networks are used for each agent. The first deep neural network, called the Critic neural network, is used as a function approximator for the states' value function V^{π}. The second deep neural network, called the Actor neural network, is used as a function approximator for road user's policy π (i.e., the adaptive sampler q_{θ}). The Critic neural network judges the actions taken by the policy neural network (Actor), and therefore encourages or discourages them. The objective of each policy neural network is to maximize the agent's expected discounted cumulative return. The Kronecker-factored approximated natural policy gradient method is used to update each agent's policy parameters θ_i with respect to the reward estimates as given by Equation 6.13 (Song, et al., 2018). This method directly adjusts each policy parameters θ_i in order to maximize its objective (expected cumulative return $J(\theta_i) = E_{\pi_i}[R_{\pi_i}]$), which is equivalent to Equation 6.10. In this algorithm, the advantage function of each agent $A^{\pi}(s_t, a_t)$ that provide a relative measure of actions values at each state is defined as given by Equation 6.14 (Song, et al., 2018).

$$\nabla_{\theta_i} J(\theta_i) = E_{\pi_i}\left[\sum_{t=0}^{\infty} \nabla_{\theta_i} log \pi_{\theta_i}(a_t|s_t).A_{\emptyset_i}^{\pi_i}(s_t, \boldsymbol{a}_t)\right] \tag{6.13}$$

$$A_{\emptyset_i}^{\pi_i}(s_t, \boldsymbol{a}_t) = \sum_{j=0}^{k-1} \gamma^j r_i(s_{t+j}, \boldsymbol{a}_{t+j}) + \gamma^k V_{\emptyset_i}^{\pi_i}\left(s_{t+k}, \boldsymbol{a}_{-i,t+k}\right) - V_{\emptyset_i}^{\pi_i}\left(s_t, \boldsymbol{a}_{-i,t}\right) \tag{6.14}$$

Where $\nabla_{\theta_i} J(\theta_i)$ is the policy gradient estimator for agent i, π_{θ_i} is the learned policy for agent i, $V_{\emptyset_i}^{\pi_i}\left(s_k, a_{-i,k}\right)$ is the baseline neural network value function approximator for agent i, which

utilizes the information of other agents policies ($\boldsymbol{a}_{-i,t}$) for variance reduction. The approximated natural policy gradient is also used to optimize the Critic neural networks parameters $\emptyset_i$ by minimizing the squared difference between the calculated and predicted returns. In this approach, the Kronecker-factored approximate curvature (K-FAC) is used to optimize the neural networks. Figure 6.2 illustrates the process of the MACK algorithm with the centralized training process.

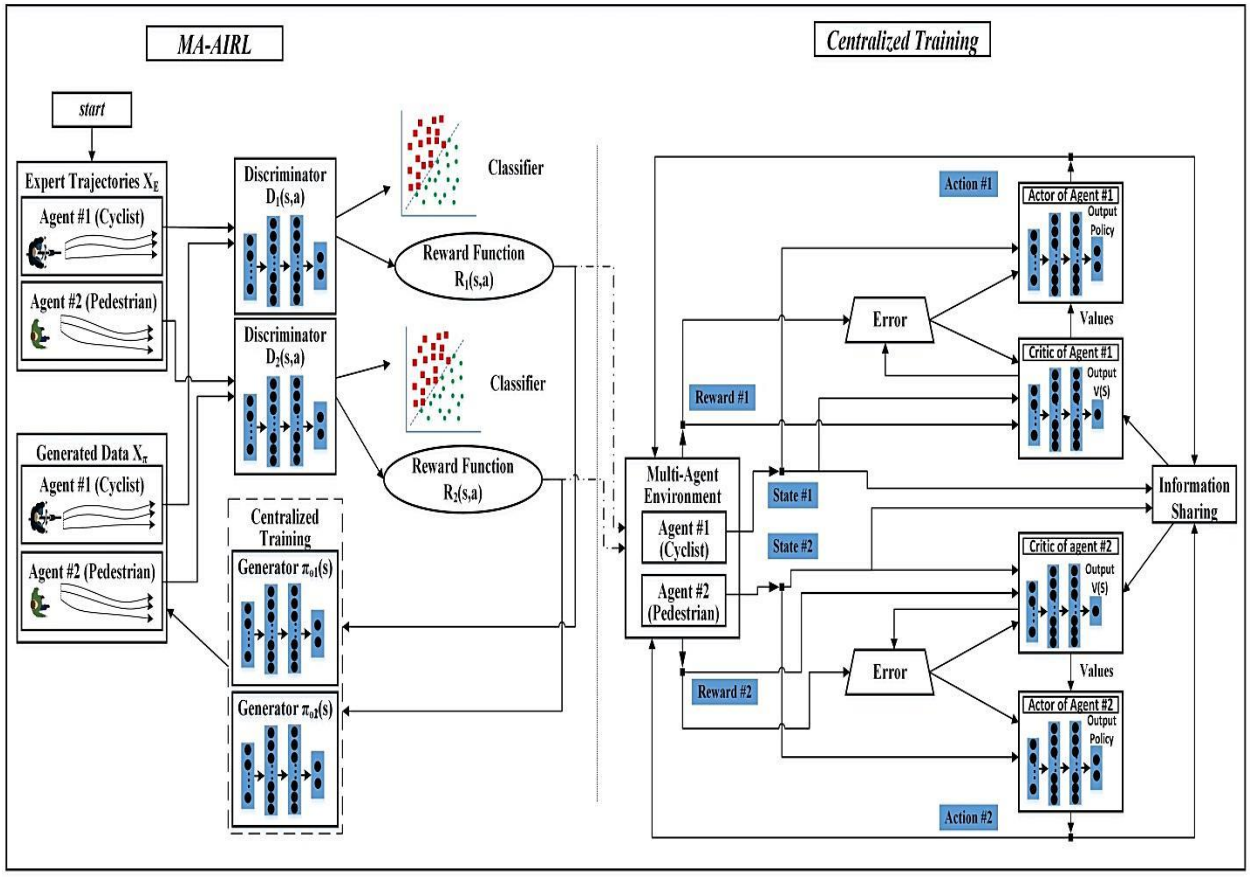


Figure 6.2 Illustration of the MA-AIRL algorithm and the centralized training process (Alsaleh & Sayed, 2021b)

6.4.5 Baseline Model: Single-Agent Gaussian Process Inverse Reinforcement Learning

In this study, the continuous Gaussian Process Inverse reinforcement learning (GPIRL) algorithm (Levine & Koltun, 2012; Levine, et al., 2011) is selected as a baseline for evaluating the performance of the proposed MA-AIRL algorithm. The GPIRL algorithm, which is based on the MDP framework, is used to model cyclist interactions with pedestrians at shared space facilities. The Gaussian Process (GP) reward function can capture the nonlinearity in road user behaviour. More details about the Gaussian Process (GP) modeling methodology are provided in chapter 5 of this thesis.

6.4.5.1 Single-Agent Actor-Critic Deep Reinforcement Learning

The cyclists' policies are learned after obtaining their GP reward functions using the single-agent synchronous actor-critic (A2C) deep reinforcement learning (Mnih, et al., 2016; Wu, et al., 2017). Similar to the multi-agent policy training, in this approach, two deep neural networks are defined for the cyclists; the Critic and the Actor neural networks. The Critic neural network approximates states' value function V^{π}, while the Actor neural network approximates cyclists' policy π. The training objective of the Actor neural network is to maximize the cyclists' expected cumulative return $J(\theta) = E_{\pi}[R_{\pi}]$. More details about the synchronous actor-critic (A2C) deep reinforcement learning are provided in chapter 5 of this thesis.

6.5 Accuracy Metrics

The accuracy of the simulated trajectories is assessed based on the relative distance from the actual trajectories. In this study, two evaluation metrics are used to assess the accuracy of the

simulated trajectories, including the *Mean Absolute Error (MAE)* and the *Hausdorff Distance (HausD)*. The methodology for calculating these metrics is presented in detail in chapter 4 of this thesis.

6.6 Results and Discussion

6.6.1 Road User Behavioural Characteristics and State and Action Identification

The trajectories of cyclists and pedestrians that are involved in head-on and crossing interactions in shared space facilities are analyzed and modeled separately. Most of the interactions analyzed in this work took place in relatively low shared space densities. The decentralized learning approach for the multi-agent learning algorithm (MA-AIRL) is adopted. The decentralized learning approach assumes no prior information is known about the type of road users' behaviour (e.g., cooperative or competitive) before the modeling. Thus, the agents' reward functions are modeled separately. Each road user's states, on which the reward function is based, and actions are defined by set of variables (as described earlier in section 6.2) that describe its behaviour during the interactions.

Figure 6.3 and Figure 6.4 illustrate the distributions of cyclist and pedestrian behaviour variables for the head-on and the crossing interactions, respectively. A negative value of the speed difference or yaw rate indicates a higher pedestrian speed than the cyclist, or a counter-clockwise rotation of the road user, respectively. A negative value of the longitudinal distance means that the conflict situation is resolved and the agent has already crossed the other interacting agent. Descriptive statistics of these distributions are elaborated in Table 6.1. Analyzing road users'

variable distributions can reveal their behavioural characteristics. Generally, cyclists and pedestrians that are involved in head-on interaction have higher average speeds and speed differences compared to those involved in the crossing interaction. In the head-on interaction, cyclists and pedestrians are associated with a slightly larger average acceleration compared to the crossing interaction. The average magnitude of pedestrian yaw rate is slightly larger in head-on interaction compared to the crossing interaction. However, cyclists in head-on interaction are associated with a slightly lower magnitude of yaw rate compared to the crossing interaction.

Variable	Head-on Interaction mean[SD*]	Crossing Interaction mean[SD]
$Cyclist\ Speed\ [m/s]$	3.735 [1.590]	3.034 [1.566]
$Pedestrian\ Speed\ [m/s]$	1.558 [0.663]	1.184 [0.576]
$Speed\ Difference\ [m/s]$	2.177 [1.712]	1.851 [1.532]
$Angle\ Difference^{*}\ [rad]$	3.131 [0.649]	1.619 [0.535]
$Cyclist\ Lateral\ Distance\ [m]$	3.197 [2.343]	2.983 [2.956]
$Cyclist\ Longitudinal\ Distance\ [m]$	3.641[7.982]	3.404 [6.839]
$Pedestrian\ Lateral\ Distance\ [m]$	3.460 [2.890]	5.722 [5.235]
$Pedestrian\ Longitudinal\ Distance\ [m]$	3.262 [7.700]	0.685 [4.873]
$Cyclist\ Distance\ to\ Destination\ [m]$	7.155 [5.690]	5.823 [5.641]
$Pedestrian\ Distance\ to\ Destination\ [m]$	3.128 [2.567]	2.590 [2.927]
$Cyclist\ Acceleration\ [m/s^2]$	0.160 [1.915]	0.123 [1.798]
$Pedestrian\ Acceleration\ [m/s^2]$	0.043 [2.731]	-0.002 [2.271]
$Cyclist\ Yaw\ Rate\ [rad/s]$	-0.019 [0.790]	-0.073 [0.853]
$Pedestrian\ Yaw\ Rate\ [rad/s]$	-0.017 [1.213]	0.014 [0.953]

* Standard deviation

* Absolute angle difference for the crossing interaction.

Table 6.1 Cyclist and pedestrian behaviour parameter descriptive statistics (Alsaleh & Sayed, 2021b)

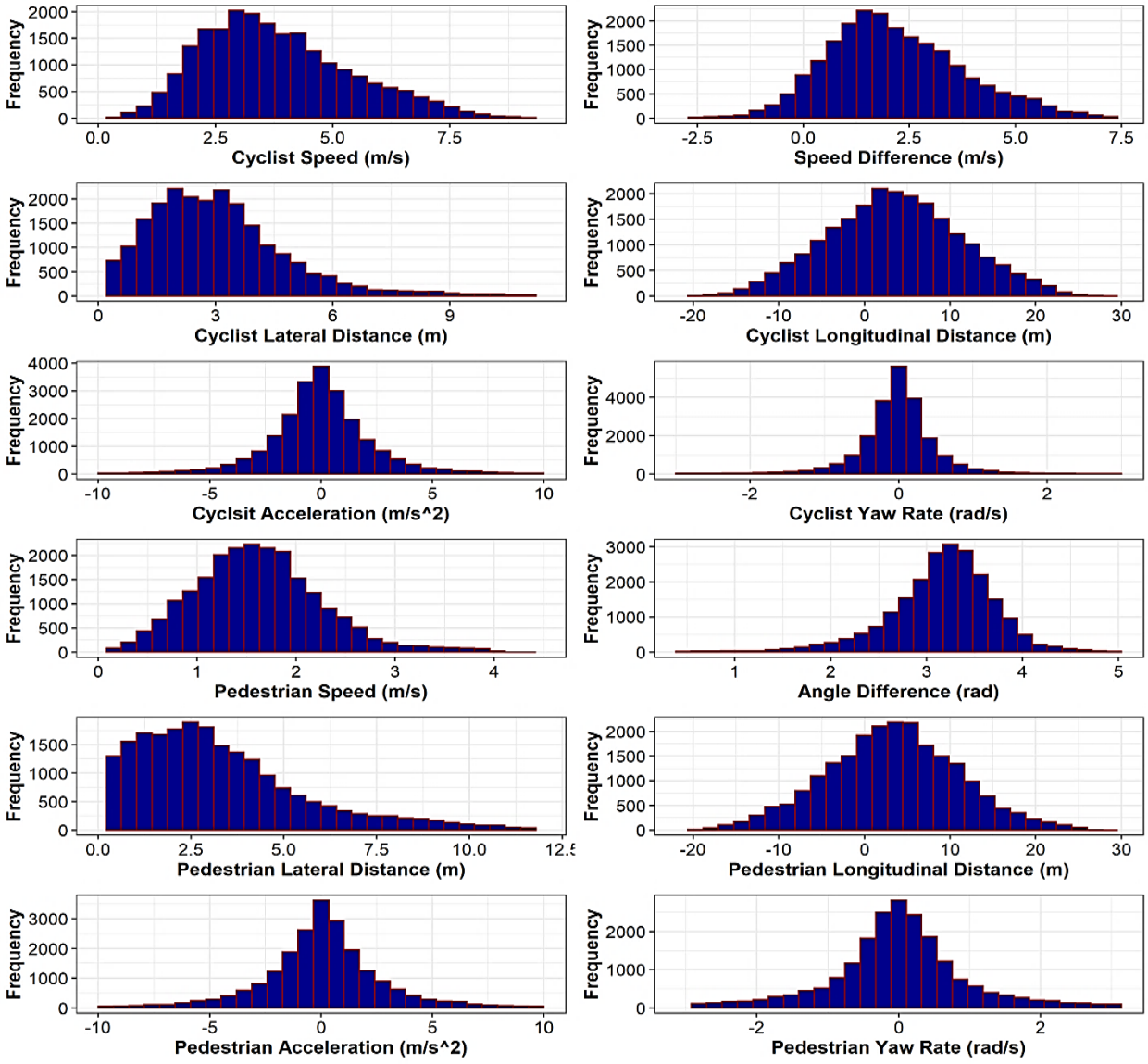


Figure 6.3 Histograms of cyclist and pedestrian behaviour parameters for the head-on interaction (Alsaleh & Sayed, 2021b)

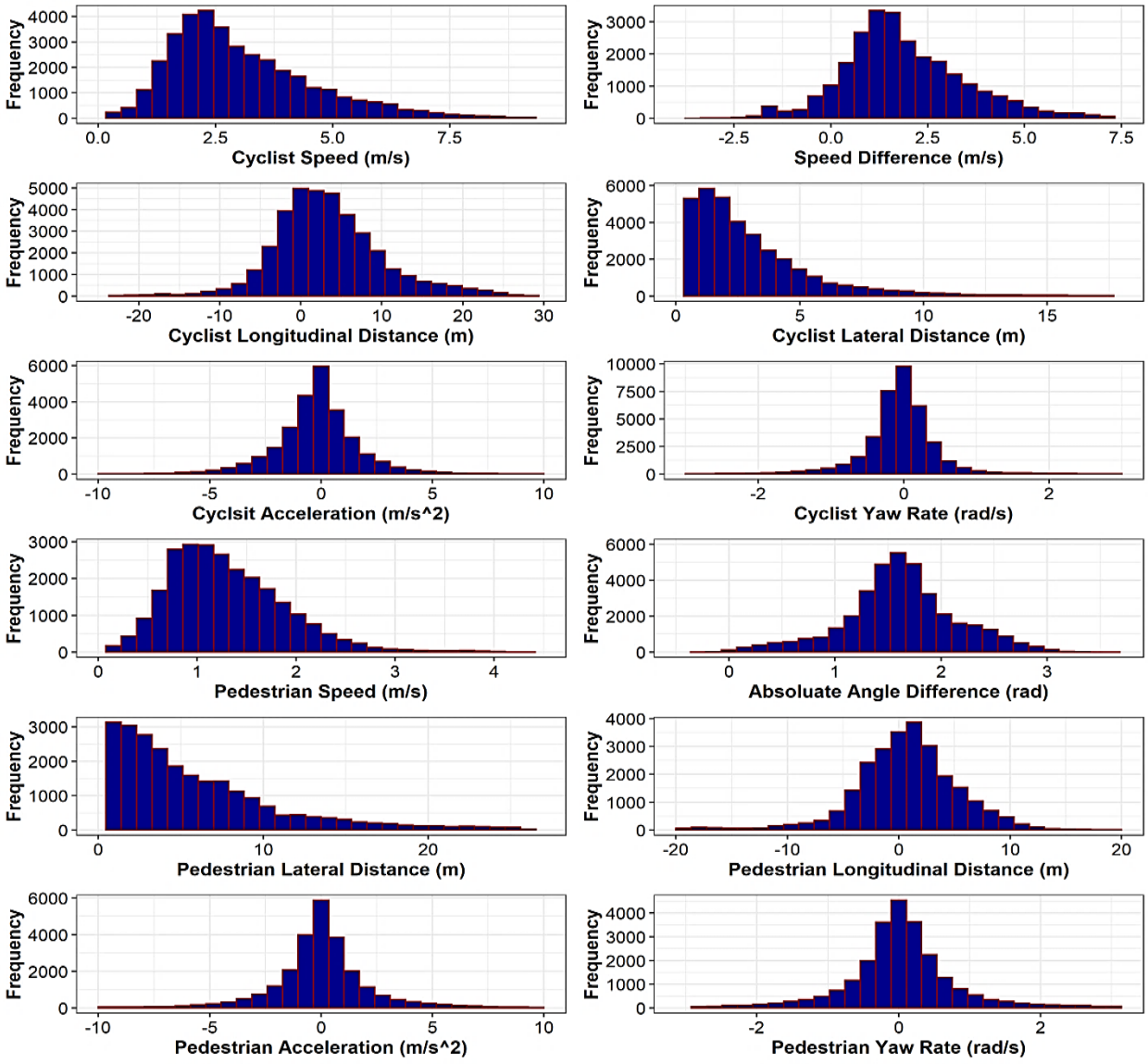


Figure 6.4 Histograms of cyclist and pedestrian behaviour parameters for the crossing interaction (Alsaleh & Sayed, 2021b)

6.6.2 Recovery of Multi-Agent and Singe-Agent Reward Functions

6.6.2.1 Cyclist and Pedestrian Head-on Interaction

The recovered multi-agent (cyclist and pedestrian) reward functions from applying the MA-AIRL algorithm are presented in Figure 6.5. The reward functions are presented as bivariate state feature spaces, while holding the other states at their mean values according to Table 6.1. Road user trajectory discriminators, which are formalized as deep adversarial neural networks, captured highly nonlinear reward structures of road user interaction. The recovered multi-agent reward functions can infer road user behaviour and give insight into their cooperative and competitive behaviour during their interactions at shared space facilities. However, inferring road user behaviour using their reward functions becomes more challenging in the multi-agent framework due to the existence of the equilibrium solution concept instead of a single optimality solution concept as in the single-agent modeling framework.

The multi-agent cyclist and pedestrian reward functions, which are presented using their distance to destination and longitudinal distance state features, suggest that states having larger distance to destination and longitudinal distance values have lower rewards compared to the states with smaller distance to destination and longitudinal distance values (i.e., large negative longitudinal distance). These results infer two preferences of road users in such interaction. First, both cyclists and pedestrians prefer to reach their destinations in shared space facilities. Second, they also prefer to resolve the conflict situation and cross each other, where a negative longitudinal distance indicates that both road users have crossed each other. These results agree with the observational behavioral study of cyclist and pedestrian interactions at shared space facilities

(Alsaleh, et al., 2020). It is noteworthy to mention that the inferred road user behaviour from the reward functions presented in Figure 6.5 is concerned with the mean values of other state features. Thus, some road user preferences may differ according to the values of other state features.

However, despite the aligned preferences of road user behaviour, which may show cooperative behaviour, a competitive behaviour can be seen in their tendency to swerve to resolve the conflict situation. A road user swerving maneuver is associated with changing its movement direction, i.e., yaw angle, leading to a sudden change (reduction) in the computed longitudinal distance and may lead to a deviation from the shortest path to its destination. The road user's deviation from its path may lead to a reduction in the cumulative total reward gained by the road user. A similar result is reported in (Liu, et al., 2014). Such cooperative and competitive behaviours can be handled by the equilibrium solution concept in the multi-agent setting. The multi-agent reward functions, which are presented using the distance to destination and longitudinal distance of each road user, show that pedestrians have a wide range of longitudinal distance preferences compared to cyclists at any given distance to destination value. This result suggests that pedestrians have more tendency and flexibility to swerve (i.e., taking swerving evasive action) than cyclists in this situation.

Moreover, the multi-agent cyclist and pedestrian reward functions, which are presented using their angle difference and longitudinal distance state features, support this finding. The multi-agent reward functions show a wide range of longitudinal distance preferences for pedestrians

compared to cyclists at any given value of the angle difference feature. This supports the higher tendency and flexibility of pedestrians to swerve than cyclists in this situation. It is noteworthy to mention that the angle difference between road users depends on their movement directions (i.e., yaw angles). However, the longitudinal distance for a road user depends on its own movement direction and the other road users' position (i.e., the movement direction of a road user does not affect the other road user's longitudinal distance computation).

The multi-agent cyclists' reward function, which is presented using their speed difference and longitudinal distance state features, indicates a higher cyclists' preference to keep intermediate speed difference related to pedestrian before resolving the conflict situations (i.e., at small positive longitudinal distances). However, after resolving the conflict situations, where the longitudinal distances become negative, cyclists prefer to increase their speed difference related to pedestrians. Moreover, the multi-agent cyclists' reward function, which is presented using their speed and longitudinal distance state features, shows a similar trend to the speed difference and longitudinal distance features reward. The reward function suggests that cyclists tend to have lower cycling speeds before resolving the conflict situations compared to after resolving the conflict situations. However, it suggests lower cyclists' preferences to have very high cycling speeds after resolving conflict situations. Instead, cyclists prefer to keep intermediate cycling speed. This finding is expected as in the head-on conflict situation, cyclists prefer to reduce their cycling speed to avoid collisions with pedestrians. However, after resolving the conflict situation, cyclists prefer to accelerate to their preferred cycling speeds. These results are consistent with (Alsaleh, et al., 2020).

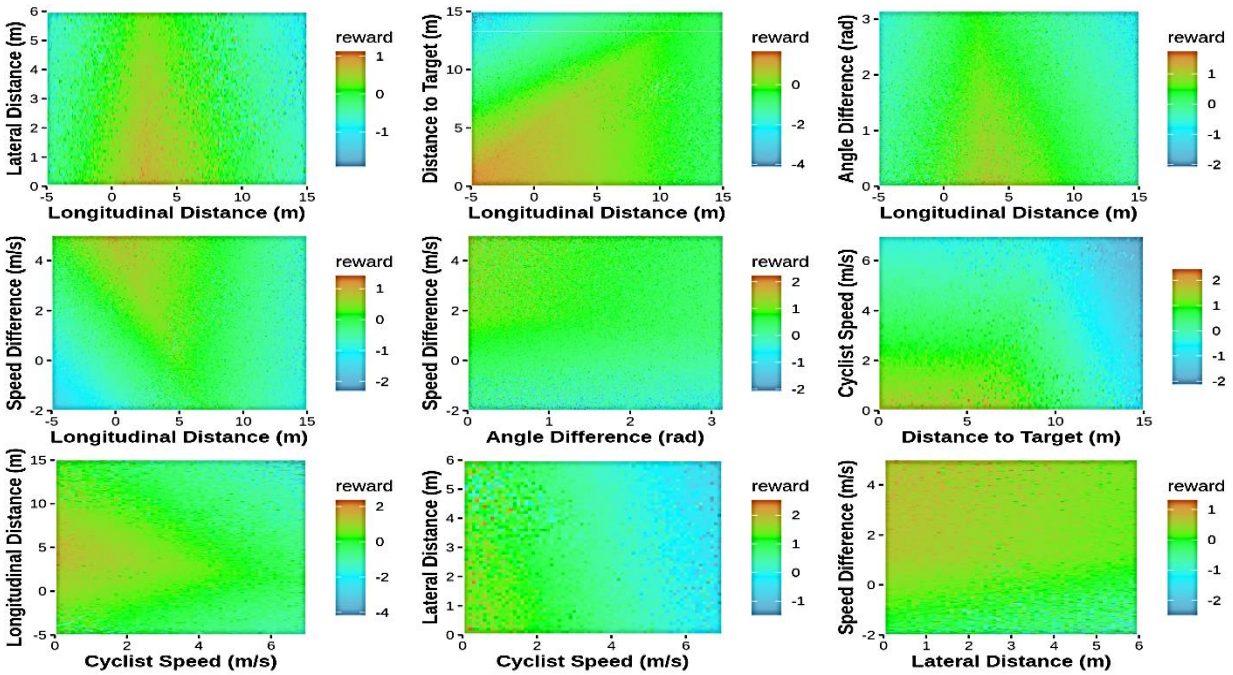
The multi-agent pedestrians' reward functions, which are presented using their speed difference or speed and longitudinal distance state features, suggest that pedestrians prefer to have slightly lower speed and speed difference before resolving the conflict situations compared to after resolving the conflict situations. This result agrees with previous observational behavioural studies (Alsaleh, et al., 2020; Hussein & Sayed, 2015). Previous studies indicated that pedestrian swerving maneuver in head-on interactions is not associated with a significant change in pedestrian speed. Despite the aligned trend of speed difference preferences between cyclists and pedestrians before and after resolving the conflict situations, pedestrians may show competitive behaviour in their preferences to have low speed differences relating to cyclists.

For the single-agent modeling approach, the reward function is recovered for cyclists using the GPIRL algorithm assuming fixed pedestrian trajectories. Similar to the multi-agent reward functions, the single-agent GP reward function is presented as bivariate state feature spaces, while holding the other states at their mean values, as shown in Figure 6.6. The optimality solution concept in the single-agent modeling framework leads to a more direct inference of cyclist behaviour using its reward function.

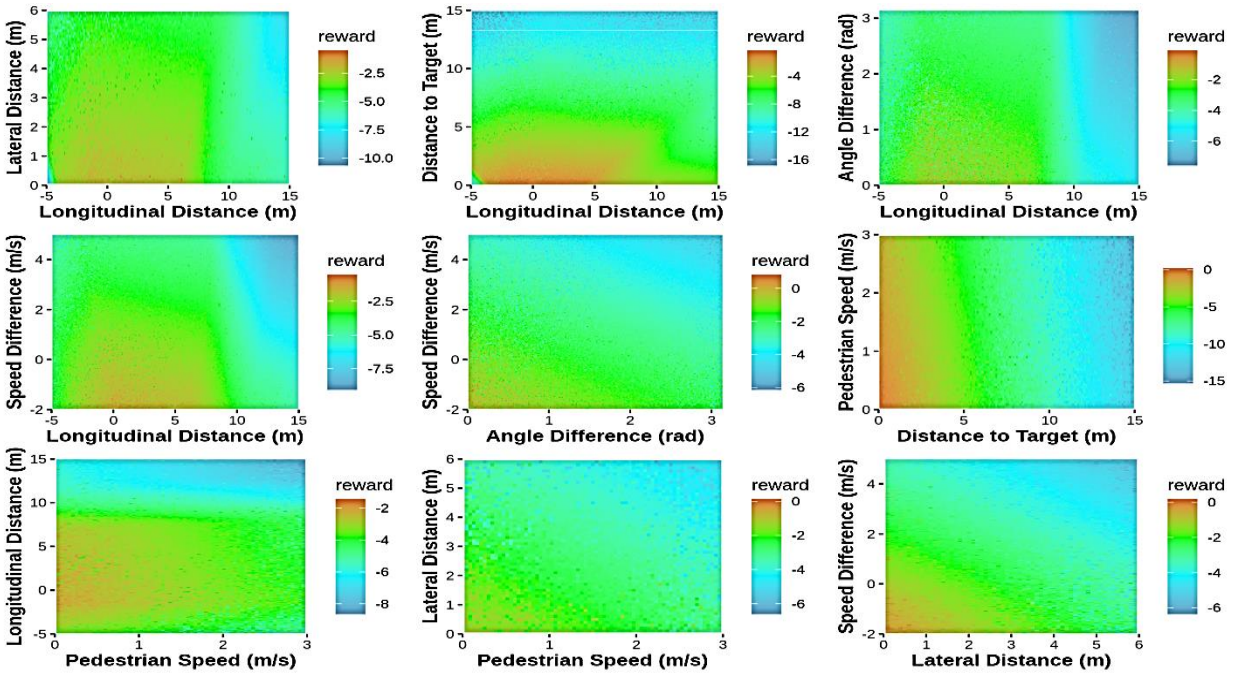
The GP reward function, which is presented using the distance to destination and longitudinal distance state features, suggests similar cyclist preferences to the multi-agent reward functions. The GP reward function suggests higher cyclists' preferences to reach their destinations and resolve conflict situations. Moreover, the GP reward function, which is presented using the speed difference and longitudinal distance state features, indicates higher cyclists' preferences to keep

intermediate speed difference related to pedestrians before the conflict points. However, after resolving the conflict situation, cyclists prefer to increase their speed difference related to pedestrians.

In the single-agent modeling framework, the recovered reward function did not capture the equilibrium solution concept similar to the multi-agent setting, which can be attributed to the assumption of fixed pedestrian trajectories. Thus, the existence of non-stationary pedestrian behaviour in the data set may affect the estimation of the single-agent reward function. For example, the single-agent GP reward functions show a wide range of longitudinal distance preferences at any given value of the angle difference or distance to destination features at the current mean value of the other state features. This result indicates higher stochasticity in cyclist's behavioural model comparing to the multi-agent model. This would likely lead to higher stochasticity in cyclists' behaviour and less certainty in their selected evasive actions to avoid collisions in conflict situations. The lack of the ability to handle the equilibrium solution between road users in the single-agent framework can limit the transferability and impact the model's accuracy.



(a) Bike Reward Function



(b) Pedestrian Reward Function

Figure 6.5 Multi-agent reward functions (at mean feature values) for the head-on interaction (Alsaleh & Sayed, 2021b)

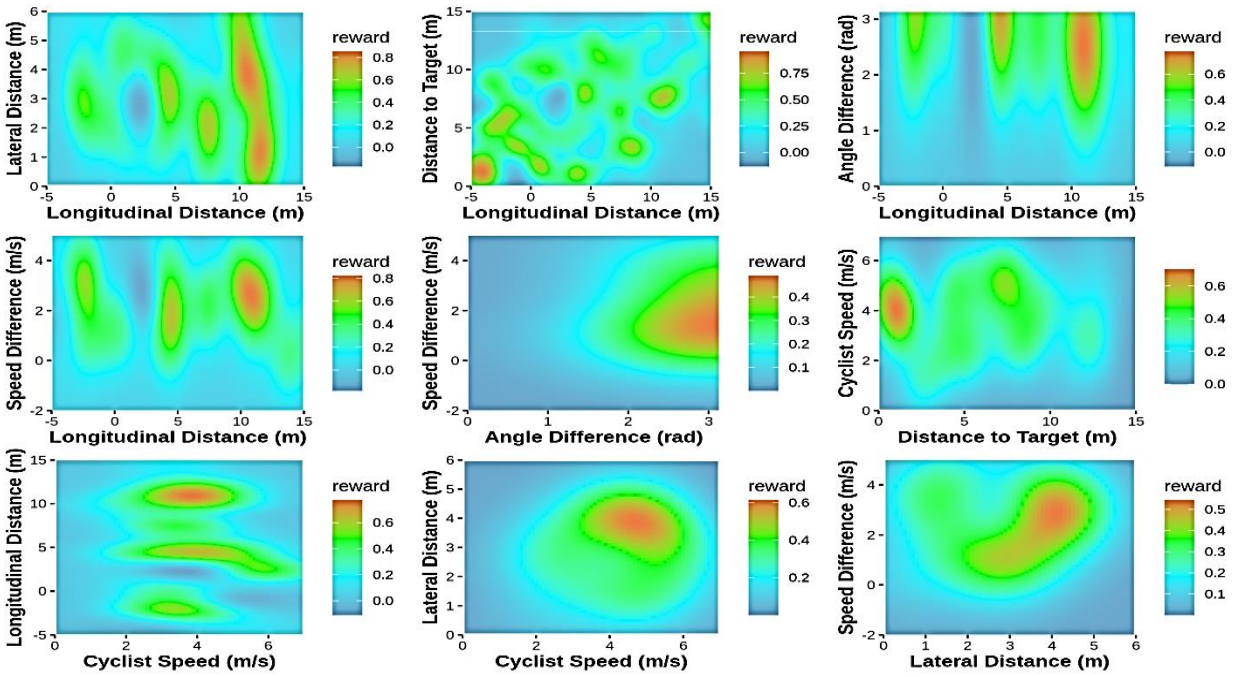


Figure 6.6 Cyclist single-agent reward function (at mean feature values) for the head-on interaction (Alsaleh & Sayed, 2021b)

6.6.2.2 Cyclist and Pedestrian Crossing Interaction

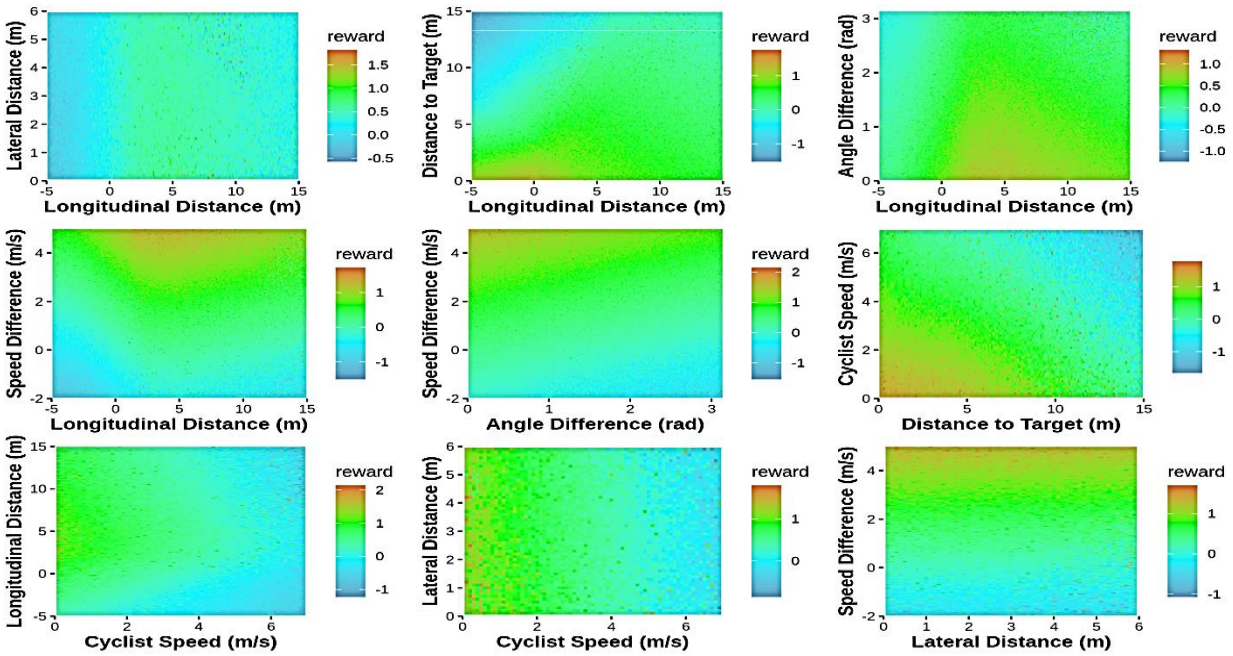
Similar to the head-on interaction, the recovered multi-agent adversarial neural network and single-agent GP reward functions of road users in the crossing interactions are presented in Figure 6.7 and Figure 6.8, respectively. The reward functions are presented as bivariate state feature spaces, while holding the other states at their mean values. The multi-agent cyclist and pedestrian reward functions, which are presented using their distance to destination and longitudinal distance state features, suggest that cyclists and pedestrians have higher preferences to reach their destinations. Moreover, they suggest that cyclists and pedestrians have different longitudinal distance preferences in the current situation of the crossing interaction. The results suggest a higher cyclist preference to keep negative longitudinal distances. However, pedestrian

highly prefers to keep small positive longitudinal distances. These preferences are associated with the cases where road users cross first (i.e., negative longitudinal distance) or yield for the other road user (i.e., small positive longitudinal distance). These results agree with (Alsaleh, et al., 2020). Road users' tendency to yield to each other may suggest cooperative behavior to avoid collisions. However, competitive behaviour may also be seen in their tendency to cross first and avoid severe changes in their motion behaviour. Such cooperative and competitive behaviours can be handled by the equilibrium solution concept in the multi-agent setting.

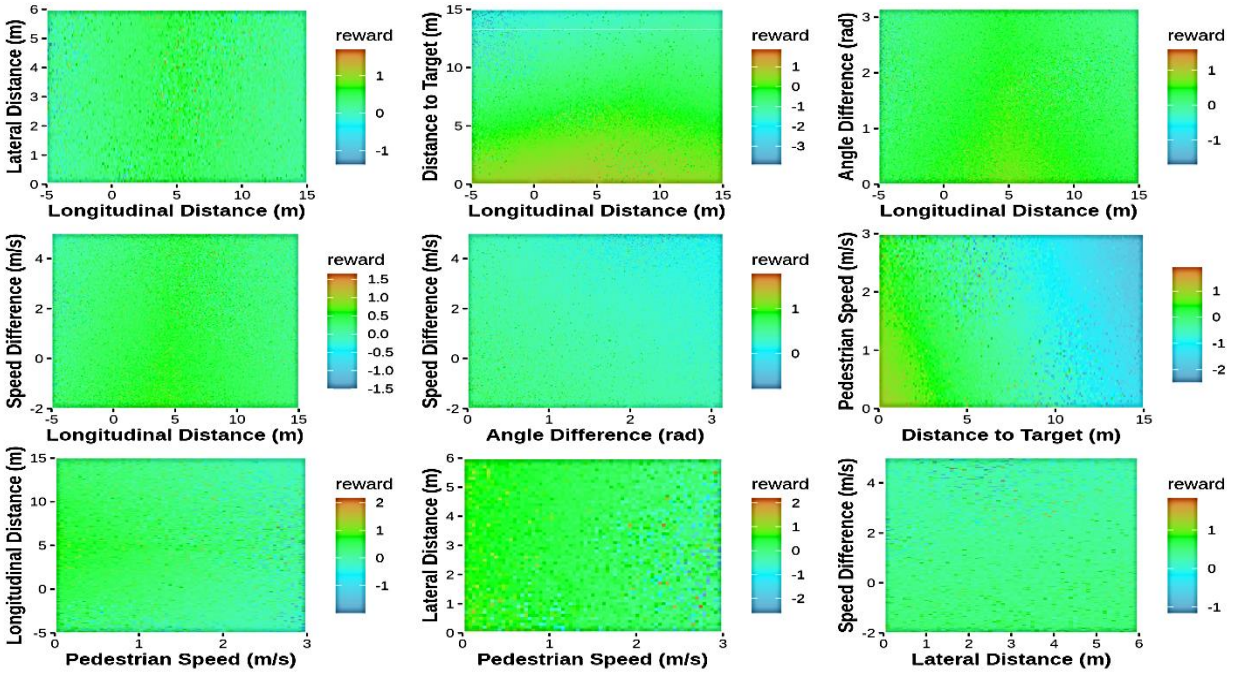
The multi-agent cyclist and pedestrian reward functions, which are presented using their speed difference and longitudinal distance state features, show different preferences for cyclists and pedestrians in the crossing interaction. The reward functions suggest that cyclists prefer to keep higher speed differences related to pedestrians than lower speed differences. However, pedestrians show the opposite preference. These different preferences for speed difference may suggest a competitive behaviour between cyclists and pedestrians, which can be handled by the equilibrium solution concept in the multi-agent setting. For road user speed, the multi-agent cyclist and pedestrian reward functions, which are presented using their speed and longitudinal distance state features, suggest that both road users tend to maintain low cycling and walking speed while being closer to the conflict point. This result is in agreement with (Alsaleh, et al., 2020). Moreover, the multi-agent cyclist and pedestrian reward functions, which are presented using their angle difference and speed difference state features, suggest that states with lower angle differences have higher rewards than the states with an intermediate angle difference of

1.57 rad (90 degrees). This result indicates that the crossing interaction involves a swerving maneuver of at least one road user in the current situation.

For the single-agent modeling approach, the GP reward function, which is presented using cyclist distance to destination and longitudinal distance state features, shows similar preferences to the multi-agent reward function. However, the GP reward function suggests that states with small positive longitudinal distance values have higher rewards than the states with negative longitudinal distance values. This result indicates a higher cyclist's preference to yield for pedestrians instead of crossing first in the current situation of the crossing interaction. This result contradicts the inferred cyclist preference from the multi-agent reward functions. This is likely due to the lack of ability of the single-agent modeling framework to capture the equilibrium solution concept similar to the multi-agent framework. Moreover, the GP reward functions, which are presented using the cyclist speed or speed difference and longitudinal distance state features, indicate higher cyclist preferences to have intermediate cycling speed and speed difference related to pedestrians. The GP cyclists' reward function, which is presented using their speed difference and angle difference state features, indicates higher cyclists' preferences to keep intermediate angle difference with pedestrians.



(a) Bike Reward Function



(*b) Pedestrian Reward Function*

Figure 6.7 Multi-agent reward functions (at mean feature values) for the crossing interaction (Alsaleh & Sayed, 2021b)

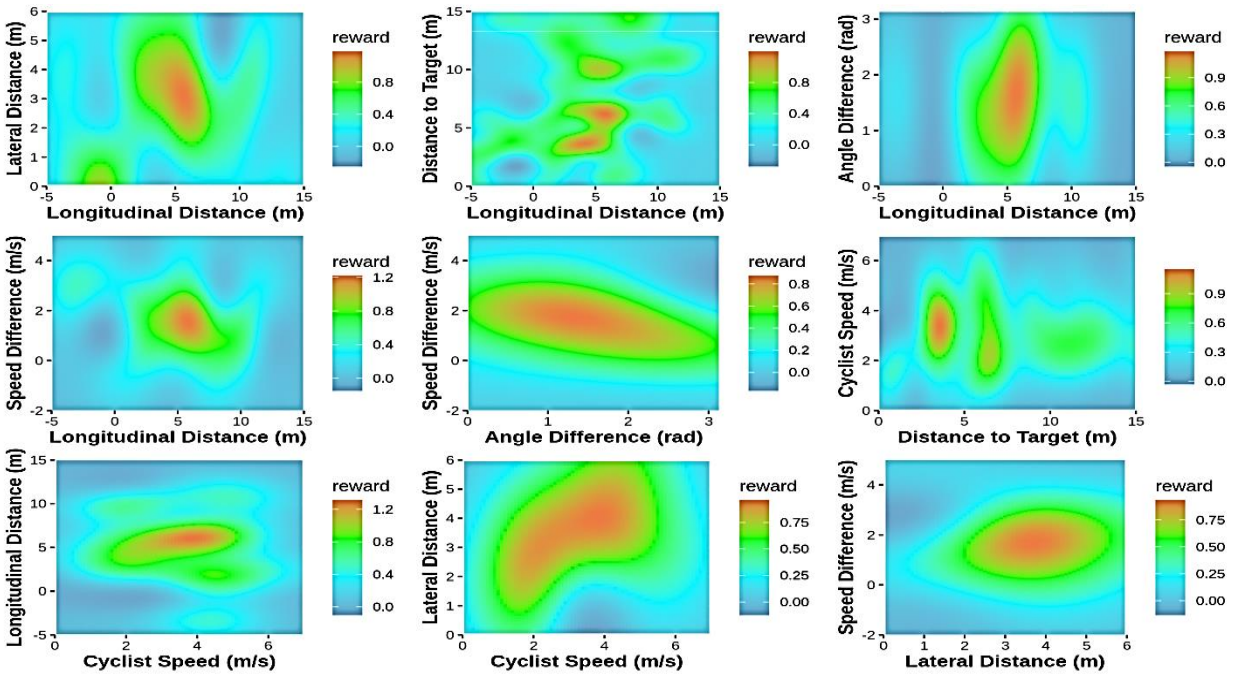


Figure 6.8 Cyclist single-agent reward function (at mean feature values) for the crossing interaction (Alsaleh & Sayed, 2021b)

6.7 Models Comparison and Validation

6.7.1 Multi-Agent Microsimulation Tool Development and Policy Estimation

In this study, the multi-agent policies (i.e., sequences of cyclist and pedestrian decisions) are learned during the training process of the MA-AIRL algorithm using the MACK DRL. In the single-agent modeling framework, cyclists' policy is learned after obtaining their GP reward function using the A2C DRL. A dual single-agent and multi-agent integrated simulation tool is developed for road user interactions at shared space facilities using Python (Python Software Foundation, 2019). A virtual customized shared space environment is developed for each case (i.e., single-agent and multi-agent) using OpenAI Gym (Brockman, et al., 2016; Lowe, et al., 2017). The virtual simulation environments operate at a rate of 30 HZ. The flow chart of the

simulation tool is presented in Figure 6.9. First, the simulation environment initializes road users with information about their initial positions, velocities, yaw angles, and destinations. The simulation environment then computes road users' state features (e.g., longitudinal and lateral distances) and their associated reward values. Then road users take appropriate decisions based on sampling from their optimal learned policies. Lastly, the states of road users are updated based on their selected decisions. In the single-agent environment, pedestrians' decisions are considered known over time.

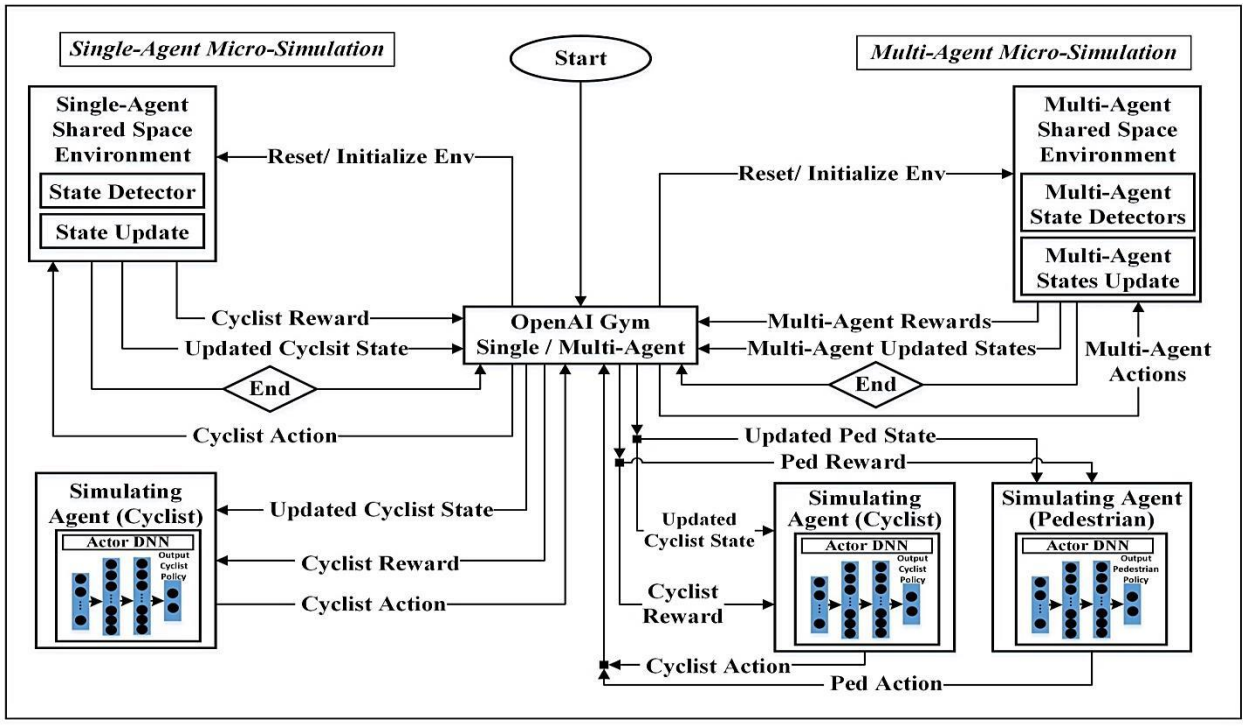


Figure 6.9 Dual single-agent and multi-agent integrated microsimulation tool flowchart (Alsaleh & Sayed, 2021b)

The developed single-agent GP and multi-agent adversarial models are used to simulate the validation dataset's trajectories. In the single-agent model, pedestrian trajectories are considered part of the passive environment. However, both road users (i.e., cyclist and pedestrian) take actions simultaneously in the multi-agent model. The models' accuracy is assessed in terms of cyclist behaviour parameters, as presented in Table 6.2. Overall, both models predict cyclist speeds more accurately than cyclist relative distances. The multi-agent model led to a significantly more accurate prediction of cyclists' behaviour compared to the single-agent model. For example, in the head-on interaction, the multi-agent model improves the prediction accuracy of cyclists' speed, longitudinal and lateral distances by about 32.7% (16%), 53% (38.7%), and 33.3% (21.1%) in terms of *MAE* (*HuasD*), respectively. The correspondent prediction improvement for the crossing interaction are found to be about 33.9% (26.5%), 56.3% (38.6%), and 43.7% (25.5%), respectively.

Moreover, besides being a more realistic approach for modeling and simulating road user interactions, the multi-agent model led to a more accurate prediction of the interaction's behavioural aspects (i.e., evasive action mechanisms). Examples of the predicted road user trajectories for the head-on and crossing interactions and the actual correspondent trajectories are presented in Figure 6.10 and Figure 6.11, respectively. Generally, the figures show that the multi-agent model predicts more accurate road user behaviour and their evasive action mechanisms in both interaction types compared to the single-agent model. For example, Figure 6.10(a) shows an example of the head-on interaction, where the cyclist takes an evasive action of a swerving maneuver and then decelerates to avoid a collision with the pedestrian.

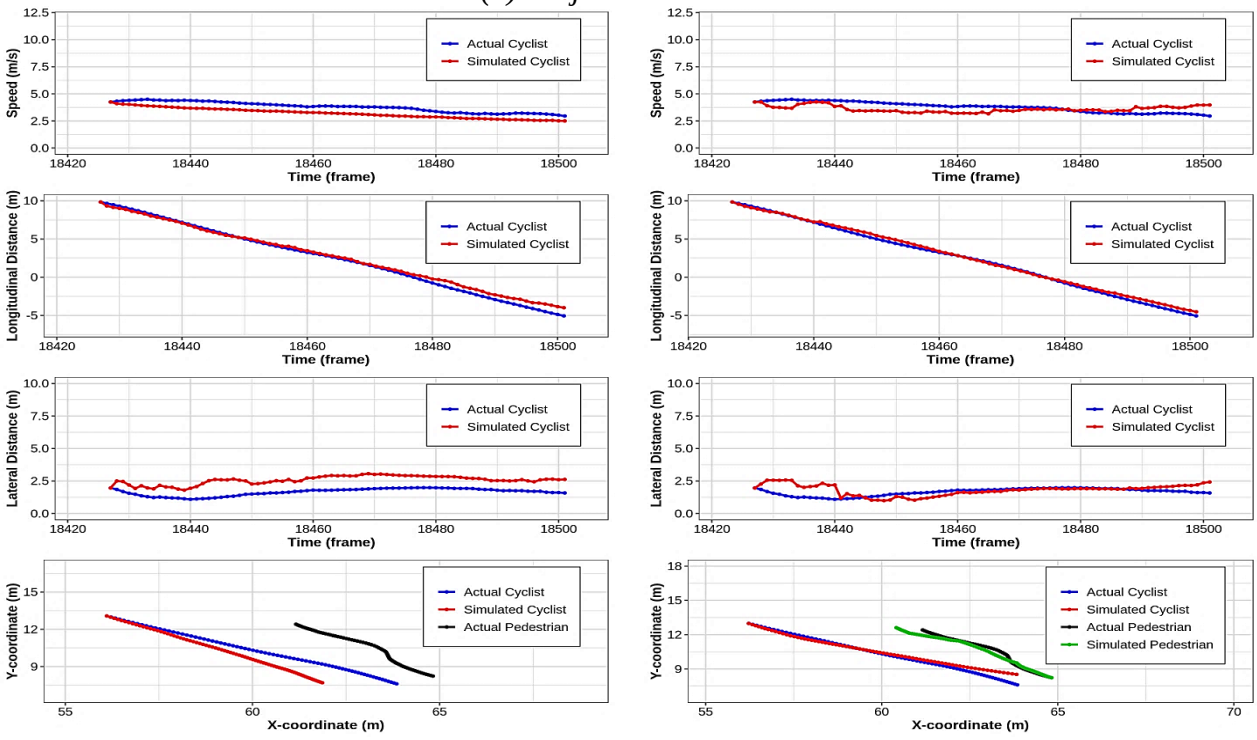
Similar evasive action mechanism is predicted by the multi-agent model. However, the single-agent model led to a less accurate prediction of the cyclist evasive action as it suggests that the cyclist would decelerate and swerve at the same time followed by acceleration. Moreover, Figure 6.10(b) shows an example of the head-on interaction, where the cyclist decelerates during the interaction, while the pedestrian takes the swerving evasive action. Both models predict cyclist decelerating behaviour; however, the single-agent model failed to account for the equilibrium between road users and predicts that the cyclist would adopt a swerving maneuver besides decelerating. However, the multi-agent model adequately learns the equilibrium between road users and predicts the swerving maneuver to be taken mainly by the pedestrian. In both cases, the single-agent model led to a higher deviation from the actual cyclist path compared to the multi-agent model.

Furthermore, Figure 6.11(a) presents an example of the crossing interaction, where the cyclist crosses first (i.e., the pedestrian yield for the cyclist). Both models capture the cyclist's behaviour to cross first; however, the single-agent model failed to capture the cyclist's accelerating behaviour during the crossing. The single-agent model led to a higher deviation from the actual cyclist path compared to the multi-agent model. Figure 6.11(b) presents the case where the cyclist yields for the pedestrian (i.e., pedestrian crosses first). Both models predict similar cyclists' preference to yield for the pedestrian. However, the single-agent model did not capture the cyclist's swerving maneuvering around the pedestrian to avoid a severe conflict. In contrast, the multi-agent model captures the cyclist's swerving behaviour.

Table 6.2 Prediction errors for the single-agent and the multi-agent simulation models (Alsaleh & Sayed, 2021b)

Interaction Type	Cyclist Variable	Multi-Agent Model		Single-Agent Model	
		$Avg.\ MAE$	$Avg.\ Hausdorff\ distance$	$Avg.\ MAE$	$Avg.\ Hausdorff\ distance$
Head-on Interaction	Speed (m/s)	0.35	0.63	0.52	0.75
	Longitudinal distance (m)	0.39	0.65	0.83	1.06
	Lateral distance (m)	0.44	0.75	0.66	0.95
Crossing Interaction	Speed (m/s)	0.37	0.61	0.56	0.83
	Longitudinal distance (m)	0.38	0.70	0.87	1.14
	Lateral distance (m)	0.40	0.73	0.71	0.98

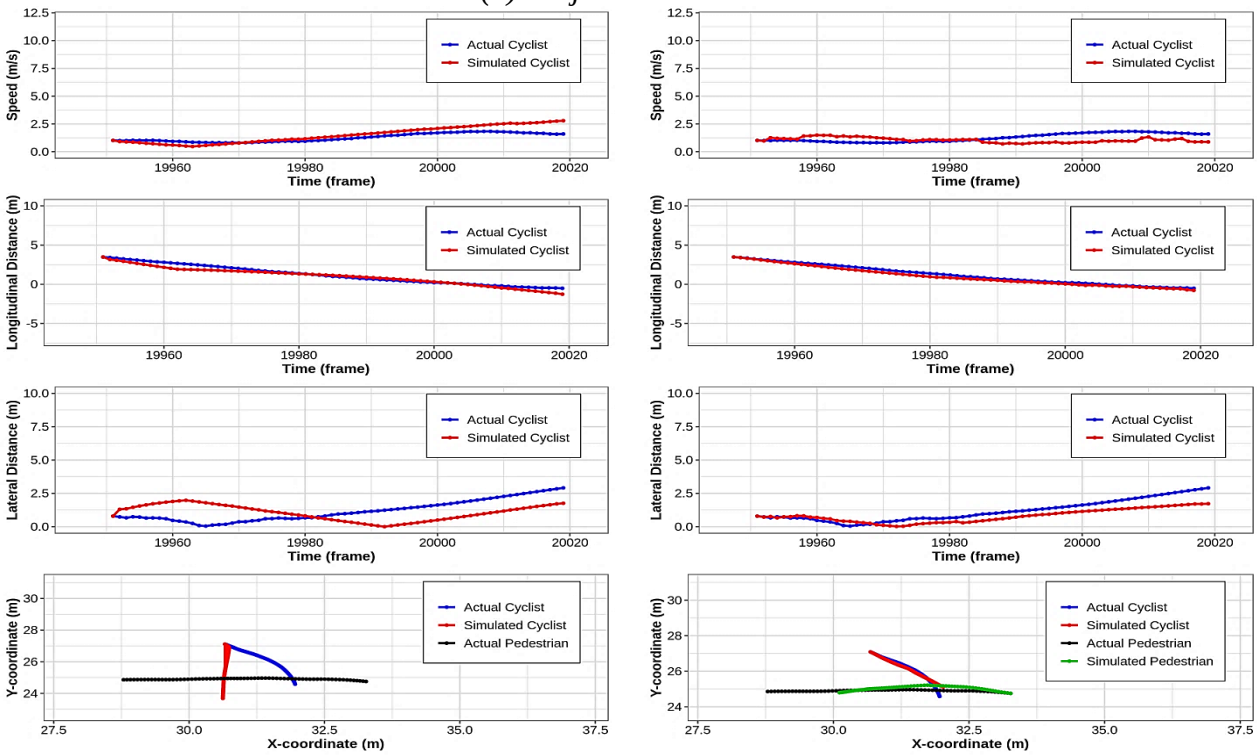
(a) Trajectories ID 196



(b) *Trajectories ID 180*

Figure 6.10 Single-agent and multi-agent simulation for the head-on interaction (Alsaleh & Sayed, 2021b)

(a) Trajectories ID 390



(a) Trajectories ID 350

Figure 6.11 Single-agent and multi-agent simulation for the crossing interaction (Alsaleh & Sayed, 2021b)

6.8 Summary and Conclusion

This chapter presented the details of developing a multi-agent microsimulation model for cyclist and pedestrian interaction in shared space facilities. This study proposed a novel Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL) model for replicating road user interactions at shared space facilities. Compared to the single-agent modeling framework, the multi-agent (MA) modeling framework presents a more realistic approach for modeling and simulating road user interactions and traffic conflicts. Moreover, the MA model can handle non-stationary road user environments and capture the equilibrium solution concept between road users (i.e., heterogonous road user behaviour and social negotiations). The MA model can handle optimal mixed strategy policies, from the perspective of game theory, which are challenging to be handled in the single-agent modeling framework. The used MA modeling approach is based on Markov Games, which considers the stochastic nature of shared space environment and models the sequential decisions (i.e., actions) of road users simultaneously instead of a single time-step payoff modeling like the traditional game-theoretic framework. Furthermore, this study bridges the gap in modeling road users' behaviour by accounting for their rationality, intelligence, and multi-agent non-stationary interaction nature. Accounting for such road user characteristics is crucial in modeling applications like road user interactions at shared space facilities and autonomous vehicle systems.

This study main contribution is in the multi-agent modeling of road user interactions using the Markov Game modeling framework. The multi-agent model's performance is compared to a baseline single-agent Gaussian Process Inverse Reinforcement Leaning (GPIRL) model. Results

show that the utilized multi-agent modeling approach led to a significantly more accurate prediction of road user behaviour and their interaction behavioural aspects (i.e., evasive action mechanisms) compared to the single-agent modeling approach. Moreover, the recovered reward functions based on the single-agent modeling approach did not capture the equilibrium solution concept compared to the multi-agent approach. This is likely attributed to the assumption of fixed pedestrian trajectories in the single-agent framework. The existence of non-stationary in road user environments would likely affect the estimation of the single-agent model reward functions and the road user optimal policies.

In conclusion, the results presented in this chapter confirm the superior performance of the Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL) algorithm under the Markov-Game (MG) framework for the multi-agent modeling of road user interactions in shared space facilities compared to the single-agent GPIRL modeling framework. This paves the way for investigating the ability of different MG-based modeling approaches that utilize different behavioural equilibrium theories in explaining/ predicting road user operational level decisions (e.g., acceleration and steering/yaw rate) and evasive action mechanisms during their interactions in multi-agent settings. The use of other multi-agent imitation learning approaches that utilize the Nash Equilibrium (NE) (Hu & Wellman, 1998; Nash, 1951) solution concept, which assumes optimal road user behavior, can be deployed to investigate the appropriate behavioural equilibrium theory in modeling multi-agent road user interaction modeling in shared space facilities. Moreover, other MG-based modeling approaches, such as the generative adversarial imitation learning, can be beneficial in learning road user optimal policies directly with less

computational cost comparing to the inverse reinforcement learning approach. The development of the multi-agent generative adversarial imitation learning for cyclist and pedestrian interactions in shared spaces facilities, based on the Nash Equilibrium (NE) theory, will be discussed in the seventh chapter of this thesis.

Chapter 7: Do Road Users Play Nash Equilibrium? A Comparison between Nash and Logistic Stochastic Equilibriums for Multi-Agent Modeling of Road User Interactions

7.1 Background

Recently, the Markov Game (MG) modeling framework (Littman, 1994) has been proposed to model the multi-agent road user interactions in shared space facilities (Alsaleh & Sayed, 2021b). In MG, unlike the Markov Decision Process (MDP), each agent's optimal policy is influenced by the other interacting agents' policies. Thus, the optimality notion is defined by an equilibrium concept like the Nash Equilibrium (NE) (Hu & Wellman, 1998; Nash, 1951) or the Logistic Stochastic Best Response Equilibrium (LSBRE) (Yu, et al., 2019; McKelvey & Palfrey, 1998; McKelvey & Palfrey, 1995). Selecting the appropriate equilibrium solution in MG is vital in specifying the agent policies and model accuracy. Under the NE, agents are assumed perfectly rational and aware of all aspects of their environment (i.e., complete information). Thus, each agent's strategy (i.e., sequence of decisions) presents the best strategy with respect to other agents' strategies (i.e., perfectly optimal behavior). However, under the LSBRE, agents' select strategies based on the relative expected utilities. Thus, better strategies have higher probabilities of being selected than the worse ones; however, the best strategies are not selected with certainty. The LSBRE can handle bounded rationality agents and sub-optimal agent behavior. Using the NE solution for bounded rationality agent systems leads to obtaining nonsensical results (Harsanyi & Selten, 1988; Harsanyi, 1968). However, there are limited studies that (1) considered the multiagent nature of road user interactions and their concurrent sequential

decision processes, and (2) investigated the ability of different equilibrium behavioral theories in predicting road user operational-level decisions and evasive-action mechanisms.

The work presented in this chapter aims to model road user interactions in shared spaces using different multiagent modeling approaches that rely on different equilibrium solution concepts. The study assessed the performance of the different behavioral equilibrium theories (i.e., NE and LSBRE) in modeling road user interactions in such facilities. Specifically, the contributions of this study are in: (1) using two multiagent approaches for modeling cyclist and pedestrian interactions in shared spaces, (2) adopting two different equilibrium behavioral theories (i.e., NE and LSBRE) for modeling cyclist and pedestrian interactions, and (3) determining a behavior-based consistent paradigm to model equilibrium in the multiagent transportation systems of road user interactions in shared spaces. The first modeling approach proposed in this study is the Multiagent Generative Adversarial Imitation Learning (MA-GAIL) (Song, et al., 2018), which is based on the NE solution concept, assuming optimal road user behavior. The second approach is the Multiagent Adversarial Inverse Reinforcement Learning (MA-AIRL) (Yu, et al., 2019), which is based on the LSBRE solution concept, assuming sub-optimal (bounded rationality) road user behavior. The second modeling approach is discussed in detail in the sixth chapter of this thesis. Both of these approaches have several advantages: (1) they are based on MG that simultaneously models the sequential decisions of cyclist and pedestrian, (2) the algorithms recover robust cyclist and pedestrian reward functions and model their reward functions and policies using deep neural networks that can capture complex and highly nonlinear behavior, (3) they are applicable for high-dimensional environments (i.e., state and action spaces) with

unknown dynamics (i.e., unrestricted movement), making them suitable for modeling shared space areas, and (4) the algorithms can model cyclist and pedestrian behavior without any prior assumption into their interaction type (e.g., cooperative or competitive).

The following sections in this chapter provide a detailed description of the proposed modeling methodologies, the utilized equilibrium solution concepts in MG, the validation process using real-world data, and the results of estimating cyclist and pedestrian reward functions and policies during their interactions in shared space facilities.

7.2 Study Locations and Dataset Description

The video data described in detail in chapter 5 of this thesis were used in the analysis. The data were recorded from three shared space locations in the United States and Canada. The first shared space is located in New York City, New York. The other two shared space locations are at Robson Square in downtown Vancouver, BC. More details about the selected study locations and the video data are provided in Figure 5.1 and chapter 5 of this thesis.

7.2.1 Data for Training and Testing

Trajectories of cyclists and pedestrians involved in head-on and crossing interactions were extracted from 51 hours of the collected video footage using the computer vision methodology described in chapter 3. The demonstration of road user tracking is shown in Figure 5.2. More details about the data used for training and testing the models presented in this study are provided in chapter 5 of this thesis.

7.3 Cyclist and Pedestrian Behavioural Profiles Extraction

The variables that are used to describe cyclist and pedestrian behaviours in shared space are based on previous studies of pedestrian behaviour, cyclist behaviour, cyclist and pedestrian interaction at shared space, and mix traffic interaction, as discussed in chapter 4 of this thesis. In this study, the behaviours of both cyclists and pedestrians are considered for modeling. Thus, the behaviour variables that describe road user state and action were calculated for each road user (i.e., cyclist and pedestrian). The variables that are used to describe each road user's state include its speed, speed difference, longitudinal and lateral distances, interaction angle, and distance to the destination. The road user's yaw rate and acceleration profiles are used to describe its action. More details about the methodology for calculating road user behaviour variables are provided in Chapters 3 and 4 of this thesis.

7.4 Multiagent Modeling Framework

7.4.1 Markov Games

The interactions between cyclists and pedestrians are modeled using Markov Game (MG) modeling framework (Littman, 1994). MG is a generalization of both Markov Decision Process (MDP) and the traditional game theory to the case of multiple interacting agents with a sequential decision-making process (Nowé, et al., 2012). More details about MG modeling framework are provided in Chapters 6 of this thesis.

7.4.2 Multi-Agent Generative Adversarial Imitation Learning (MA-GAIL)

7.4.2.1 NE Solution Concept

Nash Equilibrium (NE) (Hu & Wellman, 1998; Nash, 1951) is the most common-use equilibrium solution in MG. Under this solution concept, each agent's strategy is the best response to other agents' strategies. This solution is based on a vital assumption of rationality, where agents' are assumed entirely rational and aware of all aspects of the game. Thus, it is challenging to account for bounded rationality and non-optimality behavior using NE solution, making it inconsistent with the Maximum Entropy principle (i.e., sub-optimal agent behavior) of IRL (Ziebart, et al., 2008).

In MG, a set of policies $\{\pi_i^t\}_{i=1}^T$ is NE if no larger reward can be obtained by any agent with unilateral policy changes (Hu & Wellman, 1998). Let **MARL(r)** denotes a set of stationary Markovian NE policies under the RFs $\boldsymbol{r}$ that have the highest entropy. The **MARL(r)** policies can be found by solving the constrained optimization problem given by Equation (7.1) (Song, et al., 2018).

$$\mathbf{MARL}(\mathbf{r}) = \arg\min_{\boldsymbol{\pi}\in\prod,\mathbf{v}\in\mathbb{R}^{S\times N}} f_r(\boldsymbol{\pi},\mathbf{v}) - \mathcal{H}(\boldsymbol{\pi}) \tag{7.1}$$

$$\text{s.t. } v_i(s^t) \geq q_i(s^t, a_i^t) \quad \forall\, i \in [N], s \in S, a_i \in A_i$$

where $\mathbf{v}$ is the value function, $\boldsymbol{q}$ is the corresponding Q-function and it is given by $q_i(s^t, a_i^t) \triangleq E_{\boldsymbol{\pi_{-i}}}[r_i(s^t, \mathbf{a}^t) + \gamma \sum_{s^{t+1}\in \boldsymbol{S}} P(s^{t+1}|s^t, \mathbf{a}^t) v_i(s^{t+1})]$, the function $f_r(\boldsymbol{\pi},\mathbf{v}) = \sum_{i=1}^N [\sum_{s\in S} v_i(s) - E_{a_i\sim\pi_i(.|s)} q_i(s,a_i)]$ is a non-negative quantity for feasible solutions, and equals zero (a global minimum) at a Nash equilibrium solution, and $\mathcal{H}(\boldsymbol{\pi})$ is the γ-discounted

causal entropy for the policies $\boldsymbol{\pi}$ (Bloem & Bambos, 2014). The policy $\boldsymbol{\pi}$ is a NE if and only if $f_r(\boldsymbol{\pi}, \mathbf{v})$ reaches zero while being a feasible solution (Prasad & Bhatnagar, 2015).

7.4.2.2 MA-GAIL Modeling Approach

The imitation learning problem can be interpreted as matching two occupancy measures between each agent's policy and its expert policy. These occupancy measures include the distribution over visited states and selected actions ($\rho_{\boldsymbol{\pi}}$) while navigating the environment with a policy π (Ho & Ermon, 2016; Song, et al., 2018). In the multiagent setting, the connection between the inverse reinforcement learning (MAIRL) and the occupancy measure matching under the NE solution concept is established, where the MAIRL is a dual for the occupancy measure matching, as given by Equation (7.2) (Song, et al., 2018).

$$\textbf{MARL} \circ \textbf{MAIRL}_{\psi}(\boldsymbol{\pi}_E) = \arg\min_{\boldsymbol{\pi}\in\prod} \sum_{i=1}^{N} -\beta\mathcal{H}_i(\pi_i) + \psi_i^*(\rho_{\pi_i,\pi_{E_{-i}}} - \rho_{\boldsymbol{\pi}_E}) \qquad (7.2)$$

where β is a hyperparameter specifying the entropy regularization strength, ψ_i^* is a convex regularization function that measures the occupancy measures' similarity between the expert policy and agent's policy, and $\pi_i, \pi_{E_{-i}}$ denotes a policy π_i for agent i and $\pi_{E_{-i}}$ for the other agents.

The ψ-regularized MAIRL presented in Equation (7.2) shows that the MAIRL seeks for each agent i a policy π_i such that its occupancy measure is closer to the expert policy π_E as measured by ψ_i^*. However, the algorithm assumes no access to the expert policy π_E during the training process. Thus, the term $\rho_{\pi_{i,E_{-i}}}$ in Equation (7.2) cannot be computed. Instead, an alternative

approach for matching the occupancy measure between the expert and agent's policy can be implemented via introducing an adversarial RF regularizer while removing the effect of the entropy regularizers, as given by Equation (7.3) (Song, et al., 2018). Removing the entropy regularizer is favorable in multiagent settings as the causal entropy of the agent policy's depends on the other agents' policies.

$$\arg\min_{\boldsymbol{\pi}} \sum_{i=1}^{N} \psi_i^* \left(\rho_{\pi_i, \pi_{E_{-i}}} - \rho_{\boldsymbol{\pi}_E}\right) = \arg\min_{\boldsymbol{\pi}} \sum_{i=1}^{N} \psi_i^* (\rho_{\pi_i, \pi_{-i}} - \rho_{\boldsymbol{\pi}_E}) \qquad (7.3)$$

Over the MA-GAIL framework, matching the occupancy measures involves formulating a ω −parameterized discriminator (D_{ω_i}) and θ − parameterized generator (π_{θ_i}) for each agent i. The discriminator is a binary classifier, which maps state-action pairs to scores that are optimized to distinguish expert trajectories from the trajectories generated by the agent policy (i.e., generator) π_{θ_i}. The discriminator acts like a RF, providing the learning signal to the generator. The generator objective is to maximize the agent's reward (i.e., perform optimally under the discriminator scores) to fool the discriminator with generated trajectories that are challenging to be distinguished from real trajectories. Thus, matching occupancy measures objective in Equation (7.3) can be rewritten as the optimization objective given by Equation (7.4) (Song, et al., 2018).

$$\min_{\boldsymbol{\theta}} max_{\boldsymbol{\omega}} E_{\boldsymbol{\pi_\theta}}\left[\sum_{i=1}^{N} \log D_{\omega_i}(s, a_i)\right] + E_{\boldsymbol{\pi_E}}\left[\sum_{i=1}^{N} \log\left(1 - D_{\omega_i}(s, a_i)\right)\right] \qquad (7.4)$$

The generators $\boldsymbol{\pi_\theta}$ are updated during the training process via multiagent reinforcement learning, where a baseline value function approximator $V_{\emptyset}^{\pi}$ is utilized to reduce the learned policy variance of each agent. The steps of MA-GAIL training process are described in Algorithm 1 (Song, et al., 2018).

Algorithm 1: MA-GAIL

Input: Expert demonstrations $\boldsymbol{\mathcal{D}_E = \mathcal{T}_j^E}$; MG $(N, \boldsymbol{S}, \boldsymbol{A}, \eta, \boldsymbol{P}, \gamma)$; Initialize the parameters of policies, discriminators, and baseline value estimators with $\boldsymbol{\theta_0}, \boldsymbol{\omega_0}, \boldsymbol{\emptyset_0}$.

repeat

Sample trajectories $\boldsymbol{\mathcal{D}_{\pi_\theta}}$ from agents' generators $\boldsymbol{\pi_\theta}$

Sample the pairs of states and actions $\mathcal{X}_{\boldsymbol{\pi_\theta}}$, $\mathcal{X}_E$ from $\boldsymbol{\mathcal{D}_{\pi_\theta}}$, $\boldsymbol{\mathcal{D}_E}$

for $i = 1, \dots, N$

Update ω_i to increase the discriminators' objective function part in (Eq. 11)

end for

for $i = 1, \dots, N$

Compute value estimate V_i^* and Advantage estimate $A_{\emptyset_i}^{\pi_{\theta_i}}$ for each state-action pairs of $\mathcal{X}_{\pi_{\theta_i}}$

Update value estimator parameter $\emptyset_i$ with:

$$\min E_{\mathcal{X}_{\pi_{\theta_i}}} \left[\left(V_{\emptyset_i}(s, a_{-i}) - V_i^*(s, a_{-i}) \right)^2 \right]$$

Update generator parameter θ_i by policy gradient.

end for

until Convergence

Return: Learned policies $\boldsymbol{\pi_\theta}$ and reward functions $\boldsymbol{D_\omega}$

7.4.3 Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL)

7.4.3.1 LSBRE Solution Concept

The LSBRE solution presents a stochastic generalization of NE solution, where the bounded rationality agents or imperfect agents' behavior can be characterized (Yu, et al., 2019). The compatibility of the LSBRE solution with the Maximum Entropy principle (Ziebart, et al., 2008) in handling non-optimal agents' behavior could lead to a more robust estimation of agents' RFs and strategies. Under the LSBRE solution, the trajectory's probability is exponentially proportional to the rewards along its path. More details about the LSBRE solution concept for MG are provided in chapter 6 of this thesis.

7.4.3.2 MA-AIRL Modeling Approach

Under the LSBRE solution for MG, agent trajectories are characterized via maximizing their likelihood under the LSBRE stationary distribution generated by the ω −parameterized RFs $\{r_i(s, \boldsymbol{a}; \omega_i)\}_{i=1}^{N}$. In this approach, a ω −parameterized discriminator (D_{ω_i}) and θ −parameterized generator (q_{θ_i}) are formalized for each agent i. The discriminators and generators recover agent RFs and approximate their policies at convergence, respectively. The training process of the MA-AIRL algorithm is described in Algorithm 2 (Yu, et al., 2019). More details about this algorithm are provided in chapter 6 of this thesis.

Algorithm 2: MA-AIRL

Input: Expert demonstrations $\boldsymbol{\mathcal{D}}_E = \mathcal{T}_j^E$; MG $(N, \boldsymbol{S}, \boldsymbol{A}, \eta, \boldsymbol{P}, \gamma)$; Initialize the parameters of policies, discriminators, and potential functions with $\boldsymbol{\theta}, \boldsymbol{\omega}, \emptyset$.
repeat
  Sample trajectories $\boldsymbol{\mathcal{D}}_{\boldsymbol{\pi_\theta}}$, from agents' generators $\boldsymbol{\pi_\theta}$.
  Sample the pairs of states and actions $\mathcal{X}_{\boldsymbol{\pi_\theta}}$, $\mathcal{X}_E$ from $\boldsymbol{\mathcal{D}}_{\boldsymbol{\pi_\theta}}$, $\boldsymbol{\mathcal{D}}_E$.
  for $i = 1, \dots, N$
    Update $\omega_i, \emptyset_i$ aiming at increasing the discriminators' objective function given by (Eq. 18).
  end for
  for $i = 1, \dots, N$
    Update the reward estimates $\hat{r}_i(s, a_i, s')$ with $g_{\omega_i}(s, a_i)$.
    Update policy parameters θ_i using $\hat{r}_i(s, a_i, s')$ by policy gradient.
  end for
until Convergence
Return: Learned policies $\boldsymbol{\pi_\theta}$ and reward functions $\boldsymbol{g_\omega}$

7.4.4 Road User Policy Estimation: A Multi-Agent Deep Reinforcement Learning (MA-DRL) Approach

Cyclists' and pedestrians' optimal policies are estimated during the MA-GAIL and MA-AIRL algorithms' training processes using the MA-DRL framework. Road users' $\theta -$ parameterized

generators are optimized using the guidance of the discriminator signals (i.e., rewards) via the Multiagent Actor-Critic with Kronecker factors (MACK) algorithm (Song, et al., 2018). This algorithm is a model-free policy gradient in MA-DRL that utilizes centralized learning with decentralized execution framework. Hence, policies are trained with additional information of all road users' states and actions to reduce the training variance. However, only each road user's states are used to predict its actions at execution. For each agent, two deep neural networks are formalized (i.e., Actor and Critic). More details about the Multiagent Actor-Critic with Kronecker factors (MACK) algorithm are provided in chapter 6 of this thesis. Figure 7.1 illustrates the multiagent modeling framework considering the MA-GAIL and MA-AIRL algorithms with the MACK DRL approach.

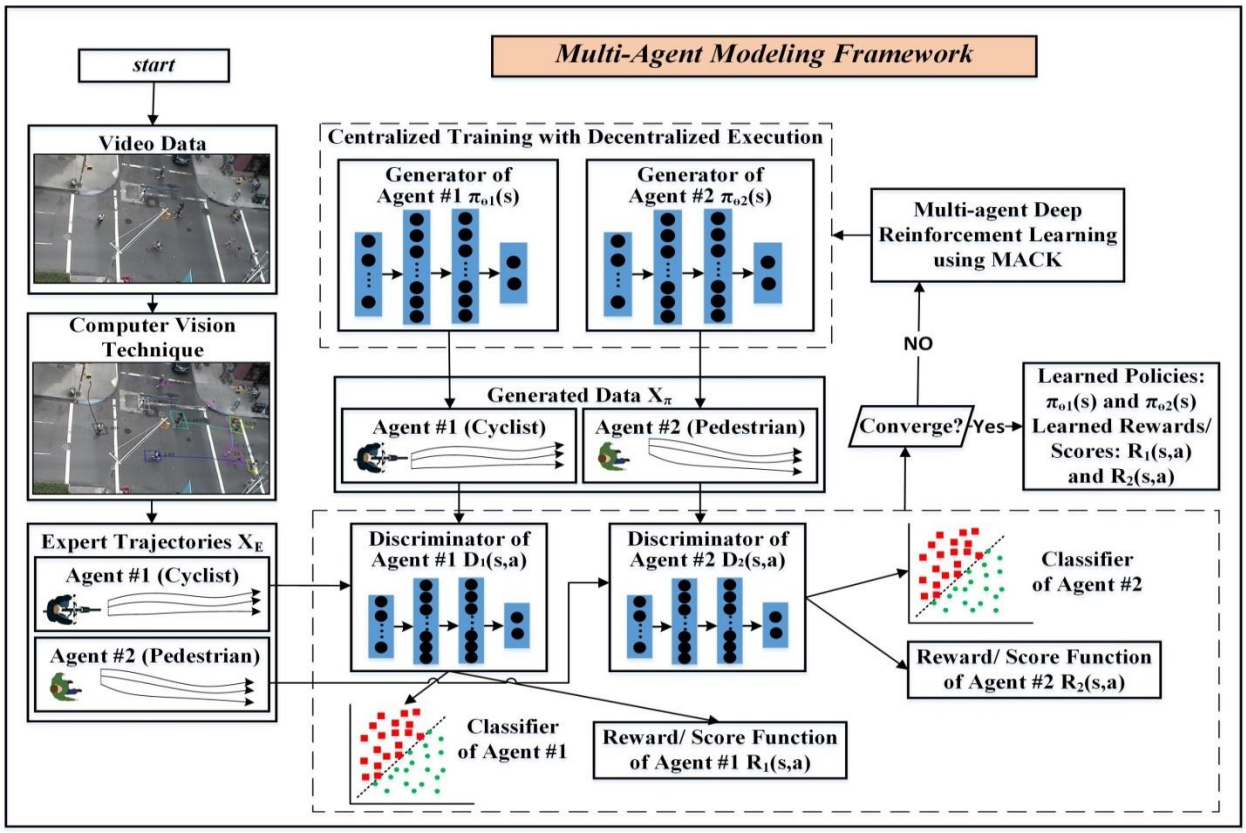


Figure 7.1 Illustration of the multiagent modeling framework for the MA-GAIL and MA-AIRL algorithms

7.5 Trajectory Accuracy Measures

The accuracy of the simulated trajectories is assessed based on the relative distance from the actual trajectories. In this study, two evaluation metrics are used to assess the accuracy of the simulated trajectories, including the *Mean Absolute Error (MAE)* and the *Hausdorff Distance (HausD)*. The methodology for calculating these metrics is presented in detail in chapter 4 of this thesis.

7.6 Multiagent Reward Function Learning Results

This study analyzes and models cyclist and pedestrian behavior in head-on and crossing interactions separately. The majority of these interactions occurred at a relatively low shared space density. In this work, the decentralized multiagent modeling approach is used to model cyclist and pedestrian interactions in shared spaces. In this approach, no prior assumption is made about the correlation between the multiagent RFs (i.e., agents' interaction type of being fully cooperative or competitive). Thus, each agent's reward function is modeled separately, where a discriminator is defined for each agent using its state features. However, agents' discriminators learn with some degree of dependency during the training process as they indirectly interact via the environment. Each road user's states, on which the reward function is based, and actions are defined by a set of variables (as described earlier in section 7.2) that describe its behaviour during the interactions. The road user states and actions distributions, along with their descriptive statistics, are provided in chapter 6 of this thesis.

7.6.1 Multiagent Reward Functions Recovery in Head-on Interaction

The learned adversarial discriminators in the MA-GAIL and MA-AIRL algorithms estimate the multiagent cyclist and pedestrian RFs. The adversarial discriminators captured highly nonlinear and complex multiagent reward functions (MARFs) of road user interaction. Figure 7.2 and Figure 7.3 present the estimated MARFs from applying the MA-GAIL and MA-AIRL algorithms, respectively. The figures present the MARFs in 2-dimensional state feature space by assigning the other features to their mean values. The estimated MARFs can be used to infer road users' preferences and give insights into their competitive and cooperative behavior. Nevertheless, inferring the MARFs could be challenging as a result of the existence of equilibrium solution concepts.

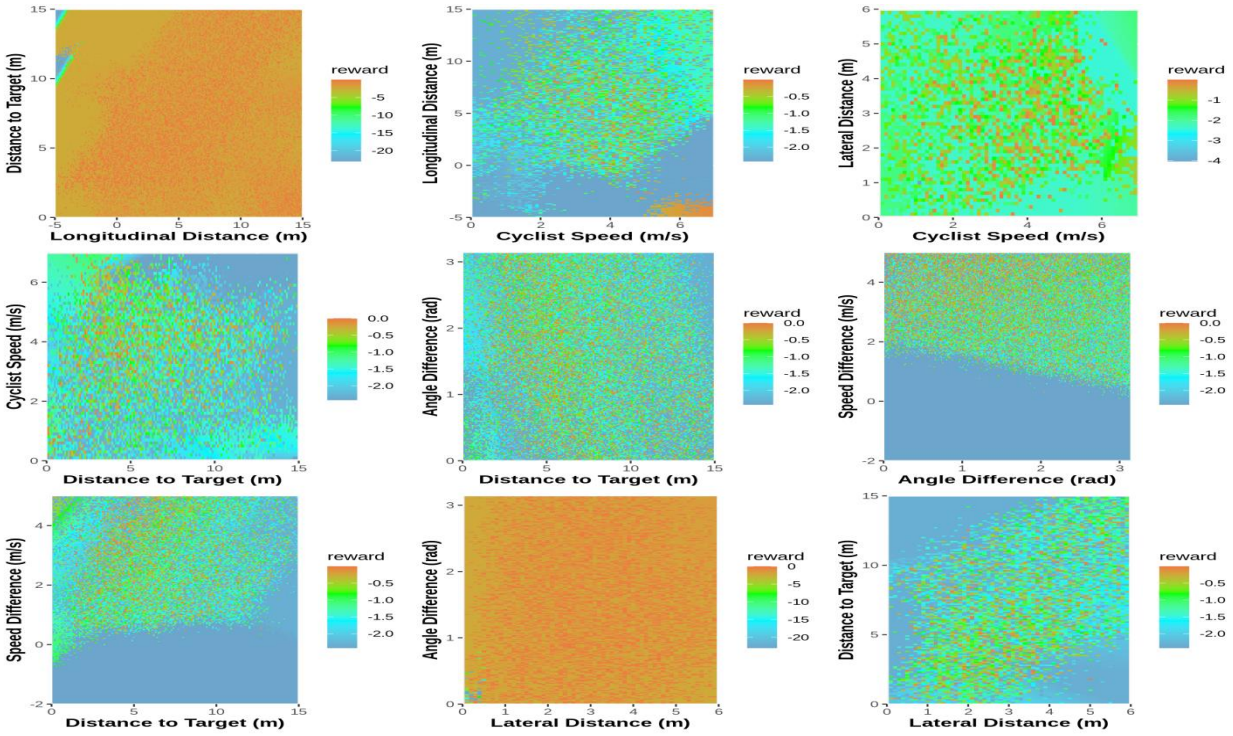
The cyclist and pedestrian MA-GAIL and MA-AIRL RFs, illustrated with the distance to destination and lateral distance features, show higher rewards for the states with smaller distances to the destination than the states with larger distances. This suggests that both road users prefer to reach their destinations. The RFs illustrated with the distance to destination and longitudinal distance features confirm this preference. Moreover, the RFs suggest that both road users have higher preferences for resolving the conflict situation, where the negative longitudinal distance values indicate that the cyclist and pedestrian have crossed each other. These results are consistent with the previous cyclist and pedestrian interaction studies (Alsaleh, et al., 2020; Alsaleh & Sayed, 2021a). It is worth mentioning that the inferred cyclist and pedestrian preferences are associated with other features' mean values. Therefore, the inferred preferences could vary based on the other features' values.

The preference of road users for reaching their destinations and cross each other may show cooperative behavior. However, their tendency to adopt evasive maneuvers (e.g., swerving) to resolve a conflict situation may show competitive behavior. The road users' swerving maneuvers are associated with sudden changes in their movement directions (yaw angles), leading to changes (reductions) in their computed longitudinal distances and deviations from their destinations' shortest paths. This may be associated with a reduction in the sum of gained rewards along each road user path. Previous studies (e.g., (Liu, et al., 2014)) reported a similar result. The equilibrium solution concept in MG (i.e., Nash and LSBRE) can handle such cooperative and competitive behaviors. For example, in both MA-GAIL and MA-AIRL models, the road user RFs illustrated with the distance to destination and longitudinal distance features indicate a wider preference range of longitudinal distance for pedestrians than cyclists. This indicates a higher pedestrians' tendency and flexibility in adopting swerving maneuvers to resolve conflict situations than cyclists.

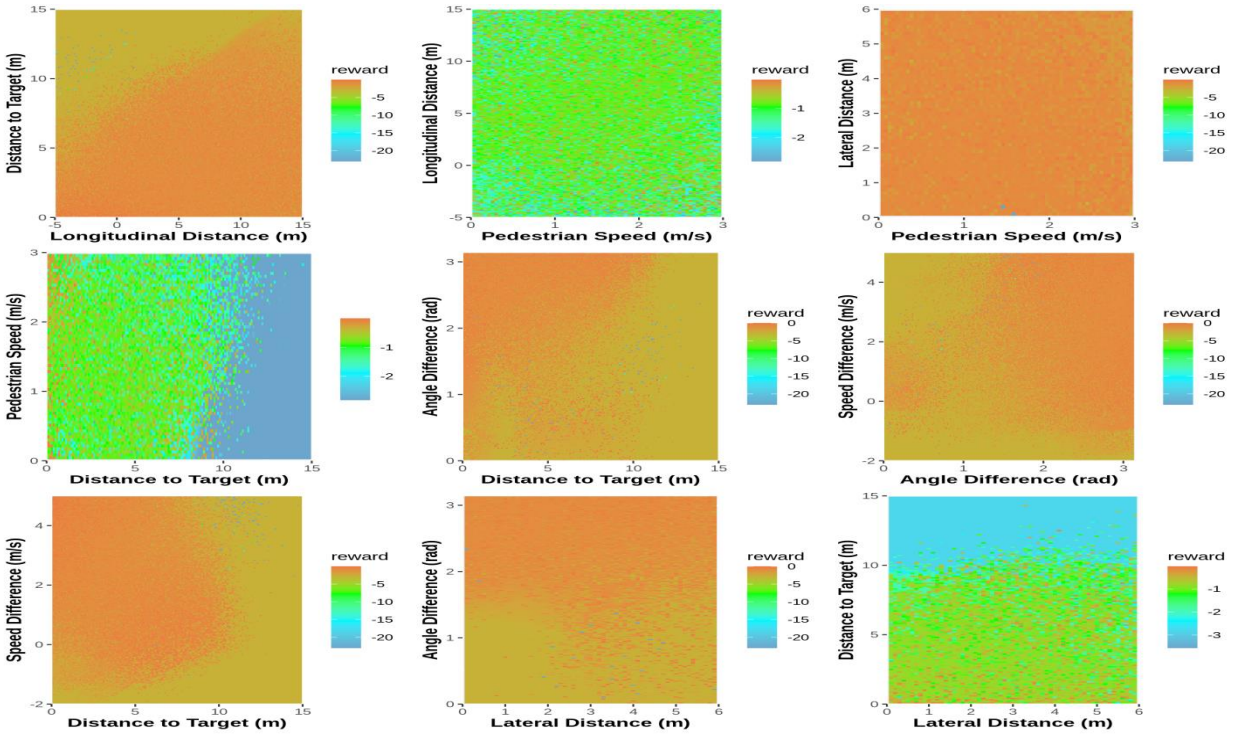
The cyclist and pedestrian MA-GAIL and MA-AIRL RFs illustrated with the speed and lateral distance features indicate that cyclists have higher preferences of increasing their lateral distances at higher cycling speeds. This result is in agreement with previous studies (Luo, et al., 2015; Alsaleh & Sayed, 2021a). The cyclist and pedestrian MA-GAIL and MA-AIRL RFs illustrated with the angle difference and speed difference features show competitive behavior in their preferences. Both the MA-GAIL and MA-AIRL RFs show high cyclist preferences for keeping high speed-differences with pedestrians. However, the MA-GAIL and MA-AIRL RFs show high pedestrian preferences for keeping intermediate and low speed-differences with

cyclists, respectively. This competitive road user behavior can be handled by the MG's equilibrium concept.

The road user MA-GAIL and MA-AIRL RFs illustrated with the speed and longitudinal distance features show that cyclists and pedestrians have high preferences to have lower and slightly lower speeds before resolving conflict situations (i.e., small positive longitudinal distances) comparing to after resolving them, respectively. These results are consistent with the previous studies (Alsaleh, et al., 2020; Hussein & Sayed, 2015). Previous studies showed that cyclists in head-on interactions prefer to reduce their speeds to resolve conflicts with pedestrians. However, swerving maneuvers of pedestrians in head-on interactions are not associated with significant changes in their speeds. Moreover, the RF indicates that cyclists prefer accelerating after resolving conflict situations to reach their preferred speeds. The MA-GAIL RFs show that cyclists and pedestrians prefer high cycling/moving speeds after resolving conflict situations. However, the MA-AIRL RFs show that cyclists and pedestrians prefer intermediate cycling/ moving speeds. The MA-AIRL RF results are in agreement with the observational behavior study of cyclist and pedestrian interactions in shared spaces than the MA-GAIL RFs (Alsaleh, et al., 2020).

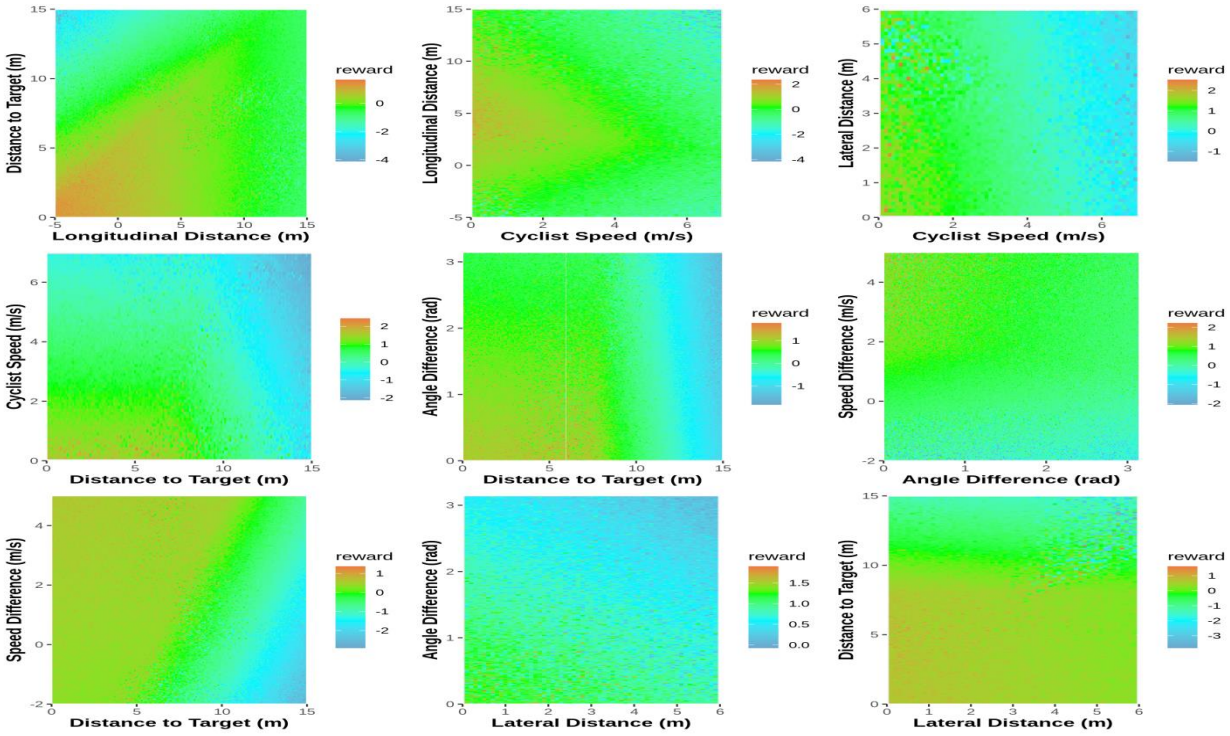


(a) Bike reward function

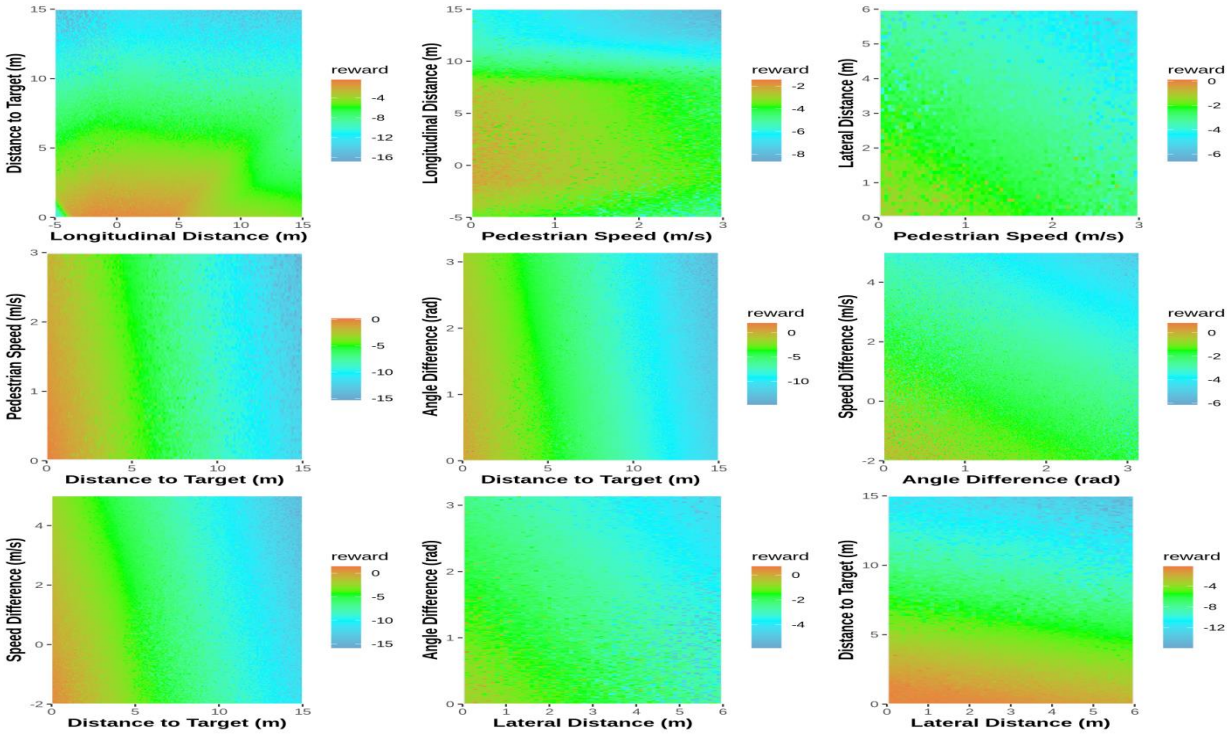


(b) Pedestrian reward function

Figure 7.2 Multiagent reward functions for the head-on interaction using MA-GAIL algorithm



(a) Bike reward function



(b) Pedestrian reward function

Figure 7.3 Multiagent reward functions for the head-on interaction using MA-AIRL algorithm

7.6.2 Multiagent Reward Functions Recovery in Crossing Interaction

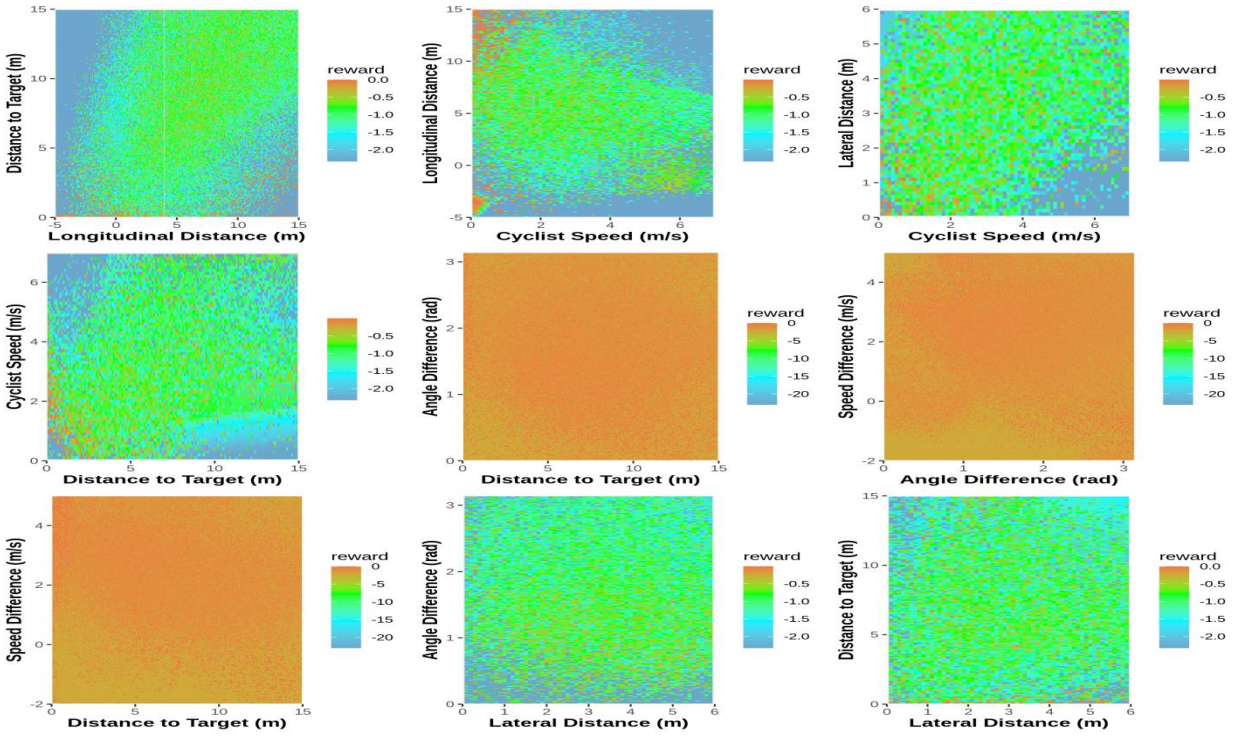
The road user MARFs in the crossing interaction are estimated using the learned adversarial deep neural network discriminators similarly to the head-on interaction. The estimated MARFs from using the MA-GAIL and MA-AIRL algorithms are illustrated in Figure 7.4 and Figure 7.5, respectively. The road user MA-GAIL and MA-AIRL RFs illustrated with the distance to destination and speed features indicate that both road users prefer to reach their destinations. Moreover, the road user MA-GAIL and MA-AIRL RFs illustrated with the lateral distance and angle difference features indicate that they prefer to keep larger lateral distances during their swerving maneuvers (e.g., at angle difference different from 90^{o}). These results are in agreement with the previous studies (Alsaleh, et al., 2020; Alsaleh & Sayed, 2021a).

Moreover, the road user MA-GAIL and MA-AIRL RFs illustrated with the distance to destination and longitudinal distance features show different cyclist and pedestrian longitudinal distance preferences (i.e., crossing and yielding behaviors) under the current crossing interaction situation. The tendency of road users to yield to each other to resolve conflict situations and avoid collisions may suggest cooperative behavior. However, competitive behavior can be observed in their tendency to cross first and avoid severe changes in their movement patterns (e.g., decelerating or swerving). The equilibrium solution concept in MG can handle these cooperative and competitive behaviors. For example, the MA-GAIL RFs, which are based on Nash equilibrium, show higher pedestrian preferences for keeping positive longitudinal distances (i.e., yield for cyclists). However, the cyclist RFs suggest uncertain preferences of crossing first (i.e., keep negative longitudinal distance) or yielding for pedestrians (i.e., positive longitudinal

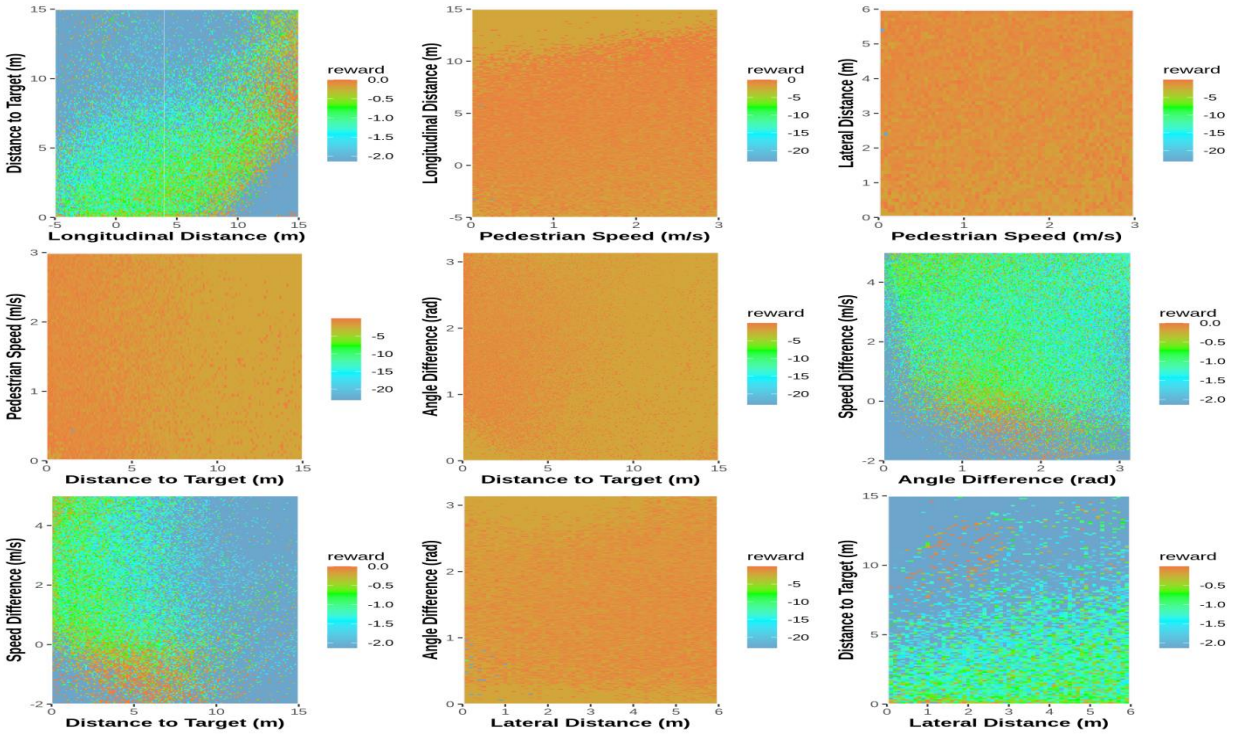
distance). However, the MA-AIRL RFs, which are based on LSBRE, indicate that cyclists highly prefer to keep negative longitudinal distances (i.e., cross first), while pedestrians have high preferences for keeping small positive longitudinal distances (i.e., yield for cyclists).

Similar to the head-on interaction, the road user MA-GAIL and MA-AIRL RFs illustrated with the speed difference and distance to destination features indicate competitive behavior in speed difference preferences. Both of the MA-GAIL and MA-AIRL RFs show that cyclists have higher preferences for high speed-differences with respect to pedestrians. However, pedestrians prefer to have low speed-differences. This competitive road user behavior can be handled by the MG's equilibrium concept. The road user MA-GAIL and MA-AIRL RFs illustrated with the speed and longitudinal distance features show high preferences of both road users for keeping low cycling/ moving speeds when being closer to conflict points, which agrees with the previous study (Alsaleh, et al., 2020).

In general, both of the multiagent models (i.e., MA-GAIL and MA-AIRL) predicted similar road user preferences, which were in agreement with the previous studies. However, the MA-GAIL RFs suggested contradictory preferences for road user speeds around the conflict points and led to higher uncertainty in cyclists crossing behaviour (i.e., crossing first or yield). This is likely attributed to road users' sub-optimal behavior, which is challenging to be accounted for by using the Nash equilibrium solution as it assumes entirely rational and optimal road user behavior. The lack of the MA-GAIL model's ability to handle bounded rationality and sub-optimal road user behavior compared with the MA-AIRL model could influence the model's accuracy.

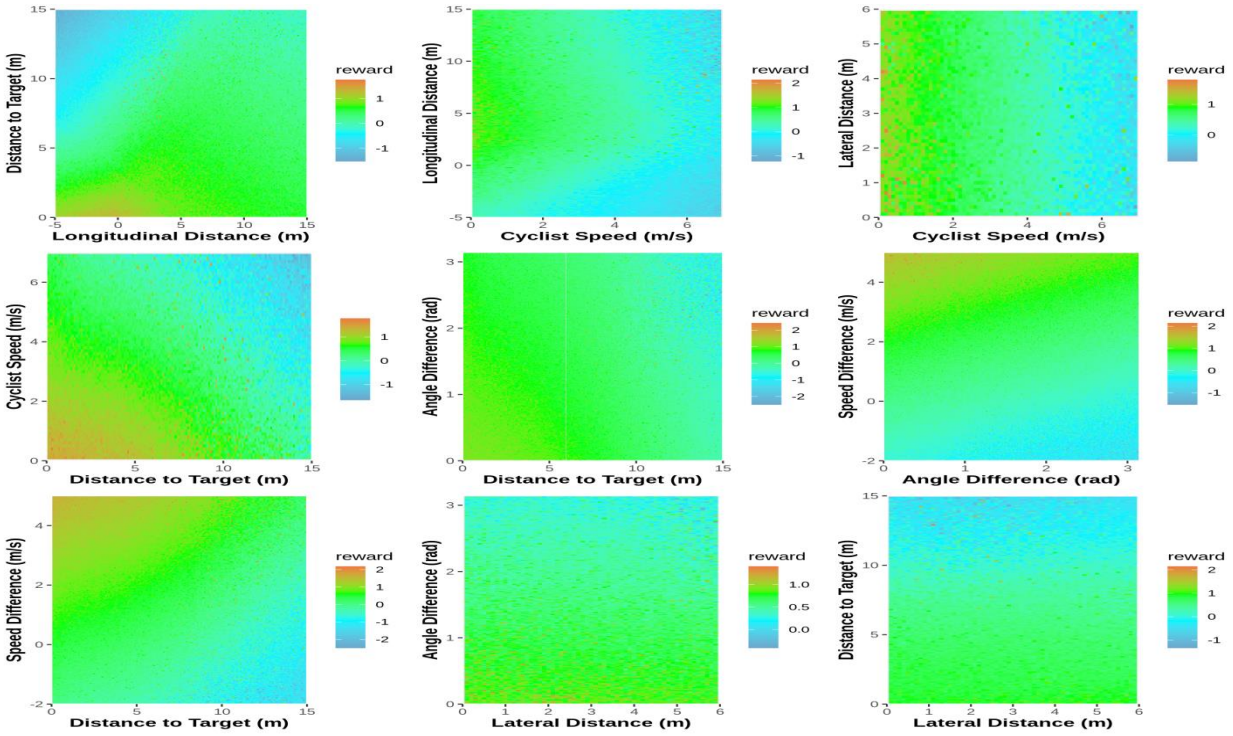


(a) Bike reward function

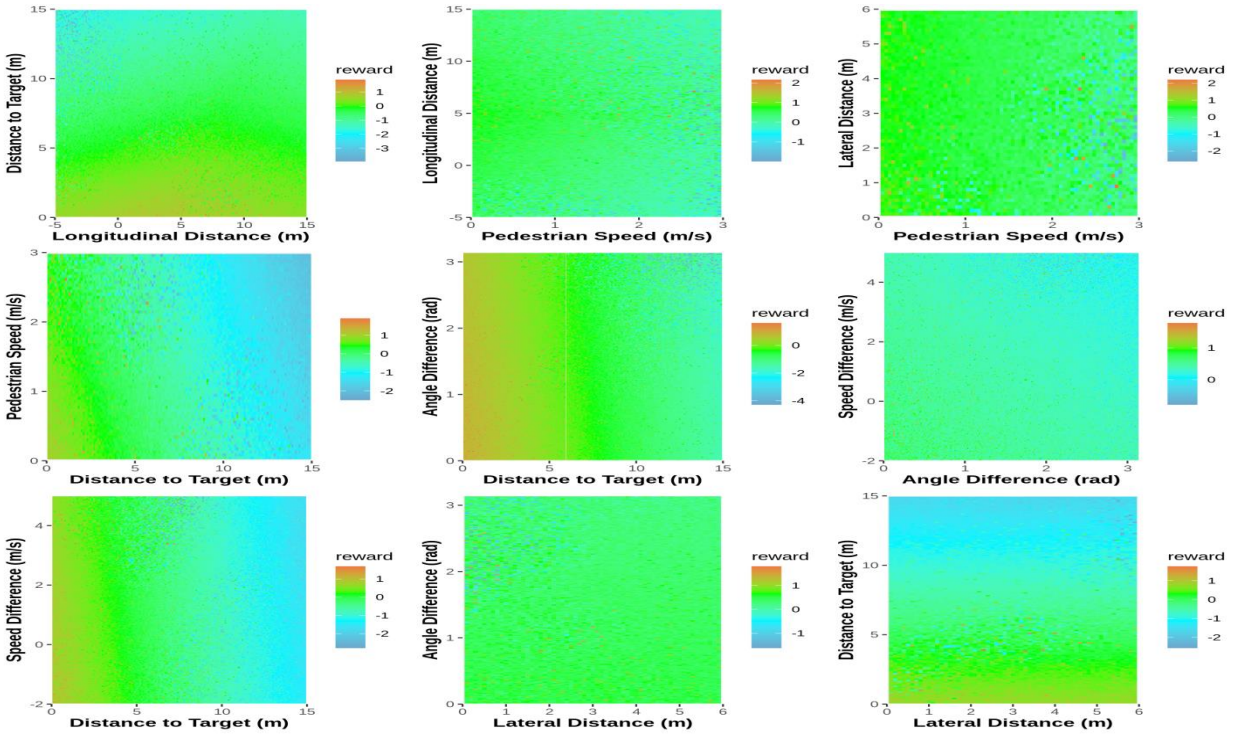


(b) Pedestrian reward function

Figure 7.4 Multiagent reward functions for the crossing interaction using MA-GAIL algorithm



(a) Bike reward function



(b) Pedestrian reward function

Figure 7.5 Multiagent reward functions for the crossing interaction using MA-AIRL algorithm

7.7 Multiagent Models Evaluation and Comparison

The multiagent cyclist and pedestrian strategies (i.e., actions) are learned during the MA-AIRL and MA-GAIL algorithms' training process using the MACK DRL approach. A multiagent simulation platform is developed for shared spaces. The learned policies are incorporated in the developed multiagent microsimulation platform to emulate road user trajectories using the testing dataset. A multiagent virtual environment of shared spaces is created for the simulation platform using OpenAI Gym (Brockman, et al., 2016; Lowe, et al., 2017). The platform runs at a simulating frequency of 30 HZ. Figure 7.6 presents the flowchart of the developed platform. Firstly, the simulation platform initiates the shared space environment with information about road users' initial states (e.g., positions, speeds, yaw angles, and destinations). Then, it calculates cyclist and pedestrian state features (e.g., relative distances) and rewards. Road users then make appropriate decisions simultaneously based on sampling from their optimal learned strategies. Finally, the selected road user decisions are used to update their states.

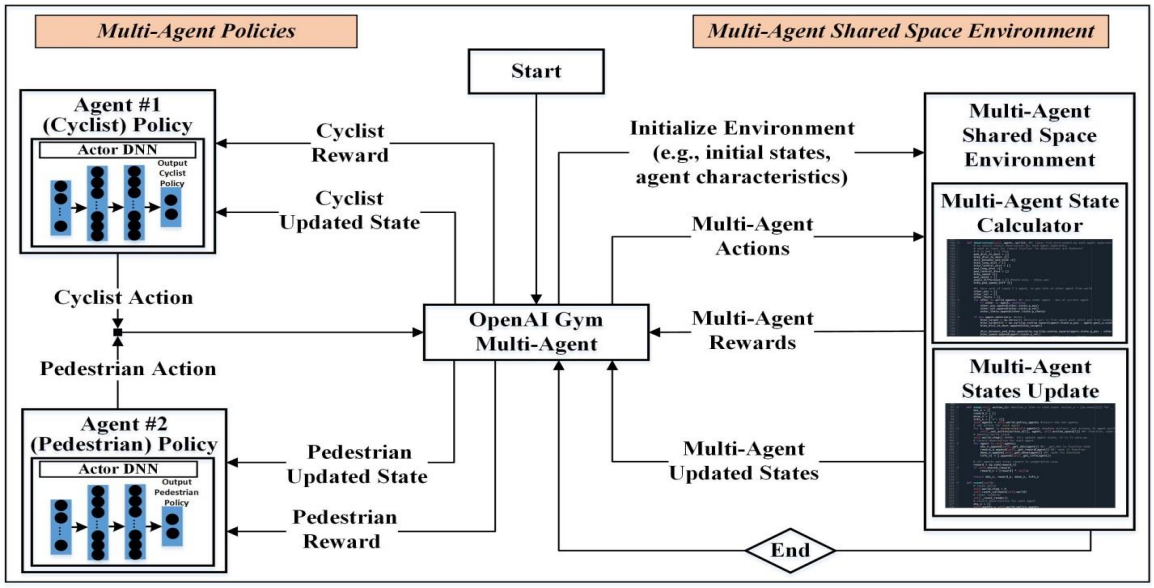


Figure 7.6 Multiagent microsimulation platform flowchart

The developed multiagent simulation platform, based on the learned MA-GAIL and MA-AIRL policies, is used to emulate cyclist and pedestrian trajectories using the testing dataset. The simulation platform is based on the multiagent setting, which enables the realistic emulating of concurrent road users' decision-making processes. The accuracy of the proposed models is evaluated with respect to road user behavioral parameters, as illustrated in Table 7.1. Generally, both models provided more accurate road user behaviour predictions than the single-agent Gaussian Process model (Alsaleh & Sayed, 2021a). Both models predicted road user speeds more accurately than their relative distances (i.e., longitudinal and lateral distances). The LSBRE solution concept in the MA-AIRL model predicted road user behavior (trajectories) more accurately than the MA-GAIL model, which is based on Nash Equilibrium. The use of the LSBRE solution concept in the MA-AIRL model increases the prediction accuracy of the speed, longitudinal and lateral distances in terms of MAE (HuasD) for cyclists and pedestrians in head-on interaction by 13.1% (12.3%), 23.2% (20.3%), 22.5% (15.0%), and 17.7% (13.0%), 11.1% (7.0%), 12.8% (6.7%), respectively. The corresponding prediction improvements in the crossing interaction are 14.2% (20.6%), 32.1% (27.0%), 22.3% (17.3%), and 22.3% (11.9%), 21.3% (20.9%), 19.5% (15.3%) in terms of MAE (HuasD) for cyclists and pedestrians, respectively.

Examples of the predicted cyclist and pedestrian trajectories and their corresponding actual trajectories for head-on and crossing interactions are illustrated in Figure 7.7-Figure 7.8 and Figure 7.9-Figure 7.10, respectively. In general, the figures show that the use of the LSBRE based model (MA-AIRL) predicts road user evasive action mechanisms more accurately than the Nash equilibrium-based model (MA-GAIL). For example, Figure 7.7-Figure 7.8 illustrate an

example of the head-on interaction, where the pedestrian takes an evasive maneuver of swerving and then accelerating to resolve the conflict with the cyclist. The LSBRE-based model predicted similar evasive action mechanisms, as shown in Figure 7.8. However, the Nash equilibrium-based model predicted road user evasive actions less accurately as it suggests that the cyclist mainly swerve and then accelerates to resolve the conflict with the pedestrian, as shown in Figure 7.7. Figure 7.9-Figure 7.10 show an example of the crossing interaction, where the cyclist accelerates to cross first and the pedestrian yields (i.e., pedestrian maintains low walking speed while adopting a swerving maneuver) to avoid the collision. The LSBRE-based model predicted similar evasive maneuver mechanisms, as shown in Figure 7.10. However, the Nash equilibrium-based model predicted road user evasive actions less accurately as it suggests that the cyclist decelerates first, then the pedestrian swerves to avoid the collision, as shown in Figure 7.9.

Interaction Type	Variable	MA-GAIL		MA-AIRL	
		$Avg.\ MAE$	$Avg.\ HausD$	$Avg.\ MAE$	$Avg.\ HausD$
Head-on Interaction	$Bike\ speed\ (m/s)$	0.40	0.72	0.35	0.63
	$Bike\ longitudinal\ distance\ (m)$	0.51	0.82	0.39	0.65
	$Bike\ lateral\ distance\ (m)$	0.57	0.88	0.44	0.75
	$Ped\ speed\ (m/s)$	0.34	0.59	0.28	0.51
	$Ped\ longitudinal\ distance\ (m)$	0.67	0.94	0.59	0.87
	$Ped\ lateral\ distance\ (m)$	0.61	0.90	0.53	0.84
Crossing Interaction	$Bike\ speed\ (m/s)$	0.43	0.77	0.37	0.61
	$Bike\ longitudinal\ distance\ (m)$	0.56	0.96	0.38	0.70
	$Bike\ lateral\ distance\ (m)$	0.51	0.88	0.40	0.73
	$Ped\ speed\ (m/s)$	0.31	0.54	0.24	0.47
	$Ped\ longitudinal\ distance\ (m)$	0.66	1.08	0.52	0.85
	$Ped\ lateral\ distance\ (m)$	0.59	0.94	0.47	0.79

Table 7.1 Microsimulation prediction errors for the multiagent models

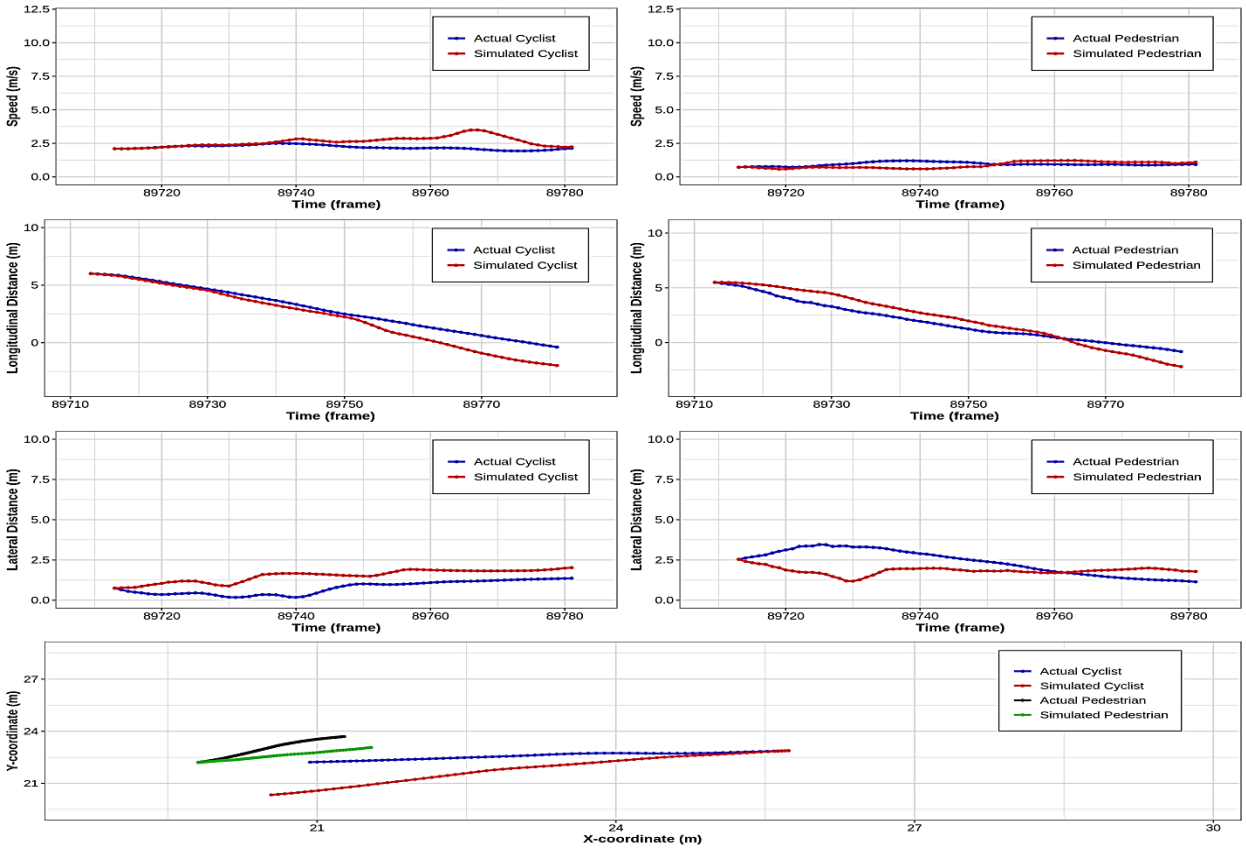


Figure 7.7 MA-GAIL (Nash Equilibrium) model's simulation for the head-on interaction

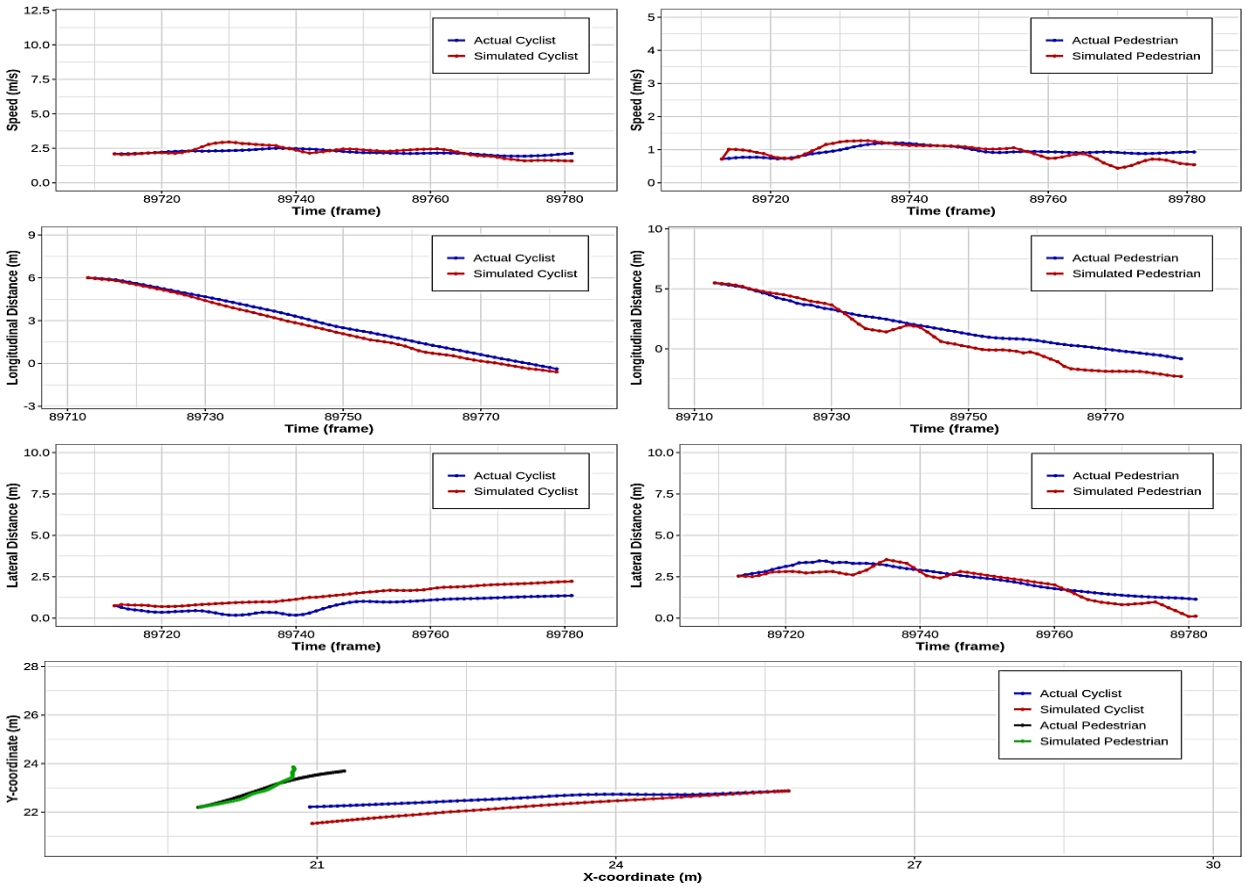


Figure 7.8 MA-AIRL (LSBR Equilibrium) model's simulation for the head-on interaction

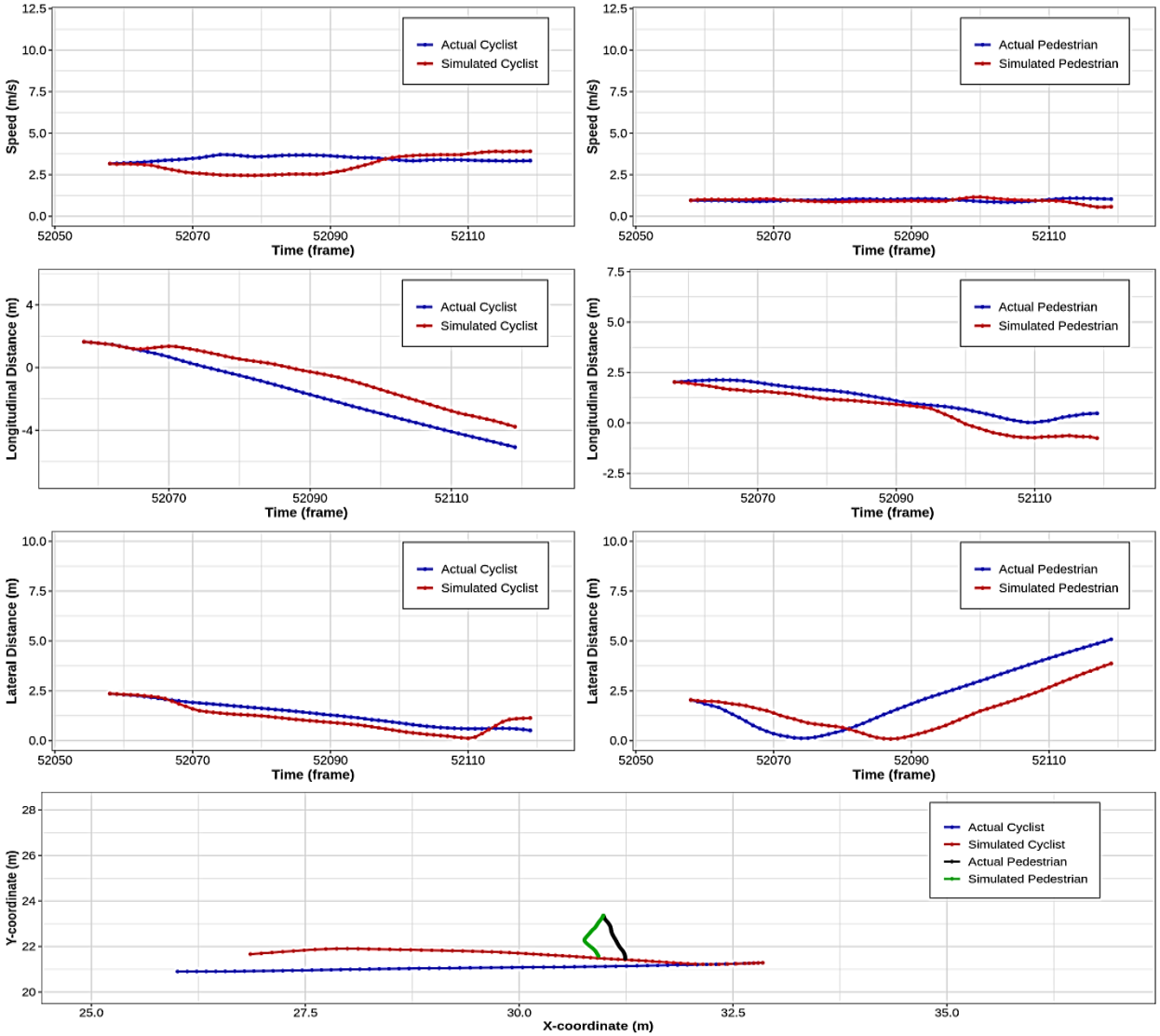


Figure 7.9 MA-GAIL (Nash Equilibrium) model's simulation for the crossing interaction

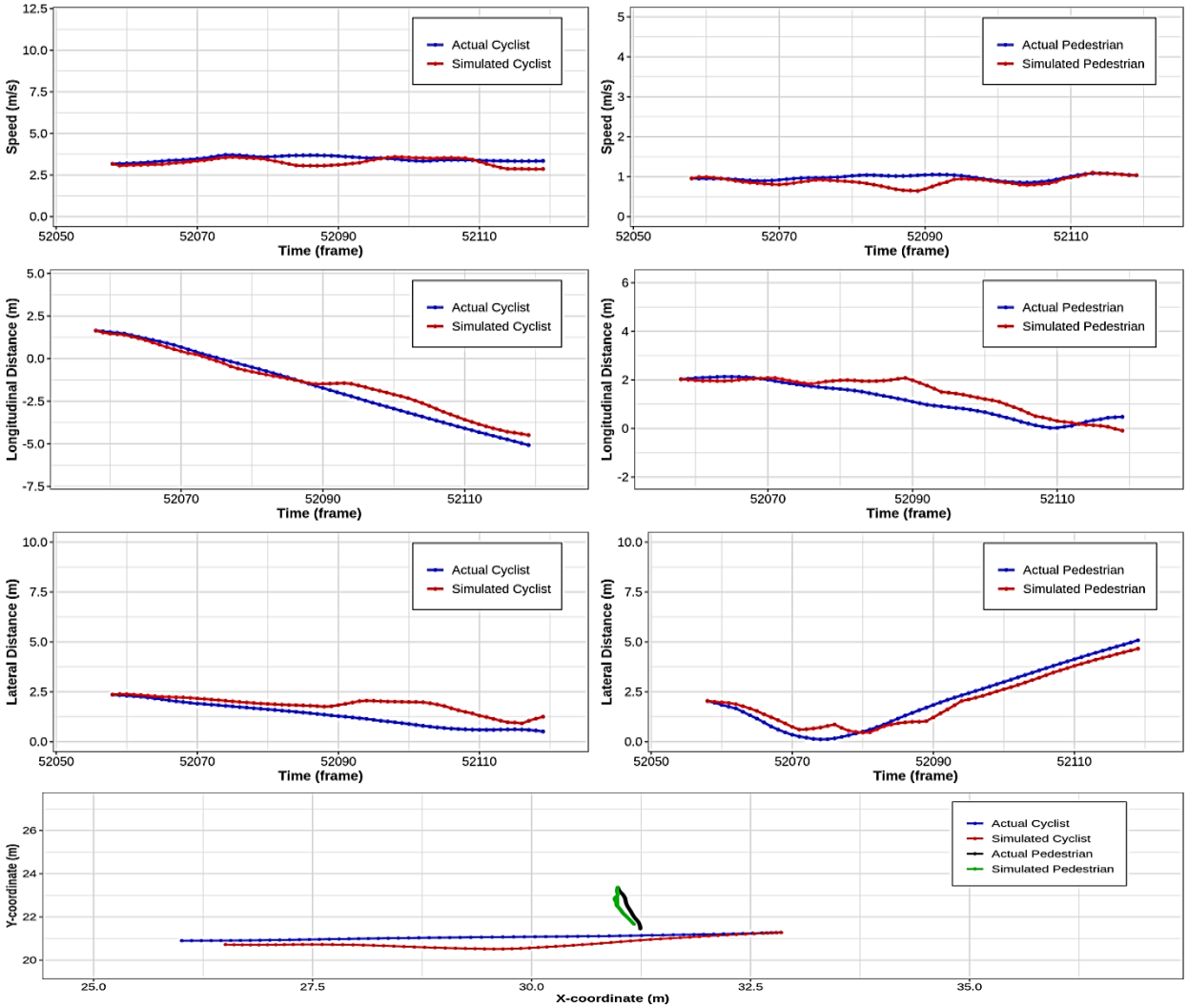


Figure 7.10 MA-AIRL (LSBR Equilibrium) model's simulation for the crossing interaction

7.8 Summary and Conclusion

This chapter presented the details of assessing the performance of different MG's behavioral equilibrium theories in modeling road user interactions in shared spaces. Specifically, this study proposed two novel multiagent approaches for modeling and simulating road users' operational level decisions concurrently in shared spaces. The first approach is the MA-GAIL, which is based on the NE theory for MG. The second approach is the MA-AIRL, which is based on the

LSBRE theory. The first modeling approach assumes entirely rational and optimal road user behavior. However, the second modeling approach relaxes the optimality assumption by assuming sub-optimal and bounded rationality (limited information access) road user behavior. Unlike the traditional game-theory modeling approach, which considers a single time-step payoff modeling, these approaches depend on MG that considers the stochastic interaction nature of road users in shared spaces and their sequential decision processes. This study has important implications for the microscopic multiagent transportation system models that aim at evaluating various transport modes' safety (e.g., cyclists, pedestrians, autonomous vehicles) and transportation facilities' operation and performance levels (e.g., shared spaces).

The contributions of this study are in the development of different equilibrium-based multiagent microsimulation models for road user interactions in shared spaces and specifying a behavior-based consistent paradigm for modeling equilibrium in the multiagent transportation systems of road user interactions in shared spaces. The MA-GAIL microsimulation model, which is based on the NE, is compared to the MA-AIRL model, which is based on the LSBRE. Results show that the LSBRE-based model predicted road users' operational level decisions and evasive action mechanisms more accurately than the NE-based model. Both multi-agent models led to more accurate road user behavior predictions compared to the single-agent Gaussian Process model (Alsaleh & Sayed, 2021a). This can be attributed to the lack of ability of the single-agent model to capture the equilibrium solution concept compared with the multi-agent models. Moreover, the recovered multiagent road user (cyclist and pedestrian) reward functions based on the LSBRE model provided more accurate inferences of their behavior and interaction, particularly around

the conflict points, than the NE-based model. This is likely attributed to the lack of ability of the NE behavioral theory to handle sub-optimal road user behavior.

Chapter 8: Conclusions and Future Research

8.1 Summary and Conclusions

Active modes of transportation, such as cycling and walking, are witnessing an increased interest in many countries worldwide. The promotion of active modes of transportation forms an important strategy for supporting cities' sustainability goals, reducing traffic-induced air pollution, and helping road users to adopt a healthy lifestyle and increase their level of physical activity. Well-designed strategies and facilities for active modes of transportation that promote their comfort and safety would likely encourage road users to increase their physical activities. Shared space facilities present a strategy that involves redesigning specific cities' roads into areas for social and recreational activities areas for active road users. Such facilities improve active transportation modes' mobility and increase their comfort level while reducing the dominance of motor vehicles. Moreover, Ciclovia (Triana et al. 2019) presents another strategy that temporarily repurposes specific cities' roads into motorized-free areas for active transportation modes. These strategies have been implemented in many cities in North America, Europe, Latin America, and Australia. However, despite the increased interest in shared space facilities, most of these facilities were built with little prior operation and safety evaluations (Karndacharuk, et al., 2016; Chong, et al., 2010). This is likely due to the significant difference between road user behaviour in shared space facilities and those in conventional roads and the lack of available tools and models for such evaluation. As such, developing a solid understanding of cyclist and pedestrian behaviours and their interactions in shared space facilities is of great interest.

Although some studies utilized computer simulation of cyclist and pedestrian dynamics as a tool to study their behaviour, existing models suffer from several shortcomings due to the complexity of active road user behaviour systems. For instance, cyclist and pedestrian systems encompass a large degree of heterogeneity and their movement behaviour in shared space facilities involves frequent changes in moving direction and/ or speed due to their interactions with other road users. Unlike conventional roads, the shared space concept allows road users to move freely in the whole area of the facility, rather than being restricted to use predetermined paths. Modeling active road user behaviour requires defining a large set of parameters, which are challenging to be extracted and measured from actual data.

In this thesis, the details of the development of microsimulation models of cyclist and pedestrian interactions in shared space facilities are presented. A detailed behaviour analysis of cyclist and pedestrian interactions in such facilities was firstly carried out. In the behavioural study, video data were collected at the Robson Square shared space in downtown Vancouver. Trajectories of cyclists and pedestrians involved in interactions were extracted using computer vision algorithms. The extracted trajectories were further analyzed to obtain several variables that describe elements of road users' behaviour, including longitudinal and lateral distances, and speed and acceleration profiles. The extracted road user variables were used to analyze their behaviour during the different phases of the interactions. Several interactions between cyclists and pedestrians were analyzed in this study, including interactions with road users moving in the same direction (i.e., following and overtaking interactions), opposing direction (head-on interactions), and crossing interaction. Moreover, the study analyzed the collision avoidance

mechanisms applied by the road users to avoid collisions with other road users during the interactions. The collision avoidance mechanisms involve either changing road user movement direction or walking/cycling speed. The study identifies several parameters that can be used to model the microscopic behaviour of cyclists and pedestrians in the microsimulation models.

Following the detailed analysis of cyclist and pedestrian interactions in shared space facilities, the details of developing agent-based microsimulation models of cyclist and pedestrian interactions were presented. The road users were modeled as utility-based intelligent and rational agents using the Markov Decision Process (MDP) modeling framework. This modeling approach accounts for road users' intelligence and their ability to logically assess the surrounding environment and make rational decisions relying on their utility functions. This is considered an important step in modeling road users' intelligence in microsimulation platforms, as most of the previous modeling frameworks ignored the intelligence of road users. Considering the intelligence of road users in microsimulation models is important, especially in shared space modeling, as they can have different degrees of freedom in accomplishing specific tasks (e.g., overtaking). Simulating platforms were developed to emulate road user interactions in shared space facilities using the R-software (R Core Team, 2018) and Python coding (Python Software Foundation, 2019) using the OpenAI Gym (Brockman, et al., 2016).

The inverse reinforcement learning (IRL) approach was utilized to recover cyclists' reward functions using their expert demonstrations (i.e., examples of their trajectories during the interactions) from two locations of a non-motorized shared space in Robson Square in downtown

Vancouver, British Columbia. The recovered reward functions from applying the IRL algorithms are important for estimating the road users' policies and developing agent-based microsimulation models. In the first phase of the MDP-based models' development, two IRL algorithms, the Maximum Entropy (ME) and the Feature Matching (FM), that recover discrete and linear cyclists' reward functions were implemented. Two types of cyclist-pedestrian interactions were considered in the first phase, the following and overtaking interactions. The recovered reward functions were used to estimate cyclists' optimal policies in their interactions with pedestrians using reinforcement learning. The developed models were validated by comparing the simulated and actual trajectories of the cyclists in the validation dataset using microscopic behaviour parameters such as speed, longitudinal distance, and lateral distances. The difference between the actual and simulated trajectories was evaluated using two error measures; the mean absolute error (MAE) and the Hausdorff distance (*HausD*). Generally, the results showed that the Maximum Entropy (ME) algorithm outperformed the Feature Matching (FM) algorithm in developing MDP models for cyclist-pedestrian interactions in shared space facilities. The ME reward function estimates were more consistent and in line with expectations than the FM estimates. This may be attributed to the ability of the ME IRL algorithm to solve the ambiguity issue in reward function estimation and accounts for imperfect (non-optimal) observed behaviour. Moreover, the ME algorithm produced more accurate estimates of cyclist speed and longitudinal and lateral distances compared to the FM algorithm. Both algorithms predict the cyclist speed more accurately than the cyclist position. For the following interaction, the ME algorithm achieved an average improvement in predicting cyclist speed by about 12.3% (10.6% using *HausD*) compared to the FM algorithm. The corresponding average improvements in the

prediction of the longitudinal and lateral distances are more pronounced and equal to 19.5% (28.7% using *HausD*) and 20.7% (22.9% % using *HausD*). For the overtaking interaction, the ME algorithm shows an average improvement in predicting cyclist speed by about 20.5% (27.9% using *HausD*) compared to the FM algorithm. The corresponding average improvements in the prediction of the longitudinal and lateral distances are 17.3% (21.6% using *HausD*) and 20.6% (17.1%), respectively. However, it was noted that these models led to abrupt changes in the cyclist behavior (e.g., speed, distance, and yaw rate), which are resulted from using linear and discrete reward functions under the MDP modeling framework. These abrupt changes can affect the safety assessment of these interactions as they may lead to more severe traffic conflicts. As such, developing more advanced models that consider the continuous nature of road user movements and the expected nonlinearity in the data, including estimating the road user reward functions and policies, were considered in the next study.

Following the development of the discrete MDP models of cyclist and pedestrian interactions in shared space facilities, the details of developing continuous state and action MDP models were presented. In this approach, continuous IRL algorithms aiming at recovering continuous road users' reward functions were implemented. Unlike the discretization approach for modeling the reward function, the continuous approach models road users' states and actions in continuous spaces. The continuous modeling approach leads to smoother predicted trajectories and avoids the abrupt changes resulted from states/ reward function discretization. The continuous IRL algorithms are implemented to recover the linear and Gaussian Process (GP) stochastic nonlinear reward functions for cyclists in head-on and crossing interactions with pedestrians in three

shared space facilities located in Vancouver, BC, and New York City, New York. The GP nonlinear reward function provides a way for learning the heterogeneity in road users' behaviour by modeling the distribution of their behaviour. The recovered reward functions are used to develop an agent-based microsimulation platform to emulate road users' interactions in shared spaces. The study utilized Deep Reinforcement Learning (DRL) using the synchronous Temporal Difference (TD) advantage Actor-Critic algorithm to estimate cyclists' optimal policies. The accuracy of the simulated trajectories was validated against actual trajectories in the validation dataset. The results show that the recovered GP nonlinear reward functions led to more accurate predictions than the linear reward functions for both the head-on and crossing interactions. The GP and the linear reward functions predicted cyclist speed more accurately compared to the cyclist position. For the head-on interaction, the GP nonlinear reward function improves the prediction accuracy of cyclist speed by about 54.0% (50.7% in terms of *HausD*) compared to the linear reward function. Moreover, the GP nonlinear reward function improves the prediction accuracy of both the longitudinal and lateral distances by about 47.8% and 50.0% (57.4% and 42.4% in terms of *HausD*), respectively. For the crossing interaction, the GP nonlinear reward function improves the prediction accuracy of cyclist speed by about 54.8% (48.4% in terms of *HausD*) compared with the linear reward function. For the longitudinal and lateral distances, the GP reward function achieves average improvement in the prediction accuracy by about 52.5% and 48.2% (57.5% and 42.0% in terms of *HausD*), respectively. The results demonstrated the potential use of the continuous nonlinear MDP framework for modeling cyclist-pedestrian interactions in hared space facilities.

Following the continuous Gaussian Process Inverse Reinforcement Leaning (GPIRL) model (i.e., a single-agent modeling framework), the advanced multi-agent microsimulation model of cyclist and pedestrian interactions in shared space facilities was developed. A novel Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL) model, based on the Logistic Stochastic Best Response Equilibrium (LSBRE) solution concept, was proposed for replicating road user interactions at shared space facilities. Compared to the single-agent modeling framework, the multi-agent (MA) modeling framework presents a more realistic approach for modeling and simulating road user interactions. Moreover, the MA model can handle non-stationary road user environments and capture the equilibrium solution concept between road users (i.e., social negotiations). The used MA modeling approach is based on Markov Games, which considers the stochastic nature of shared space environment and models the sequential decisions (i.e., actions) of road users simultaneously instead of a single time-step payoff modeling like the traditional game-theoretic framework.

The proposed MA-AIRL algorithm was used to model cyclist and pedestrian head-on and crossing interactions in three shared space facilities located in Vancouver, BC, and New York City, New York. The proposed algorithm recovers road users' multi-agent reward functions using adversarial deep neural network discriminators and estimates their optimal policies using Multi-agent Actor-Critic with Kronecker factors (MACK) deep reinforcement learning. A multiagent simulation platform was developed to emulate cyclist and pedestrian interactions in shared space facilities. The multi-agent model's performance is compared to the baseline single-agent Gaussian Process Inverse Reinforcement Leaning (GPIRL) model. Results show that the multi-

agent modeling approach led to a significantly more accurate prediction of road user behaviour compared to the single-agent modeling approach. For example, in the head-on interaction, the multi-agent model improves the prediction accuracy of cyclists' speed, longitudinal and lateral distances by about 32.7% (16%), 53% (38.7%), and 33.3% (21.1%) in terms of *MAE* (*HuasD*), respectively. The correspondent prediction improvement for the crossing interaction are found to be about 33.9% (26.5%), 56.3% (38.6%), and 43.7% (25.5%), respectively. Moreover, the multi-agent model led to a more accurate prediction of the interaction's behavioural aspects (i.e., evasive action mechanisms). The recovered reward functions based on the single-agent modeling approach did not capture the equilibrium solution concept compared to the multi-agent approach. This is likely attributed to the assumption of fixed pedestrian trajectories in the single-agent framework. The existence of non-stationary in road user environments would likely affect the estimation of the single-agent model reward functions and the road user optimal policies. The results demonstrated the superior performance of the Multi-Agent Adversarial Inverse Reinforcement Learning (MA-AIRL) under the Markov-Game (MG) framework for modeling road user interactions in shared space facilities compared to the single-agent GPIRL modeling framework.

Furthermore, this thesis presented the details of assessing the performance of different behavioral equilibrium theories in modeling road user interactions in shared spaces. Specifically, the work presented in this thesis compared the performance of two novel multiagent approaches for modeling and simulating road users' operational level decisions concurrently in shared spaces. The first approach is the Multi-Agent Generative Adversarial Imitation Learning (MA-GAIL),

which is based on the Nash Equilibrium (NE) theory for MG. The second approach is the MA-AIRL, which is based on the LSBRE theory. The first modeling approach assumes entirely rational and optimal road user behavior. However, the second modeling approach relaxes the optimality assumption by assuming sub-optimal and bounded rationality (limited information access) road user behavior. Unlike the traditional game-theory modeling approach, which considers a single time-step payoff modeling, these approaches depend on MG that considers the stochastic nature of road user interactions in shared spaces and their sequential decision processes. The accuracy of the simulated trajectories was validated against actual trajectories in the validation dataset. The results show that the LSBRE-based model predicted road users' operational level decisions and evasive action mechanisms more accurately than the NE-based model. Both multi-agent models led to more accurate road user behavior predictions compared to the single-agent Gaussian Process model (Alsaleh & Sayed, 2021a). This can be attributed to the lack of ability of the single-agent model to capture the equilibrium solution concept compared with the multi-agent models. For example, the use of the LSBRE solution concept in the MA-AIRL model improved the prediction accuracy of the speed, longitudinal and lateral distances in terms of *MAE* (*HuasD*) for cyclists and pedestrians in head-on interaction by 13.1% (12.3%), 23.2% (20.3%), 22.5% (15.0%), and 17.7% (13.0%), 11.1% (7.0%), 12.8% (6.7%), respectively. The corresponding prediction improvements in the crossing interaction are 14.2% (20.6%), 32.1% (27.0%), 22.3% (17.3%), and 22.3% (11.9%), 21.3% (20.9%), 19.5% (15.3%) in terms of *MAE* (*HuasD*) for cyclists and pedestrians, respectively. Moreover, the recovered multiagent road user (cyclist and pedestrian) reward functions based on the LSBRE model provided more accurate inferences of their behavior and interaction, particularly around the conflict points, than

the NE-based model. This is likely attributed to the lack of ability of the NE behavioral theory to handle sub-optimal road user behavior. This study has important implications in specifying a behavior-based consistent paradigm for the equilibrium modeling of road user interactions in shared spaces.

Overall, the results presented in this thesis demonstrated the validity of modeling complex road user behaviour and their interactions as intelligent and rational agents. This modeling approach accounts for road users ability to perceive the surrounding environment, think and take rational decisions relying on their utility functions. Modeling road user behaviour using the different approaches utilized in this thesis has demonstrated that road users better fellow sub-optimal and bounded rationality behaviour than being fully rational. This can be attributed to the differences in road users' environment and interaction perceptions and their demographics and characteristics. Moreover, the multi-agent modeling approach captured road user evasive action maneuvers with higher accuracy than the single-agent approach. Capturing road user evasive action maneuvers during the interactions with other road users is important in analyzing and studying their behaviour in different interactions.

8.2 Study Limitations, Implications, and Future Research

This thesis contributes to a better understanding, analysis, and modeling of active transportation modes in shared space facilities. However, this thesis is not without limitations, and several future research areas can be considered.

The active road user interaction behaviour analysis study was conducted using a dataset from a shared space facility in Vancouver, British Columbia. Thus, future research can consider investigating road user interactions in shared spaces with different layouts using a more comprehensive dataset. Moreover, the recovered road reward functions that reflect road user behavior and policies will likely vary depending on the driving environment and the driving culture. However, the data used for modeling road user interactions is obtained from three shared space locations in North America with very similar location characteristics, driving culture, and road user behaviour. Thus, considering shared spaces and other facility types from different geographical contexts and varying driving cultures (e.g., Europe, Latin America, and Asia) is important. Moreover, investigating the transferability of the developed models to other pedestrian and cyclist environments is important.

The majority of the interactions modeled in this work took place in relatively low shared space densities. Cyclist and pedestrian behaviors are defined based on several parameters, including their relative positions, speeds, and yaw angles. Some road user parameters (e.g., speed) can implicitly account for the shared space density. However, other factors that can affect road users' decisions in shared spaces, such as the neighbor condition (i.e., other pedestrians and cyclists) and shared space density, can be explicitly considered in the model in future work with more comprehensive data set with varying shared space densities.

Moreover, the behavioural study presented in chapter 3 in this thesis utilized manual methods to specify the beginning and end of the different phases of cyclist and pedestrian interactions.

Despite that the results were statistically significant, they are suffering from human subjectivity and long analysis duration. Thus, considering the implementation of clustering algorithms that automatically differentiate between the different interaction phases would significantly reduce the analysis time and provide a more concise approach for analyzing road user interactions.

The discrete linear Markov Decision Process (MDP) model developed to emulate cyclist-pedestrian interactions, which is presented in chapter 4 in this thesis, discretizes the reward function state variables into equal frequency intervals. Thus, optimization of the cut-off values of the discretization can potentially increase the accuracy of the developed models. Moreover, conducting a sensitivity analysis for some parameters that are used in developing the MDP and Markov Game (MG) models, such as the discount factor, is important.

The developed road user interaction models were concerned with cyclist-pedestrian interactions in shared space facilities. Thus, future research can consider other types of road user interactions such as cyclist-vehicle and pedestrian-vehicle interactions in mixed traffic conditions. Moreover, applying the proposed MG modeling technique to other multi-agent systems like autonomous vehicles and their interactions with active road users would be an interesting future area. Moreover, safety applications such as evaluating traffic conflicts and predicting various conflict indicators, including the time-to-collision (TTC) and Post-Encroachment Time (PET), can also be an interesting future research area.

Furthermore, the work presented in this thesis evaluated the performance of two equilibrium solution concepts, the Nash Equilibrium (NE) and the Logistic Stochastic Best Response Equilibrium (LSBRE), in MG for modeling road user interactions in shared spaces. Thus, future research can evaluate and compare the performance of other equilibrium solution concepts such as the Correlated Equilibrium (CE) (Ziebart, et al., 2011; Aumann, 1974).

The developed road user interaction microsimulation models have several safety and performance evaluation implications. The developed microsimulation models can be used to analyze the effectiveness of safety treatments (e.g., road schemes, markings, signs) of modifying regular road design schemes into shared space facilities, which are expected to modify road user behaviour. Such safety evaluation can be carried out by calculating the traffic conflict indicators from the developed microsimulation models. Moreover, statistical models that relate traffic conflicts with collision occurrence and severity can be developed. The proposed modeling approaches in this thesis can be implemented in autonomous vehicle interactions with road users. This would enable early safety and operation evaluation of the shared spaces incorporating autonomous vehicles. Furthermore, the multi-agent modeling approach proposed in this thesis can be used to design driving policies for autonomous vehicles and bicycles.

The utilized modeling approaches recover road user reward functions, which help achieve the generality and transferability of the developed models. The reward function represents the most compacted representations of road user preferences, objectives, and goals. Thus, road user policies can be re-optimized in new environments relying on the recovered reward functions. The

reward function can predict road user preferences in unseen situations, while policy training approaches are challenging to handle unseen situations and require intensive data for training. Moreover, transferring the developed models to different social and physical space contexts can be carried out using policy adaptation and transfer learning methodologies (Da Silva & Costa, 2019; Li, et al., 2019). In these approaches, the recovered reward functions can be considered a prior, and using a small dataset from the other context; the policy can be adapted. This would likely prevent the overfitting in the updated reward function using a small dataset. Moreover, the Successor Feature Transfer Learning can be utilized to transfer the interdependencies (sub-tasks) between the learned tasks and the related tasks.

Bibliography

AASHTO, 2011. *A policy on geometric design of highways and streets.* Washington, DC: AASHTO.

Abbeel, P. & Ng, A. Y., 2004. *Apprenticeship learning via inverse reinforcement learning.* Banff, Alberta, Canada, In the twenty-first international conference on Machine learning, pp. 1-8.

Adams, J., 1995. *Risk.* London: Routledge.

Adams, J., 2010. Management of the risks of transport. In: *Handbook of risk theory: Epistemology, decision theory, ethics, and social implications of risk.* Dordrecht: Springer, p. 239–264.

Akkar, M., 2005. The changing 'publicness' of contemporary public spaces: a case study of the Grey's Monument Area, Newcastle upon Tyne. *Urban Design International*, 10(2), pp. 95-113.

Alsaleh, R., Hussein, M. & Sayed, T., 2020. Microscopic behavioural analysis of cyclist and pedestrian interactions in shared spaces. *Canadian Journal of Civil Engineering*, 47(1), pp. 50-62.

Alsaleh, R. & Sayed, T., 2020. Modeling pedestrian-cyclist interactions in shared space using inverse reinforcement learning. *Transportation Research Part F: Traffic Psychology and Behaviour*, Volume 70, pp. 37-57.

Alsaleh, R. & Sayed, T., 2021a. Microscopic Modeling of Cyclists Interactions with Pedestrians in Shared Spaces: A Gaussian Process Inverse Reinforcement Learning Approach. *Transportmetrica A: Transport Science.*

Alsaleh, R. & Sayed, T., 2021b. Markov-game modeling of cyclist-pedestrian interactions in shared spaces: A multi-agent adversarial inverse reinforcement learning approach. *Transportation Research Part C: Emerging Technologies,* Volume 128, p. 103191.

Amini, R. E., Dhamaniya, A. & Antoniou, C., 2021. Towards a Game Theoretic Approach to Model Pedestrian Road Crossings. *Transportation Research Procedia,* Volume 52, pp. 692-699.

Amodei, D. et al., 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565.*

Anvari, B., Bell, M. G., Sivakumar, A. & Ochieng, W. Y., 2015. Modelling shared space users via rule-based social force model. *Transportation Research Part C: Emerging Technologies,* Volume 51, pp. 83-103.

Atkins , 2012. *Shared Use Operational Review,* London: Department of Transport.

Aumann, R. J., 1974. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics,* 1(1), pp. 67-96.

Ayres, G., Wilson, B. & LeBlanc, J., 2004. Method for identifying vehicle movements for analysis of field operational test data. *Transportation research record,* 1886(1), pp. 92-100.

Beitel, D., Stipancic, J., Manaugh, K. & Miranda-Moreno, L., 2018. Assessing safety of shared space using cyclist-pedestrian interactions and automated video conflict analysis. *Transportation Research Part D: Transport and Environment,* Volume 65, pp. 710-724.

Bellman, R., 1957. A Markovian decision process. *Journal of mathematics and mechanics,* 6(5), pp. 679-684.

Bellman, R. E., 1957. *Dynamic Programming.* New Jersey: Princeton University Press.

Bernardi, S. & Rupi, F., 2015. An analysis of bicycle travel speed and disturbances on off-street and on-street facilities. *Transportation Research Procedia,* Volume 5, pp. 82-94.

Bloem, M. & Bambos, N., 2014. *Infinite time horizon maximum causal entropy inverse reinforcement learning.* Los Angeles, California, In 53rd IEEE Conference on Decision and Control, pp. 4911-4916.

Blue, V. J. & Adler, J. L., 1998. Emergent fundamental pedestrian flows from cellular automata microsimulation. *Transportation Research Record,* 1644(1), pp. 29-36.

Blue, V. J. & Adler, J. L., 1999. Cellular automata microsimulation of bidirectional pedestrian flows. *Transportation Research Record,* 1678(1), pp. 135-141.

Bratko, I., Urbancic, T. & Sammut, C., 1995. *Behavioural cloning: phenomena, results and problems.* Berlin, Germany, International Federation of Automatic Control IFAC, pp. 143-149.

Brockman, G. et al., 2016. Openai gym. *arXiv preprint arXiv,* Volume 1606.01540.

Burstedde, C., Klauck, K., Schadschneider, A. & Zittartz, J., 2001. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications,* 295(3-4), pp. 507-525.

Campbell, R. & Wittgens, M., 2004. *The business case for active transportation: The Economic Benefits of Walking and Cycling,* Gloucester: Go for Green.

Chong, S. et al., 2010. Relative injury severity among vulnerable non-motorised road users: comparative analysis of injury arising from bicycle–motor vehicle and bicycle–pedestrian collisions. *Accident Analysis & Prevention,* 42(1), pp. 290-296.

City of Vancouver, 2012. *Transportation 2040,* Vancouver: City of Vancouver.

City of Vancouver, 2015. *Cycling safety study,* Vancouver: City of Vancouver.

Da Silva, F. L. & Costa, A. H. R., 2019. A survey on transfer learning for multiagent reinforcement learning systems. *Journal of Artificial Intelligence Research,* Volume 64, pp. 645-703.

Daamen, W., Hoogendoorn, S., Campanella, M. & and Versluis, D., 2014. Interaction behavior between individual pedestrians. In: U. Weidmann, U. Kirsch & M. Schreckenberg, eds. *Pedestrian and Evacuation Dynamics 2012.* Switzerland: Springer, pp. 1305-1313.

Dai, J., Li, X. & Liu, L., 2013. Simulation of pedestrian counter flow through bottlenecks by using an agent-based model. *Physica A: Statistical Mechanics and its Applications,* 392(9), pp. 2202-2211.

Department of Transport and Main Roads of Queensland State, 2014. *Speed management on shared paths,* Queensland: Queensland goverment.

Department of Transport, Planning and Local Infrastructure, 2012. *Bendigo town centre: Creating shared space to improve pedestrian safety,* Melbourne: state goverment of Victoria.

Dias, C., Iryo-Asano, M., Nishiuchi, H. & Todoroki, T., 2018. Calibrating a social force based model for simulating personal mobility vehicles and pedestrian mixed traffic. *Simulation Modelling Practice and Theory,* Volume 87, pp. 395-411.

Elvik, R., 2014. A review of game-theoretic models of road user behaviour. *Accident Analysis & Prevention,* Volume 62, pp. 388-396.

Figliozzi, M., Wheeler, N. & Monsere, C. M., 2013. Methodology for estimating bicyclist acceleration and speed distributions at intersections. *Transportation research record,* 2387(1), pp. 66-75.

Fujii, H., Uchida, H. & Yoshimura, S., 2017. Agent-based simulation framework for mixed traffic of cars, pedestrians and trams. *Transportation research part C: emerging technologies,* Volume 85, pp. 234-248.

Fu, J., Luo, K. & Levine, S., 2017. Learning robust rewards with adversarial inverse reinforcement learning. *arXiv preprint arXiv:1710.11248.*.

Gavriilidou, A., Daamen, W., Yuan, Y. & Hoogendoorn, S. P., 2019. Modelling cyclist queue formation using a two-layer framework for operational cycling behaviour. *Transportation Research Part C: Emerging Technologies,* Volume 105, pp. 468-484.

Georgila, K., Nelson, C. & Traum, D., 2014. *Single-agent vs. multi-agent techniques for concurrent reinforcement learning of negotiation dialogue policies.* Baltimore, Maryland, In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 500-510.

Gipps, P. G., 1981. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological,* 15(2), pp. 105-111.

Gorrini, A., Crociani, L., Vizzari, G. & Bandini, S., 2018. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation research part F: traffic psychology and behaviour,* Volume 59, pp. 269-285.

Government of South Australia, 2012. *Streets for people: Compendium for South Australian practice,* Adelaide: Government of South Australia.

Graw, M. & König, H., 2002. Fatal pedestrian–bicycle collisions. *Forensic Science International,* 126(3), pp. 241-247.

Guo, N. et al., 2020. Modeling the interactions of pedestrians and cyclists in mixed flow conditions in uni-and bidirectional flows on a shared pedestrian-cycle road. *Transportation research part B: methodological,* Volume 139, pp. 259-284.

Hamilton-Baillie, B., 2008. Shared space: Reconciling people, places and traffic. *Built environment,* 34(2), pp. 161-181.

Hamilton-Baillie, B., 2008. Towards shared space. *Urban Design International,* 13(2), pp. 130-138.

Harsanyi, J. C., 1968. Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points. *Management Science,* 14(5), pp. 320-334.

Harsanyi, J. C. & Selten, R., 1988. *A general theory of equilibrium selection in games.* Cambridge, Massachusetts: MIT Press.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika,* 57(1), p. 97–109.

Hatfield, J. & Prabhakharan, P., 2016. An investigation of behaviour and attitudes relevant to the user safety of pedestrian/cyclist shared paths. *Transportation research part F: traffic psychology and behaviour,* Volume 40, pp. 35-47.

Helbing, D. & Molnar, P., 1995. Social force model for pedestrian dynamics. *Physical review E,* 51(5), pp. 4282-4286.

Highway Capacity Manual, 1985. *Special Report 209: Highway Capacity Manual,* Washington, D.C.: Transportation Research Board, National Research Council.

Hipp, J. A., Bird, A., van Bakergem, M. & Yarnall, E., 2017. Moving targets: Promoting physical activity in public spaces via open streets in the US. *Preventive medicine,* Volume 103, pp. S15-S20.

Ho, J. & Ermon, S., 2016. *Generative adversarial imitation learning.* Barcelona, Spain, In Advances in neural information processing systems, pp. 4565-4573.

Huang, L. et al., 2016. Cyclist social force model at unsignalized intersections with heterogeneous traffic. *IEEE Transactions on Industrial Informatics,* 13(2), pp. 782-792.

Hu, J. & Wellman, M. P., 1998. *Multiagent reinforcement learning: theoretical framework and an algorithm.* Madison, Wisconsin, USA, International Conference on Machine Learning, pp. 242-250.

Hussein, M., Popescu, B., Sayed, T. & Kim, L., 2016. Analysis of road user behavior and safety during New York City's summer streets program. *Transportation research record,* 2586(1), pp. 120-130.

Hussein, M. & Sayed, T., 2015. Microscopic pedestrian interaction behavior analysis using gait parameters. *Transportation Research Record,* 2519(1), pp. 28-38.

Hussein, M. & Sayed, T., 2017. A bi-directional agent-based pedestrian microscopic model. *Transportmetrica A: Transport Science,* 13(4), pp. 326-355.

Hussein, M. & Sayed, T., 2018. Validation of an agent-based microscopic pedestrian simulation model at the pedestrian walkway of Brooklyn Bridge. *Transportation Research Record: Journal of the Transportation Research Board,* 2672(35), pp. 33-45.

Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. A., 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis & Machine Intelligence,* Volume 9, pp. 850-863.

Ismail, K., Sayed, T. & Saunier, N., 2010. Automated analysis of pedestrian–vehicle conflicts: Context for before-and-after studies. *Transportation research record,* 2198(1), pp. 52-64.

Ismail, K., Sayed, T. & Saunier, N., 2013. A methodology for precise camera calibration for data collection applications in urban traffic scenes. *Canadian Journal of Civil Engineering,* 40(1), pp. 57-67.

Jacobs, G., Aeron-Thomas, A. & Astrop, A., 2000. *Estimating global road fatalities,* Crowthorne: Transport Research Laboratory.

Jan't Hoen, P. et al., 2005. *An overview of cooperative and competitive multiagent learning.* Berlin, Heidelberg, Springer, pp. 1-46.

Jennings, N. R., 2000. On agent-based software engineering. *Artificial intelligence,* 117(2), pp. 277-296.

Jiang, R., Jia, B. & Wu, Q. S., 2004. Stochastic multi-value cellular automata models for bicycle flow. *Journal of Physics A: Mathematical and General,* 37(6), pp. 2063-2072.

Jin, S. et al., 2015. An improved multi-value cellular automata model for heterogeneous bicycle traffic flow. *Physics Letters A,* 379(39), pp. 2409-2416.

Kang, L., Xiong, Y. & Mannering, F. L., 2013. Statistical analysis of pedestrian perceptions of sidewalk level of service in the presence of bicycles. *Transportation Research Part A: Policy and Practice,* Volume 53, pp. 10-21.

Kaparias, I. et al., 2013. Analysis of pedestrian–vehicle traffic conflicts in street designs with elements of shared space. *Transportation research record*, 2393(1), pp. 21-30.

Kaparias, I. et al., 2012. Analysing the perceptions of pedestrians and drivers to shared space. *Transportation research part F: traffic psychology and behaviour*, 15(3), pp. 297-310.

Kaparias, I., Hirani, J., Bell, M. G. & Mount, B., 2016. Pedestrian gap acceptance behavior in street designs with elements of shared space. *Transportation research record,* 2586(1), pp. 17-27.

Karndacharuk, A., Wilson, D. J. & Dunn, R., 2014. A review of the evolution of shared (street) space concepts in urban environments. *Transport reviews,* 34(2), pp. 190-220.

Karndacharuk, A., Wilson, D. J. & Dunn, R. C., 2013. Analysis of pedestrian performance in shared-space environments. *Transportation research record,* 2393(1), pp. 1-11.

Karndacharuk, A., Wilson, D. J. & Dunn, R. C., 2014. Safety performance study of shared pedestrian and vehicle space in New Zealand. *Transportation Research Record,* 2464(1), pp. 1-10.

Karndacharuk, A., Wilson, D. J. & Dunn, R. C., 2016. Qualitative evaluation study of urban shared spaces in New Zealand. *Transportation Research Part D: Transport and Environment,* Volume 42, pp. 119-134.

Karndacharuk, A., Wilson, D. J. & Tse, M., 2011. *Shared space performance evaluation: quantitative analysis of pre-implementation data.* Auckland, New Zealand, s.n.

Kiyota, M., Vandebona, U., Katafuchi, N. & Inoue, S., 2000. *Bicycle and pedestrian traffic conflicts on shared pavements.* Munich, In Fourteenth velo-city international conference proceedings.

Lawson, A. R., Pakrashi, V., Ghosh, B. & Szeto, W. Y., 2013. Perception of safety of cyclists in Dublin City. *Accident Analysis & Prevention,* Volume 50, pp. 499-511.

Levine, S., 2019. *Inverse Reinforcement Learning.* [Online] Available at: http://rail.eecs.berkeley.edu/deeprlcourse-fa19/static/slides/lec-15.pdf [Accessed 28 Nov. 2019].

Levine, S. & Koltun, V., 2012. Continuous inverse optimal control with locally optimal examples. *arXiv preprint arXiv:1206.4617.*.

Levine, S., Popovic, Z. & Koltun, V., 2011. *Nonlinear inverse reinforcement learning with gaussian processes.* Granada, Spain, Advances in Neural Information Processing Systems 24 Conference, pp. 19-27.

Levinson, D., 2005. Micro-foundations of congestion and pricing: A game theory perspective. *Transportation Research Part A: Policy and Practice,* 39(7-9), pp. 691-704.

Li, A. C., Florensa, C., Clavera, I. & Abbeel, P., 2019. Sub-policy adaptation for hierarchical reinforcement learning. *arXiv preprint arXiv:1906.05862.*

Liang, X., Baohua, M. A. O. & Qi, X. U., 2012. Psychological-physical force model for bicycle dynamics. *Journal of Transportation Systems Engineering and Information Technology,* 12(2), pp. 91-97.

Lin, X., Beling, P. & Cogill, R., 2014. Comparison of Multi-agent and Single-agent Inverse Learning on a Simulated Soccer Example. *arXiv preprint arXiv:1403.6822.*

Littman, M. L., 1994. *Markov games as a framework for multi-agent reinforcement learning.* s.l., Morgan Kaufmann, pp. 157-163.

Liu, D. C. & Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. *Mathematical programming,* 45(1-3), pp. 503-528.

Liu, M., Zeng, W., Chen, P. & Wu, X., 2017. A microscopic simulation model for pedestrian-pedestrian and pedestrian-vehicle interactions at crosswalks. *PLoS one,* 12(7), p. e0180992.

Liu, Q., Sun, J., Tian, Y. & Xiong, L., 2020. Modeling and simulation of overtaking events by heterogeneous non-motorized vehicles on shared roadway segments. *Simulation Modelling Practice and Theory,* Volume 103, p. 102072.

Liu, S., Lo, S., Ma, J. & Wang, W., 2014. An agent-based microscopic pedestrian flow simulation model for pedestrian traffic problems. *IEEE Transactions on Intelligent Transportation Systems,* 15(3), pp. 992-1001.

Lowe, R. et al., 2017. *Multi-agent actor-critic for mixed cooperative-competitive environments.* Long Beach, California, Advances in neural information processing systems, pp. 6379-6390.

Lucas, B. D. & Kanade, T., 1981. *An Iterative Image Registration Technique with an Application to Stereo Vision.* Vancouver, BC, International Joint Conference on Artificial Intelligence, pp. 674-679.

Luo, Y. et al., 2015. Modeling the interactions between car and bicycle in heterogeneous traffic. *Journal of advanced transportation,* 49(1), pp. 29-47.

Lu, R. et al., 2020. Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management. *Applied Energy,* Volume 276, p. 115473.

Mataric, M. J., Sukhatme, G. S. & Ostergaard, E. H., 2003. Multi-robot task allocation in uncertain environments. *Autonomous Robots,* 14(2), pp. 255-263.

Ma, X. & Luo, D., 2016. Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data. *Transportation research part F: traffic psychology and behaviour*, Volume 40, pp. 130-144.

McKelvey, R. D. & Palfrey, T. R., 1995. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1), pp. 6-38.

McKelvey, R. D. & Palfrey, T. R., 1998. Quantal response equilibria for extensive form games. *Experimental economics*, 1(1), pp. 9-41.

Michon, J. A., 1985. *A critical view of driver behavior models: what do we know, what should we do?*. Boston, MA: Springer.

Mnih, V. et al., 2016. *Asynchronous methods for deep reinforcement learning.* New York, United States, International conference on machine learning, pp. 1928-1937.

Mohammed, H., Sayed, T. & Bigazzi, A., 2021. Microscopic modeling of cyclists on off-street paths: A stochastic imitation learning approach. *Transportmetrica A: Transport Science*, pp. 1-39.

Montufar, J., Arango, J., Porter, M. & Nakagawa, S., 2007. Pedestrians' normal walking speed and speed when crossing a street. *Transportation Research Record*, 2002(1), pp. 90-97.

Nagel, K. & Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. *Journal de physique I*, 2(12), pp. 2221-2229.

Nash, J., 1951. Non-cooperative games. *Annals of mathematics*, 54(2), pp. 286-295.

Ng, A. Y. & Russell, S. J., 2000. *Algorithms for inverse reinforcement learning.* Stanford, CA, USA, In the International Conference on Machine Learning, pp. 663-670.

Nowé, A., Vrancx, P. & De Hauwere, Y. M., 2012. Game theory and multi-agent reinforcement learning. In: *Reinforcement Learning.* Berlin, Heidelberg: Springer, pp. 441-470.

OECD, 2014. *The cost of air pollution: Health impacts of road transport.* Paris, France: OECD Publishing.

Papadimitriou, E., Yannis, G. & Golias, J., 2009. A critical assessment of pedestrian behaviour models. *Transportation research part F: traffic psychology and behaviour,* 12(3), pp. 242-255.

Peng, P. et al., 2017. Multiagent bidirectionally-coordinated nets for learning to play starcraft combat games. *arXiv preprint arXiv:1703.10069 2.*

Prasad, H. L. & Bhatnagar, S., 2015. A study of gradient descent schemes for general-sum stochastic games. *arXiv preprint arXiv:1507.00093.*

Python Software Foundation, 2019. *Python Language Reference, version 3.7.* [Online] Available at: http://www.python.org

R Core Team, 2018. *A language and environment for statistical computing,* Vienna, Austria: R Foundation for Statistical Computing.

Rabl, A. & De Nazelle, A., 2012. Benefits of shift from car to active transport. *Transport policy,* 19(1), pp. 121-131.

Rahmati, Y. & Talebpour, A., 2018. Learning-based game theoretical framework for modeling pedestrian motion. *Physical Review E,* 98(3), p. 032312.

Rockafellar, R. T. & Wets, R. J. B., 2009. *Variational analysis.* 3rd ed. s.l.:Springer Science & Business Media.

Rudloff, C., Schönauer, R. & Fellendorf, M., 2013. Comparing calibrated shared space simulation model with real-life data. *Transportation research record*, 2390(1), pp. 44-52.

Saunier, N. & Sayed, T., 2006. *A feature-based tracking algorithm for vehicles in intersections.* Quebec, Canada, The 3rd IEEE Canadian Conference on Computer and Robot Vision, pp. 59-59.

Savitzky, A. & Golay, M. J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry,* 36(8), pp. 1627-1639.

Schönauer, R. et al., 2012. Modeling concepts for mixed traffic: Steps toward a microscopic simulation tool for shared space zones. *Transportation research record,* 2316(1), pp. 114-121.

Schonauer, R., Stubenschrott, M., Schrom-Feiertag, H. & Mensik, K., 2012. *Social and Spatial Behavior in Shared Spaces.* Schwechat, Austria, 17th International Conference on Urban Planning and Regional Development in the Information Society, pp. 759-767.

Shahhoseini, Z. & Sarvi, M., 2019. Pedestrian crowd flows in shared spaces: Investigating the impact of geometry based on micro and macro scale measures. *Transportation research part B: methodological,* Volume 122, pp. 57-87.

Shiwakoti, N., Sarvi, M. & Rose, G., 2008. *Modelling pedestrian behaviour under emergency conditions–State-of-the-art and future directions.* Gold Coast, Australia, In 31st Australasian Transport Research Forum (ATRF), pp. 457-473.

Shu, S. et al., 2016. Air quality impacts of a CicLAvia event in Downtown Los Angeles, CA. *Environmental Pollution,* Volume 208, pp. 170-176.

Song, J., Ren, H., Sadigh, D. & Ermon, S., 2018. *Multi-agent generative adversarial imitation learning.* Montreal, Canada, In Advances in neural information processing systems., pp. 7461-7472.

Sutton, R. S. & Barto, A. G., 1998. *Introduction to reinforcement learning.* 1st ed. London, England: MIT press.

Swinburne, G., 2005. *Report on Road Safety in Kensington High Street,* London: Royal Borough of Kensington and Chelsea.

Tageldin, A. & Sayed, T., 2019. Models to evaluate the severity of pedestrian-vehicle conflicts in five cities. *Transportmetrica A: transport science,* 15(2), pp. 354-375.

Taherifar, N., Hamedmoghadam, H., Sree, S. & Saberi, M., 2019. A macroscopic approach for calibration and validation of a modified social force model for bidirectional pedestrian streams. *Transportmetrica A: Transport Science,* 15(2), pp. 1637-1661.

Talebpour, A., Mahmassani, H. S. & Hamdar, S. H., 2015. Modeling lane-changing behavior in a connected environment: A game theory approach. *Transportation Research Procedia,* Volume 7, pp. 420-440.

Teknomo, K., 2006. Application of microscopic pedestrian simulation model. *Transportation Research Part F: Traffic Psychology and Behaviour,* 9(1), pp. 15-27.

Tierney, L. & Kadane, J. B., 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association,* 81(393), pp. 82-86.

Tomasi, C. & Kanade, T., 1991. *Detection and Tracking of Point Features,* Pennsylvania, USA: Carnegie Mellon University.

von Stackelberg, H., 1952. *The Theory of the Market Economy.* s.l.:Oxford University Press.

Wang, W. L., Lo, S. M., Liu, S. B. & Kuang, H., 2014. Microscopic modeling of pedestrian movement behavior: Interacting with visual attractors in the environment. *Transportation Research Part C: Emerging Technologies,* Volume 44, pp. 21-33.

Wiedemann, R., 1974. *Simulation des straßenverkehrsflusses, Schriftenreihe Heft 8,* Karlsruhe, Germany: Institute for Transportation Science.

Williams, C. K. & Rasmussen, C. E., 2006. *Gaussian processes for machine learning.* Cambridge, MA: MIT press.

Wooldridge, M., 1997. Agent-based software engineering. *IEE Proceedings-software,* 144(1), pp. 26-37.

World Health Organization, 2004. *World Report on Road Traffic Injury Prevention,* Geneva: World Health Organization.

World Health Organization, 2009. *Global health risks : mortality and burden of disease attributable to selected major risks,* France: World Health Organization.

World Health Organization, 2018. *Global status report on road safety 2018,* France: World Health Organization.

Wu, Y. et al., 2017. *Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation.* Long Beaah, CA, Advances in neural information processing systems, pp. 5279-5288.

Xue, S. et al., 2017. An improved Burgers cellular automaton model for bicycle flow. *Physica A: Statistical Mechanics and its Applications,* Volume 487, pp. 164-177.

Yeung, C. S., Poon, A. S. & Wu, F. F., 1999. Game theoretical multi-agent modelling of coalition formation for multilateral trades. *IEEE Transactions on Power Systems,* 14(3), pp. 929-934.

Yu, L., Song, J. & Ermon, S., 2019. Multi-Agent Adversarial Inverse Reinforcement Learning. *arXiv preprint arXiv:1907.13220v1*.

Zacharias, J., 1999. The Amsterdam experiment in mixing pedestrians, trams and bicycles. *ITE journal,* Volume 69, pp. 22-29.

Zamboni, S. et al., 2020. Pedestrian Trajectory Prediction with Convolutional Neural Networks. *arXiv preprint arXiv,* p. 2010.05796.

Zhang, P. et al., 2020. Potential field cellular automata model for overcrowded pedestrian flow. *Transportmetrica A: transport science,* 16(3), pp. 749-775.

Zhang, Y., Chai, Z. & Lykotrafitis, G., 2020. Deep reinforcement learning with a particle dynamics environment applied to emergency evacuation of a room with obstacles. *arXiv preprint arXiv:2012.00065.*

Zhao, D. et al., 2013. Modeling of passing events in mixed bicycle traffic with cellular automata. *Transportation research record,* 2387(1), pp. 26-34.

Zhou, Z., Zhou, Y., Pu, Z. & Xu, Y., 2019. Simulation of pedestrian behavior during the flashing green signal using a modified social force model. *Transportmetrica A: transport science,* 15(2), pp. 1019-1040.

Ziebart, B. D., Bagnell, J. A. & Dey, A. K., 2011. *Maximum causal entropy correlated equilibria for Markov games.* Taipei, Taiwan, International Conference On Autonomous Agents and Multi-Agent Systems, pp. 207-214.

Ziebart, B. D., Maas, A. L., Bagnell, J. A. & Dey, A. K., 2008. *Maximum Entropy Inverse Reinforcement Learning.* Chicago, Illinois, In the Twenty-Third AAAI Conference on Artificial Intelligence, pp. 1433-1438.