

SOCIAL NAVIGATION FOR A QUADRUPED MOBILE ROBOT

SEBASTIAN ÆGIDIUS
B.Eng(Hons)

A Thesis submitted in fulfilment of requirements for the degree of
Master of Science
Applied Machine Learning
of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College London
September 1, 2022

Abstract

Social navigation for mobile robots at its core is navigation through a human populated space ensuring the social comfort of interacting pedestrian while following as efficient a path as possible. As with any human robot interaction, social navigation is a multi-faceted task. The field of social navigation is a popular research area with a plethora of approaches to specific elements of social navigation, however the current state of the art is still far from a generalized solution that enables a mobile robot to walk alongside pedestrians.

This project successfully implements the social force model for pedestrian dynamics onto a quadruped mobile robot system using visual perception for human pose detection. By understanding the key elements in social navigation and evaluating the social force model implementation, a novel augmented social force model is designed to heighten the mobile robot's ability to perform more effective and efficient social navigation based on the drawbacks of the original social force model. The model augmentations deal with repulsive force interpretation, avoidance behavior, and enables the robot to follow a person given certain conditions. A thorough evaluation of both model implementations is conducted over four test scenarios designed to test the most important sides of navigating a pedestrian walkway – behavior towards oncoming pedestrians, crowds, and blocked pathways. The results shows that the augmented social force model successfully outperforms the original model.

Acknowledgment

I want to thank my supervisor Prof. Yiannis Demiris who let me be part of his laboratory over the course of this project and gave me the opportunity to use the exotic robot equipment handled in this project.

I also want to thank Rodrigo Chacon Quesada for acting as my day to day supervisor always ready to help whenever I asked and providing excellent advice whenever I struggled with challenges whether they were theoretical or practical robot problems.

My thanks also goes to the rest of Phd students and Postdoctoral researchers at the personal robotics laboratory for their help and advice whenever I needed it.

I would also like to extend my gratitude to Jia Du who played a crucial role in the experiments conducted in this project and for being a huge support to me through this project.

Lastly, but not least I wish to express my gratitude towards my sister and parents for their constant support and understanding throughout this project and also during my time as an MSc student at Imperial college London.

Contents

Abstract	3
Acknowledgment	5
Contents	7
List of Figures	11
List of Tables	13
Abbreviations	15
Chapter 1. Introduction	17
1.1 Background	17
1.2 Project Motivation	18
1.3 Project Aim	19
1.3.1 Assumptions	19
1.4 Contributions	21
1.5 Implementation Platform and Hardware	22
Chapter 2. Literature Review	25
2.1 Challenges of Social Navigation	25
2.2 Path Planning for Social Navigation	27
2.3 Human Behavior in Social Navigation	30
2.3.1 Personal Space	31
2.3.2 Display of Intention	32
2.3.3 Position and Formation in Social Spaces	32
2.4 Evaluating Social Navigation Models	34
2.4.1 Metrics	34
2.4.2 Evaluation Methods	37
Chapter 3. Hardware, System Design, and Implementation	39

3.1	Mobile Robot	39
3.2	Visual Perception	41
3.2.1	Sensor Choice	41
3.2.2	Human Detection	42
3.2.3	Camera Placement on the Robot	44
3.3	Computing Unit	45
3.4	Implementation	46
Chapter 4.	The Social Navigation Model	49
4.1	Choice of Model	49
4.2	The Social Force Model for Pedestrian Dynamics	50
4.3	Social Force Model Implementation	52
Chapter 5.	Model Augmentation	57
5.1	Social Force Augmentations	58
5.1.1	Decreasing prediction error and model complexity	58
5.1.2	Side-stepping	60
5.1.3	Follow Mode and Crowd Interpretation	61
5.1.4	Redefining Direction Priority	63
5.1.5	Human Friendly Velocity Control	64
5.2	Augmented Social Force Model Implementation	65
5.2.1	Logging Pedestrian Information	65
Chapter 6.	Experiments and Results	69
6.1	Evaluation of Model Behavior	69
6.1.1	Test Scenario 1	70
6.1.2	Test Scenario 2	74
6.1.3	Test Scenario 3	77
6.1.4	Test Scenario 4	78
6.1.5	Scenario 5	80
6.1.6	Hyper Parameter Tuning	82
6.2	Discussion	85
6.2.1	Limitations	85
6.2.2	Comments On Scenario Evaluation	85
6.2.3	Comfort	86
6.2.4	Social Force Model in Real Life	87
Chapter 7.	Conclusion and Future work	89
7.1	Conclusion	89
7.2	Future Work	90

Bibliography	93
Appendix A. Access to Project Resources	105
A.1 Access to model code	105
A.2 Project Video Demonstration	105

List of Figures

1.1	Social navigation for a mobile robot	20
3.1	The Mobile Robot System	40
3.2	The Human Pose Detection Performed by the Zed2i	43
3.3	3D Printed Camera Mount	44
3.4	System Execution Flowchart	46
4.1	Path planning approach for model implementation.	55
5.1	ASFm change in force field	59
5.2	Repulsive force measured from a moving pedestrian	60
5.3	ASFm follow mode conditions	64
5.4	the maximum forward velocity allowed depending on how far away the closest pedestrian is.	65
6.1	Test scenario 1.	70
6.2	Scenario 1 results.	72
6.3	Test scenario 2	73
6.4	Time to goal, SFM vs. ASFm	76
6.5	Test scenario 3	77
6.6	Test scenario 4	79
6.7	Time to goal, SFM vs. ASFm	81
6.8	The radial force field visualized	83
6.9	The elliptic force field visualized	84

List of Tables

2.1 Metrics for evaluating a robot’s social navigation ability	35
5.1 Matrix format of the log that the model keeps for every currently detected pedestrian.	66

Abbreviations

- NN:** Neural Networks
- HRI:** Human-Robot Interaction
- POV:** Point Of View
- FOV:** Field OF View
- IRL:** Inverse Reinforcement Learning
- GPU:** Graphical Processing Unit
- SSH:** Secure Shell
- ROS:** Robot Operating System
- TOF:** Time Of Flight
- SFM:** Social Force Model
- ASFM:** Augmented Social Force Model
- RGB-D:** Red, Green, Blue and Distance
- CCA:** Cooperative Collision Avoidance
- MOMDP:** Mixed Observability Markov Decision Processes
- POMDP:** Partially Observable Markov Decision Processes
- GAN:** Generative Adversarial Networks
- LSTM:** Long Short-Term Memory
- ADE:** Average Displacement Error
- FDE:** Final Displacement Error
- CPU:** Central Processing Unit
- IMU:** Inertial Measurement Unit
- SDK:** Software Development Kit

Chapter 1

Introduction

1.1 Background

THE task of walking through a crowd of people whether it be down the street or through a shopping mall without colliding into another person requires an incredible amount of information processing, and yet it is almost autonomous for humans. Strategies to tackle and navigate dynamic environments require more than path planning alone. An environment change such as a partial path obstruction can often be solved by object avoidance without much regard to what the obstruction is. But when it comes to domains including humans, altering and continuous environment change must be accounted for. Human-aware navigation or social navigation is the crossing of robot navigation with human-robot interaction (HRI). Social navigation is navigation through a space influenced and structured by the activities of humans within the given space. A major task in social navigation is predicting human behavior. Visual input is one of the most prominent sensory inputs humans have and rely on for navigation in a social setting. Although vision is the most dominant sense, sound and touch also compliment visual cues to perceive environmental changes. Visual input through a camera is one of the most replicated and applied sensory input source for robotics that has been prompted significantly more since the emergence of neural networks and growing computational power.

1.2 Project Motivation

Pre-existing robots and robot models that can perform social navigation are mainly developed for interacting with humans rather than only navigating through a crowd. Robots that can accomplish tasks like being interactive tour guides [1], [2] or information assistants [3] are wheeled robots with large and heavy-built bodies or of a size that does not adapt well in a social setting, making them stiff and slow. More agile robots like holonomic and nimble quadruped robots are advantageous over large wheeled robots in an urban setting with pedestrians. Wheeled and autonomously driven robots are more suitable for road environments as wheels can achieve a much higher speed than legged robots.

For social robots, being able to interpret the surrounding environment that is constantly changing through a first person point of view (POV) like humans present several challenges. In indoor and certain outside areas, it is possible to set up bird-eye-view cameras displaying an overview capable of predicting upcoming changes in the crowd compared to first person POV cameras that mimic the human eye. Another motivation to develop first person POV camera carrying robot systems over static mounted POV cameras is their ability to follow the robot into unknown territories.

A robot similar in size to human pedestrians that is able to traverse an urban environment have a multitude of possible applications. However, a robot that can be set to complete tasks such as food delivery or perform surveillance in various facilities that require navigation through humans has yet to reach mainstream deployment. This is mainly due to a lack of safety assurance, behavioral robustness, economic liability, and social acceptance. Although these barriers are difficult to overcome, today's easily accessible mobile computational capabilities embedded in hardware such as Nvidia's Jetson edge graphical processing unit (GPU) units allow the creation of a behaviorally robust solution that ensures safety and the implementation of stronger and more demanding solutions in the future.

1.3 Project Aim

Human behavior is challenging to predict as the motivation behind different human reactions are not easily classified and subsequent behavioral changes may be drastic in different situations. This project aims to create a model that will enable a quadruped mobile robot to perform social navigation using a camera for visual feedback in a first person POV. More specifically, this model can react and interact with pedestrians and maneuver through an unpredictable environment where path obstruction by pedestrians is probable. This project aims to create a model that considers the comfort of nearby pedestrians to ensure consistent harmonious human-robot interaction (HRI) while navigating a social environment. The proposed model will also tackle social navigational scenarios that require more than simple avoidance of humans, such as densely populated hallways or walkways where no free path is present.

1.3.1 Assumptions

As this project encompasses pedestrians behavior, the vast plethora of variables for the human aspect and demeanor must be placed under basic postulations to allow for a reasonable descriptive system. The assumptions that will enable a more direct approach to social navigation are listed below.

Pedestrian Behavior. To narrow the scope of the HRI covered in this project, the term social navigation is formulated such that pedestrian populated scenarios where navigation is performed only consists of humans with certain types of behaviors. This means that movement of pedestrians only include standing still, walking, or running, and exclude any radical movement or reactions. This assumption is placed to ensure continuous reaction regardless of feedback delay or sensor error.

Acceptance. Today there are few to no mobile robots walking alongside pedestrians on the street. Thus, a mobile robot of any form trying to navigate the streets would attract unavoidable and unpredictable attention. If humans tend to give more space and



Figure 1.1: Social navigation for a mobile robot

The images show how a mobile robot is enabled to correctly perform social navigation in key scenarios by using the social navigation model developed in this project. The left image is a depiction of a scenario where the robot must move to the other side of the hallway to avoid a collision with the oncoming pedestrian while also not needing to stop. The right image depicts a scenario where there is a crowd blocking the pathway in the hallway, but the mobile robot can maneuver comfortably around this by following another pedestrian that walks past the crowd.

leeway to people in wheelchairs, the a robot would disrupt human walking patterns even more. To make the project more realistic, it is assumed that pedestrians will accept a robot amongst them and display little to no reaction towards a quadruped robot as if it is another pedestrian.

Sensor constraints. While the robot is able to fully grasp the environment it is placed in, there are angles where the sensor hardware setup on the robot is limited. The robot would have a hard time getting accurate readings especially in the edges or borders of the sensor input where interpretation becomes unstable. The project therefore assumes that all scenarios are perceived and read correctly.

1.4 Contributions

The main contributions of this project are:

Enable a robot to perform social navigation. A mobile robot system with no navigational or social awareness had a social navigation model developed and tailored for it, granting it the ability to infer adequate behavior when navigating in a social environment. Given the nature of the project, a challenge and contribution is the implementation of a social navigation model on the chosen robot system.

Evaluating the social force pedestrian dynamics model. This project makes a detailed evaluation of the very popular and well established social force pedestrian dynamics descriptive model to identify and recognize its flaws. By analyzing model performance in different test scenarios in a pedestrian walkway, the model's force field approach was proved impractical for the chosen robot system. The magnitude of the forces and the consideration of direction were difficult to configure on the proposed robot system. The robot's avoidance behavior towards approaching pedestrians also proved inefficient with a considerable amount of unnecessary movement and no given speed control. The model further demonstrated a weakness for crowd formations, often

resulting in the freezing robot problem.

Designing a model solution for the found shortcomings. Building upon the shortcomings of the social force model (SFM), model augmentations were designed and proposed enabling the chosen robot system to execute quality social navigation. The proposed augmentations extended the SFM by redefining the repulsive forces and direction priority to allow for easier configuration and hyper parameter tuning, and smoother social interactions. The model extension incorporated a stronger emphasis on sideways motion for collision avoidance to increase the robot's path efficiency. The model was also modified to control its max speed based on the population of the environment. To solve the freezing robot problem the augmentations integrated the ability for the robot to be guided by another pedestrian through a crowd.

1.5 Implementation Platform and Hardware

The following summary of hardware make up the full robot system platform set to perform social navigation in this project.

The Mobile Robot: Spot. The mobile robot used in this project is the quadruped robot Spot designed and manufactured by Boston Dynamics. Spot is a quadruped robot with 12 degrees of motion, three on each leg. It is fitted with 360° vision field from five built-in black and white stereo cameras. The robot can perform accurate padding for an obstacle avoidance boundary. In this project this boundary padding will be set to the lowest possible distance as the social navigation would not rely on the brute force of obstacle avoidance. The robot balances itself autonomously and can walk with a speed up to $1.6ms^{-1}$.

Visual Perception: ZED2i Stereo Camera. While the mobile robot has full 360° of vision from its own built-in cameras, the more accurate and higher resolution stereo camera ZED2i from Stereolabs will be used as the main sensor for social interpre-

tation. The camera is mounted on top of the Spot robot on a custom camera stand to achieve an elevated POV. The camera will perform human detection and depth extraction to identify the behavior of humans in front of the robot.

Computer Solution: Nvidia Jetson Orin Edge GPU. Spot comes with a built-in computer running on an Ubuntu 18 Linux system. The processing power available through the built-in system is enough for very accurate locomotive calibrations for Spot, however, the required processing power is not enough. Thus, the newest edge GPU from Nvidia, the Jetson Orin, will be employed to take care of the computer vision along with the calculations of the actual social model.

Chapter 2

Literature Review

SOCIAL navigation encompassing the task of describing human behavior and robot navigation for HRI is a well-established field that has progressed immensely in the last 30 years. Despite the topic's popularity with the rising interest in the neighboring field of autonomous driving skyrocketing, the task of social navigation still faces significant challenges that are yet to be overcome. To present the background on social navigation, and to gain an understanding of the underlying nature of social navigation along with its challenges, this chapter will provide an in depth and up-to-date review of previous literature.

2.1 Challenges of Social Navigation

To present a holistic view of social navigation, the concept must be dissected into specific sub-topics. Mavrogiannis et al. [4] identified three main challenges within the complex problem of human-aware navigation:

- **Motion Planning**

Which must be considered when planning a route through a human populated area?

- **Interpretation of Human Behavior**

In order to be able to interpret human behavior an understanding of what social cues and signals humans use are necessary to ensure a robust navigation system.

In addition, what social behavior should a robot display as it maneuvers through a social scenario?

- **Model Evaluation**

What are the appropriate metrics and techniques for evaluating the efficiency of a social navigation model?

The tasks surrounding the three challenges stated above will in a broad sense create the full depth of what social navigation entails. As the introduction motivates, the project will mainly focus on social behavior and robot behavior along with motion planning. The width of the project scope does not include as much work on model evaluation, but to aid the readers understanding of the topic, an adequate summary of model evaluation will still be given.

Social navigation for a robot is, simplified, the task of behaving as a human would when maneuvering through a human populated area. In a sense it is the task of reaching a goal in the most efficient way staying free of any collision while also remaining neutral and cooperative in interactions towards people encountered in the environment. In order for a robot to reach a final destination, it must design a plan to reach this goal. This process can be divided into two parts, a general task of seeking the end point through checkpoints, and a more local small scale task of interacting with the immediate met obstacles along the way while maneuvering towards the next checkpoint.

The global objective deals with landscape obstacles in the environment and considers the terrain to plan the desired checkpoints towards the goal. The local objective manages the interpretation of and reaction to human behavior with a goal to ensure safety and efficiency. The task of safety is achieved through collision avoidance, social restraints, and policies regarding the robotic systems physical limitations.

In a dynamic social setting, humans have public and visually apparent objectives, as well as private objectives that are more difficult to interpret. In a crowd, human behavior is often motivated by others' actions and interactions. Such coordination is seen in letting a person pass by or two people holding hands [4].

Previously published surveys generated cost functions to describe a person's movement of global and local objectives rooted in the concept of social navigation [5, 6]. These surveys postulated that all social navigation models reward or punish on the basis of efficient global planning, coupled with local planning involving human-aware based considerations. Despite this the main difference between such a problem formulation and a problem of general navigation, is that in the case of social navigation any given agent do not have access to the full description of other agents planning costs. This has been the main rational behind incorporating prediction systems and techniques into social navigation models in recent years [3, 7, 8].

2.2 Path Planning for Social Navigation

Planning a route that leads to a chosen goal destination in a social setting requires an awareness of the surrounding environment and human actions. The evolution of path planning in social navigation traces back to navigation models that treated human as static and non reactive objects. The earliest demonstration of these models are on the famous tour guide robots MINERVA [1] and RHINO [2] that were deployed in museums to guide people from one attraction to another without colliding or stopping. They used obstacle avoidance as the main tool for path planning and treated humans as static obstacles without differentiating them from non-human obstacles. Although robot models have been successful at social navigation, these solutions lacked cooperation and negatively impacted humans through interference [9].

In the first generation of social robot models which lacked predictive capabilities, human-awareness was crucial for creating robust navigation models. This paradigm shift introduced model architecture emphasized the uncertainty of the operating domain. The focus has now shifted to infer human behavior in the engineering of socially aware models [10–13]. The first models that accommodated this shift performed complex inference of human behavior, but still failed to consider human response. The most common problem found in these models is the freezing robot effect where the robot is not able to alleviate itself when stuck and instead stops indefinitely to wait for clearance.

Even though socially unaware models were successful in their task, the field of social navigation has moved towards inferring human intention and model human reaction. New navigation models proposed that separate planning and prediction such as [14–16] and are all based on inverse reinforcement learning (IRL). Today most social navigation algorithms are based on the principle of human interaction as the first principle and approach prediction and planning as two aspects of the same task.

When path planning is coupled with motion prediction it allows for cooperation. The paradigm shifts further and becomes Cooperative Collision Avoidance (CCA) where scenarios with multiagent interactions are seen as a collective effort to avoid collision. The inference of human interactions is extremely complex. Gregory et al. [17] showed that the computational complexity of conducting probabilistic inferences on Bayesian belief networks proved to be NP hard. For human behavior the belief networks could represent inference on human behavior thus showing that the task of interpreting behavior has the same level of difficulty.

Approaches that couple human inference and motion planning use both explicit and implicit methods. Explicit models provide collision avoidance and inference data on its surroundings, which it feeds to the coupled path planning and prediction. There are several ways of implementing explicit approaches [18–21, 21–23], one of them is representing multi agent collective navigation inferences as topological invariants and braid groups by Mavrogiannis et al. and a few others. Presenting such inference using braid groups take inferences into a new domain. It gets rid of the complex dynamics of multiagent systems, which in turn they perform new interpretations of. This is then fed to the path planning and future agent prediction for social compliance. The braid theory uses neural network models to enable interpretation of domain shift.

Other models of explicit collision avoidance input are founded in game theory and used for multiplayer games. Turnwald et al. [24] created a multiagent motion planner that would enable a multiagent crowd to appear human-like solving for a Nash equilibrium for collision avoidance with the specific purpose of making computer animation convincing to humans. There are other less known explicit approaches not mentioned in this project.

Implicit models does the same as explicit ones, but does not actively set constraints on the coupled prediction and planning framework, but rather implicitly. This implicit influence are in most cases from neural network models that have been trained on some sort of multi agent collision avoidance simulation and will create a certain behavior from the model implicitly for any scenario it encounters. Most such neural network navigation solutions are within in the IRL paradigm. IRL is in general how a self learning agent should take action to maximize its cumulative reward. Cumulative reward is a sort of measure of the expected return, or scoring of outcome based on a given reward function. Essentially it is the technique of teaching an agent to abide by a certain set of rules using weighted reward and punishment. IRL is essentially the reinforcement learning paradigm backwards; the model is trying to learn the reward function that is behind the given some demonstrations [25]. IRL is mainly used in model design where the problem is not concretely defined or is unclear, yet there is data available or simulation possibility to show a model what actions are correct. IRL fits the paradigm of human-aware navigation perfectly as human behavior is ambiguous and difficult to fully describe.

The literature on implicit cooperative based collision avoidance in social navigation using IRL is extensive. Some of the More notable IRL frameworks are Ziebart et al. [9] IRL solution to promote robot behavior minimizing the disruption to predicted human paths following the ideology that humans collectively try to prevent divergence of other peoples trajectories. Another notable IRL model purposed by Henry et al. [26] successfully revolving around dynamics for human crowds. Their model used IRL to predict only partially observable environments with the motivation that most robot sensory input is subpar to human level sensory richness. Kreutzschmar et al. [15] created a model that could interpret minute human movements features (slight movement right or left) along with cooperative behavior in speed among others.

Although the presence of IRL solutions are significant in the social navigation community, there are perhaps as many utilizing deep reinforcement learning models. This area of the community has grown more recently with the considerable advances in the field of neural network and can create models with less defined goals, putting more focus on the

unsupervised learning [27–34]. The models pointed to here all have successfully attained an understanding of social navigation in more densely human populated areas with the goal predetermined or a set objective to reach. despite the large social awareness, by giving the models a predetermined goal the models become inaccurate or right out unable to perform pure navigation as they are not capable of path planning themselves.

Other model architectures that have received great attention for their success use Mixed Observability Markov Decision Processes (MOMDP) and Partially Observable Markov Decision Processes (POMDP) [35, 36]. The use of Generative Adversarial Networks (GAN) and Long Short-Term Memory (LSTM) networks have also taken up recently especially within the field of autonomous driving as the focus on this field has skyrocketed over the recent years [31, 37–42].

What can be summarized for the advances in the task of path planning for social navigation, is that a complex problem requires a complex solution and most such successful solutions are neural network based with implicitly learned strategies on behavioral concepts. While this is the case for the majority of research, the models that make inferences without weighted learning like POMDP, MOMDP, attention mechanisms, and Gaussian processes. There is simply no correct or preferred model architecture when going about path planing for social navigation.

2.3 Human Behavior in Social Navigation

Although the task of social navigation as explored in the previous section is possible to do without any consideration for human interaction or reaction modeling, they are vital for robust social navigation model design. As such this section will look at what and how inference of human behavior is tackled in research, what is required to model pedestrian behavior, and what implications arise in interactions with pedestrians. Despite human behavior analysis being more of a field in psychology, to add to the holistic view of the topic, human behavior can be approached in a more engineering oriented way. Mavrogiannis et al. [4] postulate that there are mainly three areas to consider in human

behavior when considering pedestrians; the personal space around a pedestrian, display of intention for future behavior, and pedestrian placement and formations in social spaces. All three comprise the main influences for reassuring human comfort while a robot is navigating in a social setting.

2.3.1 Personal Space

Proxemics is the field of study on peoples personal space and the distance people feel necessary to set between themselves and others. The cultural anthropologist Edward T. Hall first coined the term proxemics [43, 44] and noted that robots should refrain from coming too close to humans or pedestrians as perhaps the strongest social norm to follow. Hall described personal space as a circular area concentric with a pedestrians midpoint with several circles increasing in radius presenting different comfort zones. Later the description of proxemics where proposed as the shape of an egg by Kirby [45] and Hayduk [46] postulating that most pedestrians value their frontal space more than the space behind them. A community accepted and very popular way to describe the proxemics of a person is through repulsive force fields [47]. The social force model modelled the social space of a pedestrian to take the shape of an ellipsoid with the human being situated in one of the focal points with the other focal point directed in the same direction as the humans velocity. Further development on proxemics theorized that the prioritized personal space around a person is not a rigid space, but rather dynamic and asymmetric [48, 49]. Personal traits such as height, age, and gender can also impact the proxemics of a person as as considered by Amaoka et al. [50].

Another way of modeling proxemics is by using probability based models on human signaling when approached by an agent (human or robot) to delineate the appropriate approachable distance. Mead et al. [51] did this through computer vision rendered skeletons for humans poses, using simple Bayesian Inference to compute probability of interagent distance and orientation.

2.3.2 Display of Intention

The most complex and challenging inference to perform on a pedestrian is interpreting and understanding intention based movement. Intentions are often the most useful signals to fully predict pedestrian actions. The cognitive and psychological aspects involved in the interpretation of human intention offers great insight for social navigation. Psychological studies [52–54] show that human actions are always initiated with a purpose, and humans analyze others' actions for to understand the intended goal behind an action. This can translate into equipping a robot with the ability to indicate a robot's intention as a measure to gain human cooperation. Carton et al. [55] found that enabling a robot to display physical intention cues lead to a lower effort required for path planning as pedestrians better understood the intentions of the robot and were more willing to cooperate. Such cues can be facial expressions, hand and arm gestures, relative posture and position. Most robots, however, do not look like humans, making human-like indications unavailable. Hence, many social navigation models try to create legible robot motion that can be interpreted by surrounding pedestrians. Dragan et al. [56] showed that the legibility of an action is proportional to the understanding of the intention. This may seem obvious to humans, but robots tend to cut corners where humans would not to follow the shortest path. Dragan et al. showed that more articulated movements increased legibility and thus the pedestrian's understanding of the robot's intentions. Other publications on maximizing the legibility of a robot's movement and intention highlighting the creation of different curvatures of movements, increasing pedestrian perception of safety towards a robot, acceleration, and attention seeking behavior [57–61].

2.3.3 Position and Formation in Social Spaces

Social formation of pedestrians is infused with social signals and information. Majority of the research in this area concerns pedestrian crowds and grouping. A crowd is defined as a group of people gathered for a common purpose. For social navigation, a gathering of two or more pedestrians that are interacting, actively or subliminally, are considered. Grouping can be seen as a collective effort towards cooperation as pedestrians

adjust to the behavior of other pedestrians. However, social navigation is more focused on the interaction with pre-existing formed groups and not the process of group formation.

There are two types of pedestrian groups - static groups and dynamic groups. Static groups do not move. Examples include conversational groups or queues. A mobile robot is expected to respect static groups and avoid intruding on the space formed inside the group. A conventional tool used to analyze the formation of a static crowd is Kendon's F-formation [62] which describes the naturally occurring spatial organization that arises from human interaction. The space inside a static group is recognized as off limits, and can be used to detect the presence of groups. F-formation and deep learning approaches can be used for the detection of static groups [63–65].

Dynamic groups are groups where pedestrians are moving. Although not static, dynamic groups also exhibit social space that should not be intruded upon. MINERVA [1] and RHINO [2] mentioned earlier are examples of robots that are actively affecting pedestrian behavior in dynamic groups. Recent research by Garrell et al. on robots accompanying pedestrians [66] used several robots to guide people and analyzed the type of behavior that motivated people to leave or stay in the group. Their proposed method is similar to herding animals with robots, which may seem intrusive.

Observation overview is crucial for analyzing dynamic groups. Thus there are differences between the approaches based on sensor placement. Data from sensors placed in bird's eye POV over a pedestrian crowd would observe much more compared to a first person POV sensor mounted on a mobile robot. Research on behavioral analysis in pedestrians using externally placed sensors are frequently employing graphing methods [67, 68], and probabilistic progression models [69–71], and social force based methods [72, 73]. Despite the interest for the social force model in this project, work done using social force based methods for crowd analysis are mostly on pedestrian tracking and not on path prediction.

For mobile robots equipped with sensors, analysing groups becomes a difficult task as occlusion errors from pedestrians who are behind other pedestrians are inevitable. Other sources of error can be sensor inaccuracy caused by a robot's movement. For first person

view sensor input, a number of solutions have been proposed using probabilistic learning models on multi hypothesis tracking [74–76]. These solutions group and identify what observed moving pedestrians make up a social group, with the solutions from Luber et al. [76] having a consistent accuracy rate of over 80 percent when identifying groups in crowded areas.

even though solutions to understanding and interpreting pedestrian behavior often employs deep learning approaches, these solutions cannot be used for the analysis of pedestrians for social navigation to understand pedestrian behavior. Deep learning solutions present limitation when they learn to decipher very complex social cues, and lack the adaptability to be used in a partially applicable context. Henry et al. as mentioned earlier, used IRL to successfully navigate a "swarm" of people in simulation with only giving the agent first person POV input.

To efficiently navigate a social area, it is essential for a robot to interpret people and for them to interpret the motion of the robot. In social navigation research has shown that for a robot to navigate a space efficiently, it should consider a person's personal space and movement as a clue for their next intended action, and how people are grouped in a social space.

2.4 Evaluating Social Navigation Models

2.4.1 Metrics

Even though the primary objective of social navigation is path planning, there are several secondary objectives that are also required for ensuring effective social navigation. This makes concise and objective evaluation of this multi-faceted problem a challenge. The metrics discussed in the following sections primarily revolve around measuring a single agent's performance in social navigation.

To address the primary objective, the success rate of reaching a desired destination is used as a metric to evaluate the performance of a chosen social navigation model. This is based on a model's ability to close the physical distance between the start point and the

desired end point. However, it is also subjective in accordance to chosen constraints or preferences, such as the compliance of social norms. Measuring social compliance includes analyzing the comfort of the pedestrian(s), the safety of execution, the naturalness of the HRIs, or other psychosocial aspects. Depending on the objective, different metrics can be used to evaluate the quality of the model's performance. despite the lack of metric supremacy, the most common metrics used for evaluating social navigation according to Mavrogiannis et al [4] is listed in table 2.1.

Metrics	Definition
Arrival Rate	Percentage of trials where the agent ends up at the goal
Collision rate	Percentage of trials where collisions occurred
Path irregularity	How irregular was the movement of the robot compared to a humans
Time to goal	How long the robot used to reach the goal point
Safety	How safe were the motion and maneuvers of the robot
Path efficiency	How efficient was the path the robot followed
Naturalness of Behavior	How human-like was the robot behaving during HRI
Social Comfort	How comfortable were surrounding pedestrians
Success rate	How well did the robot navigate in a social space given its performance when encountering other pedestrians along the way

Table 2.1: Metrics for evaluating a robot's social navigation ability.

Success Rate

The success of a robot's execution of social navigation is not entirely ensured even if it reaches the final destination. Although arrival rate is an important metric used in both social navigation and robot navigation [77, 78], the success rate is the arrival rate but with a set of added assumptions. Time is often used alongside arrival rate and if the robot takes too long to reach the goal point, the success rate would be lowered. similarly, time is also subjective as the environment encountered can have a high crowd density meaning navigation would take longer. Given a robust social navigation model, the time of arrival can be an indicator of crowd density. Success rate mainly depends on reaching the target goal point, but also incorporates other aspects of the robot's performance.

Collision Rate and Safety

The most crucial aspect of a robots performance is the safety of the robot's action. Safety can be a collective measure of the number of violations on chosen constraints. The most common constraint is the number of collisions, and the number of intrusions on pedestrian personal space [79]. Three safety criteria a robot must comply with when performing navigation around humans are defined in [80]. A robot must consider the consequences of its own dynamics and predict the behavior of encountered pedestrians. lastly, perhaps the hardest to implement, a robot must make reason over a time-horizon spanning from the presence and infinitely into the future.

Path Optimality

While success rate is based on reaching the goal, path optimality is a measure of the quality of the robot's trajectory. The efficiency and irregularity of movement are indicators of path optimality. Path efficiency is the measure of the time taken to maneuver through the environment and path irregularity is the measure of deviation from the chosen path. Less obvious measures such as average acceleration and average energy were first proposed by Mavrogiannis et al. [79] to measure subliminal motion changes for a pedestrian's motion.

Naturalness of Behavior

How natural or human-like a robot's behavior is is not a direct measure of path planning. However, but for social navigation it indicates how well a robot can adapt to a human environment. Measure of comfort indicates how well a robot blends in with other pedestrians, but lacks objectiveness since comfort is highly subjective from human to human. comfort is often based on the acceptance of how comfortable people are with a robot appearing in a public space.

An objective way to evaluate at naturalness of a robot's behavior would be to compare the robot motion with a dataset of recorded human motion performing the exact same task. Deviation from human motion is commonly measured and two metrics are

accepted in the community of social navigation - Average Displacement Error (ADE) [81], and Final Displacement Error (FDE) [82]. ADE is the mean difference between path points aligned in time set by the navigation model and the ground truth represented by human motion. The FDE measures the difference in displacement at the goal point temporally aligned between robot navigation algorithm and human data set.

2.4.2 Evaluation Methods

As mentioned earlier, a valid way to evaluate the social navigation capabilities of a navigation algorithm is by comparing it to a data set of human motion in the same environment as the robot. While this is true, a real life data set presents several challenges. First, humans in a recording of social scenarios will almost always demonstrate a change in behavior, based on the awareness of the given context. Second, scenarios do not always show all possible behaviors, thus leading to some valid robot behavior being measured wrongly. Third, when recording human social navigation behavior, there is a lack of extreme constraint violations which are necessary to evaluate a robot's behavior in extreme scenarios.

Another commonly used method for assessing social navigation models is simulation. Simulation has the advantage of being scaled to magnitudes bigger than any real life data. This project does not include any simulations. However, crowd modelling in simulations is commonly based on the SFM [47], Velocity Obstacle (VO) model [83], the reciprocal velocity obstacles model [84], and the optimal reciprocal collision avoidance model [85].

Chapter 3

Hardware, System Design, and Implementation

THE mobile robot used in this project, as first introduced in chapter 1, is the Boston Dynamics Spot quadruped robot. The four legged robot Spot is a holonomic system and has the advantage of being able to move in the x and y direction at any given time in comparison to a nonholonomic wheeled robot. Mounted on top of Spot's back, as a payload, is its eyes and brain; a Zed2i stereo camera and a Nvidia Jetson AGX Orin edge GPU. The final mobile robot system can be seen in figure 3.1.

3.1 Mobile Robot

The already introduced mobile root Spot has 12 degrees of freedom, three on each leg, two shoulder/hip joints and a knee joint. This makes the back of Spot align horizontally with very little deviation even when the robot walks. The robot has two odometry frames; a local odometry frame based on locomotive movement and its accelerometer, and a vision odometry frame based on vision localization from its five built in stereo cameras. These two odometry frames are merged together to create a more accurate world frame. The center of Spot's body is the location of the robot's base link which is where the robot denotes its whereabouts in the odometry frame. The world frame origin is set as the robot's base

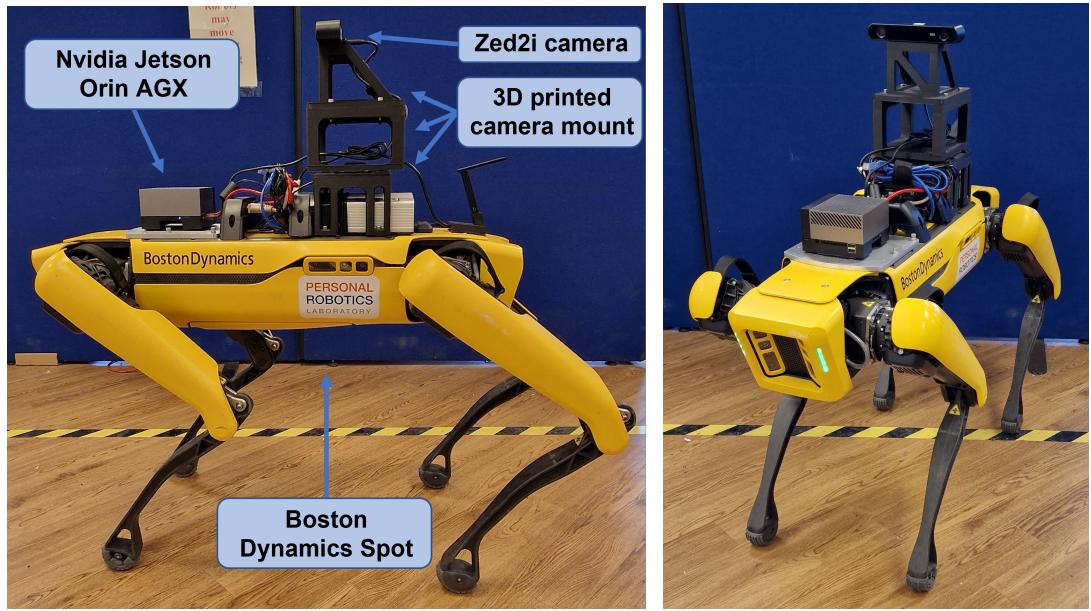


Figure 3.1: The Mobile Robot System

The mobile robot system consists of the quadruped mobile robot Spot from Boston Dynamics. Spot is carrying the systems computing unit, the edge GPU Nvidia Jetson Orin, along with the stereo camera Zed2i form Stereolabs mounted on top of custom 3D printed camera mount.

link upon boot. The odometry frame of Spot follows the universal axes notation that can be described by the right hand rule (index = x , middle = y , thumb = z).

Through Boston Dynamic's API for Spot, the robot takes two types of robot motion commands; a velocity value in either x or y direction from its current position or a pose containing a certain rotation and odometry coordinates to where Spot will assume the pose.

Spot has an onboard computer. The only way to get into the system is by connecting an external display, but given the mobile nature of the robot having an external display physically connected is highly inconvenient. As such the solution to get onto the Spot system is through a Virtual Network Computing (VNC) link from a desktop connected directly to Spot's internet. A VNC connection essentially gives cross-platform screen access to the system from a remote computer through a virtual control connection.

3.2 Visual Perception

3.2.1 Sensor Choice

The choice to give the robot a visual perception as a way to perceive its environment, and what sensor this was done through was based on several reasons. The most common environment perceiving sensors to put on a mobile robot are time of flight (TOF) sensors and cameras. TOF sensors can be laser, infrared, ultrasound, and LiDAR sensors and while a TOF sensor can give very accurate distance measurements for mapping, it is difficult to make accurate readings of a dynamic environment using only one dimensional input as the changes cannot be readily interpreted. That being said TOF sensors are ideal for less complex tasks such as obstacle avoidance as a source for accurate distance reading from obstacle to robot. The most dominant sense for humans is the visual input from the eyes. The objective of how to successfully perform social navigation as motivated in chapter 2, is in large a task of replicating human behavior. This motivates the use of vision sensors for perception as neural network solutions allow for accurate human pose detection. This, as shown in chapter 2, was also the main sensory input in the majority of past attempts at social navigation for mobile robots.

There are several types of cameras that can grant a robot visual perception; regular RGB camera, RGB-D camera, depth camera, and stereo cameras. While it is possible to deploy detection networks to any live visual feed and get accurate object detection, a depth reading made onto the same FOV as the RGB camera feed of the surroundings is of vital importance in this project as it will give an understanding of the spacial orientation for a perceived object.

The chosen camera as mentioned in chapter 1 was the Zed2i stereo camera [86], but before this was decided, the built-in stereo cameras on the Boston dynamics Spot were first tried and tested to see if the robot's own cameras were enough. Spot's five stereo camera setup has the advantage of giving 360° vision. The stereo cameras are however gray-scale and previous work done on this by a member of the PRL lab at Imperial college found that the camera resolution, being only 720p, along with the gray-scale made for a relatively

poor solution to run object detection on. The object detection on a simple object such as a teddy bear would not be detected at all times, which in turn would introduce the liability of not being able to detect humans at all times when around the robot. Another shortcoming of Spot's own stereo cameras are their more downward facing POV. A person standing next to Spot would only be shown as their legs and hips making it difficult for accurate human pose estimation. The built-in stereo cameras are mainly used for accurate distance reading and positional tracking in the robots vision odometry frame hence the focused POV on the sides of the robot and its feet. A last point that would not allow for good social navigation capabilities is the limited range of 4 meter accurate distance readings.

Another camera that was also considered was the Intel RealSense L515 RGB-D camera. This RGB-D camera would allow to overcome the additional adjustment of fusing a point cloud from a depth sensor to a camera image feed as the RealSense camera delivers full HD images with depth data at a rate of 30 frames per second (fps). The drawbacks of the RealSense was its relatively narrow field of view with only 70° horizontal view and 55° vertical view along with a max depth range of only 9 meters.

The chosen Zed2i stereo camera from Stereolabs has a field of view of 110° horizontal and 70° vertical, and a depth and detection range of up to 20 meters making it superior to the RealSense. The frame publish rate of the Zed2i camera is however limited to 15 fps. The ZED2i stereo camera has a significant advantage to the other tried cameras as it comes with its own out of the box human pose detection model built-in and optimized. The object detection model created by Stereolabs does not rely on just classical stereo matching algorithms, but utilizes trained deep neural network models to improve the depth extraction process with improved stability and quality from its latest generation built-in Inertial Measurement Unit (IMU), barometer, magnetometer, and thermometer.

3.2.2 Human Detection

When the Zed2i is set to perform human pose detection the camera identifies as many people as there are in the frame as long as their detection certainty is above 70% allowing

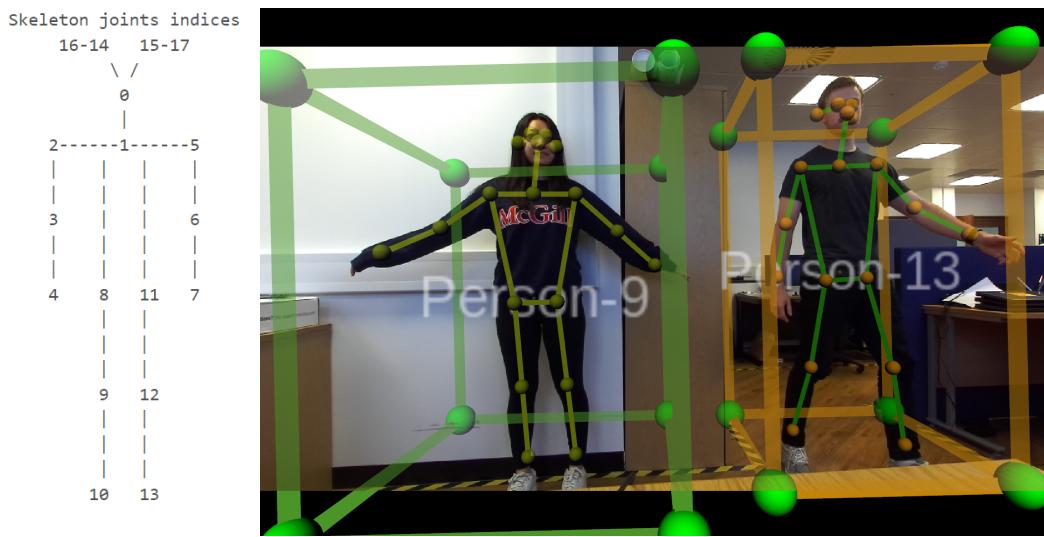


Figure 3.2: The Human Pose Detection Performed by the Zed2i

The image is a visualization of the 18 keypoints the Zed2i camera tracks to detect human poses. The image shows the skeleton joint indices (left) and the skeleton overlaid over two detected humans in the image frame (right). This also shows that the Zed2i stereo camera model is capable of detecting multiple people within the same picture frame. The camera tracks limb position relative to the torso of the detected human as well as the position of their eyes, nose and ears.

for some occlusion to occur. the camera fits onto the detected humans a skeleton with 18 coordinate keypoints, see figure 3.2 for the placement of the human skeleton keypoints. The Zed2i detection model also outputs accurate information of the x, y, z coordinates of the main body, which is the position of skeleton keypoint zero. The detection model also provides an accurate reading of what the velocity of any detected human is in the frame of positive and negative x and y direction in the camera's world frame. The camera's odometry frame like Spot's odometry frame is set with the universal vector axes following the right hand rule. Although the resulting distance and velocity values for a detected human come with a percentage error, it is small and performs very accurately. The camera is also capable of accurately labeling a person as standing still or moving with a simple Boolean variable. The camera's object detection model provides more information however the rest of the information output were not used in this project.

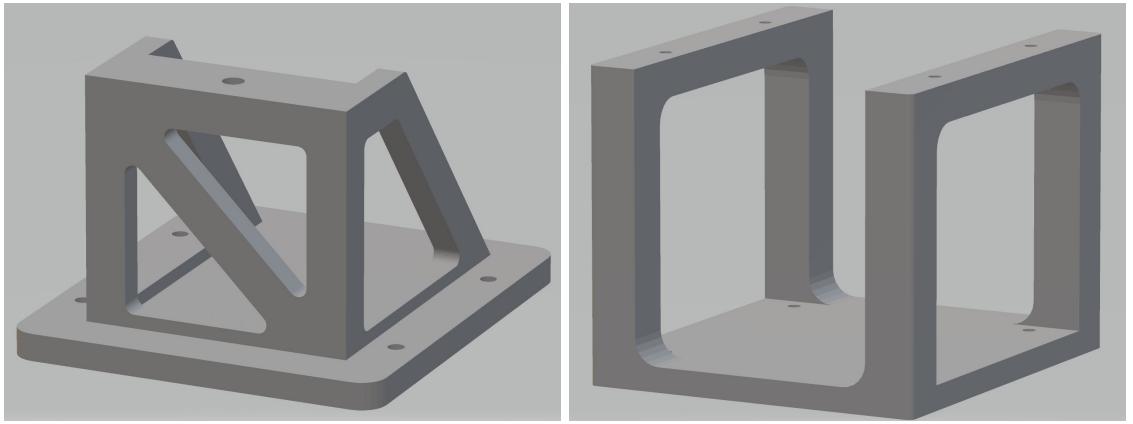


Figure 3.3: 3D Printed Camera Mount

To elevate the camera to a higher field of view to get better detection of humans close to the robot, a mount was 3D printed. The mount consisted of upper (left) and lower (right) parts.

3.2.3 Camera Placement on the Robot

As was discussed in chapter 2, the camera placement is detrimental to performing social navigation. The better the camera overview is of the environment the more data one can utilize for navigating among pedestrians. As the robot is to be able to navigate unknown environments, the camera cannot be remotely placed, but must be placed onto Spot itself.

The strategy for the placement of the camera as seen in 3.1 placed on top of Spot further back than the robots middle point was to give the camera as wide a view as possible of what is in front of the robot, including as much peripheral vision as possible.

Just as Spot's own cameras were not fit for human pose detection as their field of view was angled downwards, if the Zed2i camera was placed directly on top of Spot's back plate it would be unable to detect humans too close to itself. To overcome this problem the camera was placed on top of two custom 3D printed parts placed onto Spot's back that would elevate the camera up by $0.28m$ which created a much better viewing angle more fit for human pose detection. Figure 3.3 shows the two parts printed creating the camera mount. The parts were printed using the thermoplastic monomer Polylactic Acid (PLA) and proved quite sturdy with the camera not shaking noticeably more after being elevated further away from Spot's center of mass.

3.3 Computing Unit

Spot's with its computer inside only allow for custom code to be run through its Software Development Kit (SDK) over a WiFi connection, but an additional edge Central Processing Unit (CPU), SpotCore, can be mounted onto the robot to enable software run locally with much higher bandwidth and reduced latency. The Spot used in this project had a SpotCore mounted onto it. The SpotCore is an i5 Whiskey lake-U Intel CPU (3.9GHz), with 16Gb DDR4 RAM operating on an SSD. These specifications make for fast computer performance, but as a large part of the data processed from the Zed2i is image data, SpotCore not having a dedicated GPU would not perform as well as a computer with a GPU. Therefore the option of using an edge GPU was explored.

Nvidia's Jetson Xavier NX edge GPU with 384 CUDA cores and 48 Tensor cores was tried. The performance of the Xavier was better than the SpotCore itself, but when running the Zed2i object detection inference the Xavier struggled to keep the frame rate at 15Hz consistent and would drop. The next and newest generation of Nvidia edge GPUs, the Jetson Orin AGX, was also tried in comparison. The Jetson Orin AGX has a total of 2048 CUDA cores and 64 Tensor cores. The Jetson Orin has a promised minimum 3.3 times performance increase in performance over its predecessor, the Xavier, and it was very apparent as the Jetson Orin kept the object detection publishing rate at 15Hz consistently with few to no frame drops over time. As such the chosen processing unit was NVidia's Jetson Orin AGX edge GPU.

Since the Jetson Orin provided a higher performance it also required more power than the Jetson Xavier NX. The Xavier at its highest performance settings draws 30W, while the Orin at its highest performance settings draws 50W. This introduced some complications as Spot's external payload power port only delivered 12V on four of its HD15 connector. Now if a single wire is set to supply the power to the Jetson Orin, the cable has to carry $50W/12V = 4.167$ Amps of current. Most household wires are not of a wire diameter large enough to carry this high a current. Therefore a wire with the current capacity had to be used to connect the payload power pins and the Orin. Two 16 AWG rated (or $1.3mm$ in diameter) wires were used to deliver the power from the HD-15

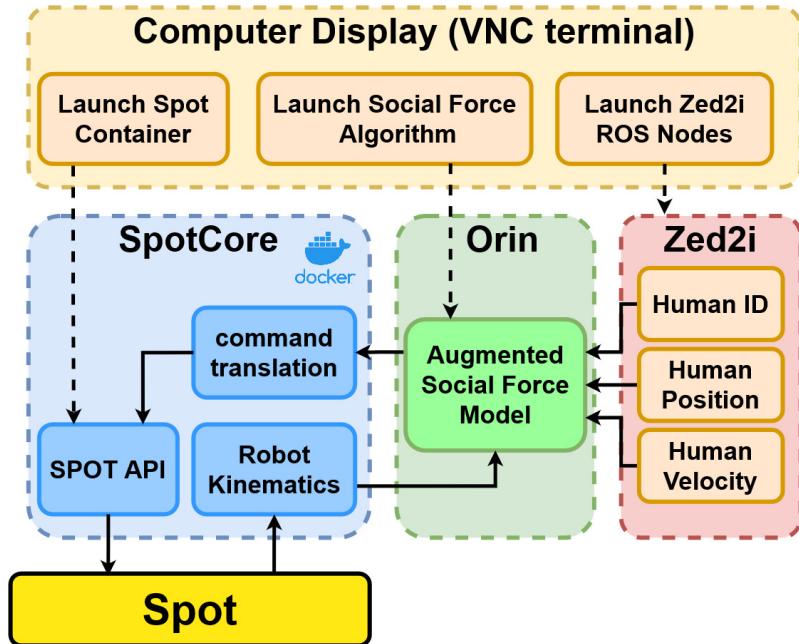


Figure 3.4: System Execution Flowchart

Flowchart over how the hardware system executes the social navigation. The VNC terminal link launches the docker container on the SpotCore starting the master roscore, the ZED2i camera, and the social navigation model on the Jetson Orin. Both the Social navigation model ROS nodes and the Zed2i camera ROS nodes are publishing to topics visible to the SpotCore container which subscribes to them and executes commands to Spot based on the information published. The Social force model along with the command translation and Spot API are all written in python.

payload connector to the Orin. Although a wire rating of 16 AWG is rated for 3.7Amps at 30° Celsius, the wire could tolerate a higher heat so drawing the final 0.367 Amps was possible despite the temperature increase.

The external edge GPU, being external, had to be mounted onto Spot's back as part of the payload. This was done using Velcro tape to make it easy to remove.

3.4 Implementation

The communication between the robot, its sensor(s), the external processing unit, and its computer interface were all done in Robot Operating System (ROS) along with Secure Shell (SSH) protocol links between the control laptop and Spot, and the control laptop and

Jetson Orin. The systems digital integration can be seen in 3.4 where all black arrows are ROS nodes publishing to ROS topics, which in turn the receiving block subscribes to. The black dashed arrows going from the VNC terminal are ROS terminal launch commands.

The box with dashed outline labeled SpotCore on the system integration diagram, is a docker container launched from the actual SpotCore. The need for a docker container on the SpotCore machine was because the SpotCore itself does not have ROS installed, and so using a docker container would allow for installation of all packages needed without having to install it on the actual SpotCore itself.

To allow for the Jetson Orin to see the ROS topics published by the nodes from the SpotCore and to allow the SpotCore access to the topics that the Jetson Orin posted, the SpotCore was set as the master roscore.

Another hardware hurdle that came with the SpotCore not having anything installed on it, was a discrepancy in date and time settings between the Jetson Orin and the SpotCore container. To have two different topics of data as input to the social navigation model that was running on the Orin, the time for when each topic publishing had to be the same in order for the script to process them at the same time. By setting the Orin as slave to the time of the SpotCore, the time was synchronized.

The Zed2i camera was connected to the Jetson Orin by cable and had its own open source ROS wrapper developed by the manufacturer Stereolab. The wrapper made the Zed2i camera a plug and play solution - publishing to the roscore that the Jetson Orin was slave to after it was given its launch command.

Chapter 4

The Social Navigation Model

ATTEMPTING to describe pedestrian behavior is a well established area of research as was explored in chapter 2, and perhaps the most robust and accepted foundational model to describe such behavior is Helbing's social force model [47]. In this project the main behavioral descriptive model used will be based on the social force model.

4.1 Choice of Model

A number of models prior to the force model were based on strong analogies between kinematics in gasses and pedestrian behavior, the creator of the social force model, Dirk Helbing, tried along with others to describe pedestrian behavior founded in the similarities between fluid dynamics and human crowd behavior although not very successful for wide case use [47].

More recent models rely on training some type of neural network to learn social navigation abilities, as discussed in chapter 2, with many proposed models successfully doing so utilizing learned methods. This project did not try a method based on any type of required learning as the time horizon of the project was too short to incorporate any type of data set recording to learn on. Some datasets exist as mentioned earlier, but they are vastly different from the camera input that the robot system has, making for a poor fit. Another common approach to train a learning model is through simulation. Again,

the project was too limited to include any sort of integration of real life robot system to simulation. Therefore a general pedestrian dynamic descriptive model was chosen to facilitate for better integration. Most successful navigational models are not considered for first person POV robots. POMDP and MOMDP are very successful but require a static camera with a complete view of the environment. The Velocity objective model [83] was considered as a contestant to the SFM, but the Velocity Objects model puts less focus on pedestrian interaction and is mainly concerned with avoiding dynamic obstacles. As the project is on social navigation the SFM was thus deemed more human-like in behavior and chosen.

4.2 The Social Force Model for Pedestrian Dynamics

The social force model describes pedestrian movement by a sum of "social" forces acting upon the pedestrian. The model, despite excluding many aspects of actual motivations for a humans movements, proves very robust at describing what objectives motivates and drives core pedestrian movement [47]. The social model takes its inspiration from the way Newtonian gravitational force fields behave and tries to model a pedestrian's motivations for movement through attractive and repulsive forces. The simplest social force model deploys forces as a means of acceleration influence on the pedestrian's velocity and direction.

The attractive force is directed from the pedestrian towards the nearest visible goal the pedestrian is trying to reach. As such a checkpoint sort of approach motivates a constant pull of the pedestrian directly towards the next checkpoint. Helbing's force equations are modelled similarly to how Newtonian physics describe force. In Newtonian physics $F = ma$, while in Helbing's proposed social forces, the mass aspect is taken out of the equation with $m = 1$ at any time making force the acceleration a pedestrian has in a specified direction.

The attractive force influencing pedestrian α is described by the following equation:

$$\mathbf{F}_\alpha^{attr}(v_{desired}, \hat{\mathbf{e}}, v_\alpha) = \frac{1}{\tau} (v_{desired}\hat{\mathbf{e}} - \mathbf{v}_\alpha). \quad (4.1)$$

where τ is what Helbing calls the relaxation time, $v_{desired}$ is the desired maximum speed the pedestrian wants to reach, $\hat{\mathbf{e}}$ is the unit vector of the desired direction towards the field source, and \mathbf{v}_α is the current velocity of the pedestrian. Although 4.1 is based on velocity, it is only a number related to ratio and can be thought of as acceleration. When the pedestrian has zero current velocity, 4.1 will be greatest meaning the acceleration towards the force source will be the largest. When the current velocity of the pedestrian eventually reaches the desired velocity $v_{desired}$, the acceleration will be zero meaning no further velocity increase. The relaxation time given by τ is what is responsible for a slower increase in velocity based on the force or as it decreases the overall acceleration.

The repulsive forces deployed by the model comes from other pedestrians and barriers where they all exert a force directed away from themselves. Helbing's proposed pedestrian repulsive force from pedestrian β onto pedestrian α is represented by

$$\mathbf{F}_{\alpha\beta}(\mathbf{d}_{\alpha\beta}) = -\nabla_{\mathbf{d}_{\alpha\beta}} V_{\alpha\beta}[b(\mathbf{d}_{\alpha\beta})]. \quad (4.2)$$

Where $\mathbf{d}_{\alpha\beta}$ is the distance between pedestrian α and β , $\nabla_{\mathbf{d}_{\alpha\beta}}$ is the gradient of the field at the given distance, V is an exponential decreasing function with respect to b taking the shape of an ellipse. V is given by

$$V_{\alpha\beta}(b) = Ae^{-b/B}. \quad (4.3)$$

for certain parameters A and B . b represents the semi-minor axis of the ellipse the pedestrian is surrounded by and is given by

$$2b = \sqrt{(||d_{\alpha\beta}|| + ||d_{\alpha\beta} - v_\beta \Delta t \mathbf{e}_\beta||)^2 - (v_\beta \Delta t)^2}. \quad (4.4)$$

Where Helbing introduces the term $v_\beta \Delta t$ which signifies the step size of pedestrian β . For the b term as described in 4.4, the ellipse grows in size the higher pedestrian β 's speed

is, but does not change shape with a fixed semi-minor to semi-major ellipse axis ratio. To understand Helbing's repulsive force for pedestrians 4.2 can be rewritten into a more descriptive format

$$\mathbf{F}_{\alpha\beta}(\mathbf{d}_{\alpha\beta}) = Ae^{\frac{-b_i}{B}} \frac{||d_{\alpha\beta_i}|| + ||d_{\alpha\beta_i} - \mathbf{Y}_i||}{4b_i} \left(\frac{d_{\alpha\beta_i}}{||d_{\alpha\beta_i}||} + \frac{d_{\alpha\beta_i} - \mathbf{Y}_i}{||d_{\alpha\beta_i} - \mathbf{Y}_i||} \right) \quad (4.5)$$

Helbing also theorized that walls and borders have a similar repulsive effect on pedestrians and proposed an additional repulsive force between pedestrian α and border B written as

$$\mathbf{F}_{\alpha B}(\mathbf{d}_{\alpha B}) = -\nabla_{\mathbf{d}_{\alpha B}} U_{\alpha B}(||\mathbf{d}_{\alpha B}||). \quad (4.6)$$

Where $U_{\alpha B}$ is another decreasing exponential function as 4.3 tuned for walls and borders. Helbing also gave direction of pedestrian motion an influence on the forces. As the social force formulas hold proper effectiveness if observed in the desired direction, unwanted influence from other directions is handled by a direction dependent weighting function

$$w(\mathbf{e}, \mathbf{F}) = \begin{cases} 1 & \text{if } \mathbf{e} \cdot \mathbf{F} \geq \mathbf{F} \cos(\phi) \\ c, & \text{otherwise} \end{cases}. \quad (4.7)$$

Where c is $0 < c < 1$ to weaken forces acting behind the pedestrian, and ϕ is set based off of the effective angle of vision (2ϕ). Helbing originally also included another attractive force describing an attraction towards other pedestrians, but as this is something that lies outside the scope of this project, it will not be included. As such Helbing's *social force* model can be formulated as the summation of all the aforementioned forces

$$\mathbf{F}_\alpha = \mathbf{F}_\alpha^{attr}(v_{desired}, \mathbf{e}, v_\alpha) + \sum_{\beta} w(\mathbf{e}, \mathbf{F}_{\alpha\beta}) \mathbf{F}_{\alpha\beta}(\mathbf{d}_{\alpha\beta}) + \sum_B w(\mathbf{e}, \mathbf{F}_{\alpha B}) \mathbf{F}_{\alpha B}(\mathbf{d}_{\alpha B}). \quad (4.8)$$

4.3 Social Force Model Implementation

The pseudocode for the model implementation on the robot system can be seen in algorithm 1. To implement the algorithm on the mobile robot system, some tailoring were needed, and as can be seen in the algorithm, the main function is a loop that runs every

time it receives data from both the camera and the robot kinematics. This is capped by the cameras ROS topic output as the cameras output publish frequency is 15Hz whereas the Spot kinematics ROS topic has a publish rate of 400Hz.

The path planning approach of the navigation model is based on predetermined checkpoints that have been recorded in the robots odometry frame in advance for the given environment the robot is set to navigate through. The robot will thus go from checkpoint to checkpoint with the condition of coming within one meter distance away from the coordinates of the current checkpoint goal before it sets the goal to the next checkpoint. An illustration of the checkpoint approach can be seen in figure 4.1

The implementation as shown in algorithm 1 outputs a velocity which is then sent to Spot through its API. Something that the social force model does not take into account as a pedestrian motion model that must be accounted for in order for the robot system to function, is the orientation of the camera. The goal is to reach the next given checkpoint and as such the camera should be facing the direction of the goal to detect humans that appear before reaching the goal. This means that at all times the robot will be rotating - adjusting to always face the goal location regardless of velocity input direction from the social force model.

Algorithm 1 Social force Model

Parameters: $v_{desired}$, τ , $v_\beta \Delta t$
Function Attractive_Force(*goal*, *position*, \mathbf{v}_α , $\hat{\mathbf{e}}$):

$$\mathbf{F}_\alpha = \frac{1}{\tau} (v_{desired} \hat{\mathbf{e}} - \mathbf{v}_\alpha)$$

return \mathbf{F}_α
Function Social_Force(*Obj_det*, $\hat{\mathbf{e}}$):

for *i* in number of humans **do**

$$\mathbf{d}_{\alpha\beta_i} \leftarrow \text{distance to pedestrian } i$$

$$v_{\beta_i} \leftarrow \text{magnitude of pedestrian } i\text{'s velocity}$$

$$\hat{\mathbf{e}}_{\beta_i} \leftarrow \text{unit vector of pedestrian } i\text{'s direction}$$

$$\mathbf{Y}_i = v_{\beta_i} \Delta t \hat{\mathbf{e}}_{\beta_i}$$

$$b_i = \frac{1}{2} \sqrt{(\|\mathbf{d}_{\alpha\beta_i}\| + \|\mathbf{d}_{\alpha\beta_i} - \mathbf{Y}_i\|)^2 - (v_{\beta_i} \Delta t)^2}$$

$$\mathbf{f}_{\alpha\beta_i}(\mathbf{d}_{\alpha\beta_i}) = A e^{\frac{-b_i}{B}} \frac{\|\mathbf{d}_{\alpha\beta_i}\| + \|\mathbf{d}_{\alpha\beta_i} - \mathbf{Y}_i\|}{4b_i} \left(\frac{\mathbf{d}_{\alpha\beta_i}}{\|\mathbf{d}_{\alpha\beta_i}\|} + \frac{\mathbf{d}_{\alpha\beta_i} - \mathbf{Y}_i}{\|\mathbf{d}_{\alpha\beta_i} - \mathbf{Y}_i\|} \right)$$

$$\mathbf{F}_{\alpha\beta} = \mathbf{F}_\alpha + w(\hat{\mathbf{e}}, \mathbf{f}_{\alpha\beta_i}) \mathbf{f}_{\alpha\beta_i}(\mathbf{d}_{\alpha\beta_i})$$

end
return $\mathbf{F}_{\alpha\beta}$
Function Main(\mathbf{F}_α , $\mathbf{F}_{\alpha\beta_i}$):

while receiving robot kinematics & humans detected **do**
 $\text{position} \leftarrow \text{odometry from robot kinematics}$
 $\text{goal} \leftarrow \text{predetermined path points}$
 $\hat{\mathbf{e}} \leftarrow \text{unit vector from robot to goal}$
 $\mathbf{F}_\alpha \leftarrow \text{Attractive_Force}(\text{goal}, \text{position}, \mathbf{v}_\alpha)$
 $\mathbf{F}_{\alpha\beta} \leftarrow \text{Social_Force}(\text{detected_humans})$
 $t_{now} \leftarrow \text{time of instance}$
 $\Delta t = t_{now} - t_{past}$
 $\mathbf{v}_{new} \leftarrow \mathbf{v}_{current} + (\mathbf{F}_\alpha - \mathbf{F}_{\alpha\beta}) \Delta t$

 Publish \mathbf{v}_{new} to robot

 $\mathbf{v}_{current} = \mathbf{v}_{new},$
 $t_{past} \leftarrow t_{now}$
end
end

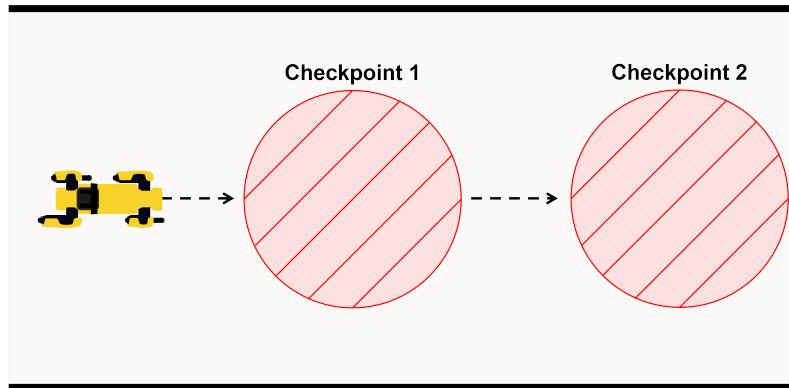


Figure 4.1: Path planning approach for model implementation.

The path the robot will take through an environment is a set of prerecorded checkpoints in the form of odometry coordinates relative to the odometry reference frame. If the robot's base link comes within a meter distance away from the current goal checkpoint, the goal point is changed to the next checkpoint.

Chapter 5

Model Augmentation

THE original SFM by D. Helbing [47] was created as a means to accurately describe the behavior of pedestrians and the dynamics motivating pedestrian motion. As a result, translating such a descriptive model into a social navigation solution, as shown in chapter 4, fit to be deployed on the robot system presented in chapter 3, will have shortcomings/flaws/issues/failings. The SFM flaws specific for the application of social navigation, are the reasons and motivation behind creating the Augmented Social Force Model (ASFM). The ASFM is this project’s answer to a social navigation solution for mobile robots with a model framework based on the force-field theory introduced by the social field model.

The social force model’s shortcomings will be discussed more in detail in chapter 6, but they are the direct motivations behind the augmentations developed in the ASFM. The major drawbacks of the social force model were

1. Hard to tune elliptical pedestrian personal space shape.
2. Non-intuitive backwards motion when approached head on.
3. No particular awareness of human formations.
4. Poor practical solution for lowering repulsive forces using the direction dependent weighting function.

5.1 Social Force Augmentations

5.1.1 Decreasing prediction error and model complexity

The first augmentation made is to create a model with force fields that are more robust and easier to tune to deal with shortcoming 1. The proposed augmentation redefines the repulsive forces and utilizes the robot hardware better. The Zed2i camera makes an accurate classification if a human observed is moving or not, so to lower error in detected human pose movement, the repulsive force was implemented to utilize different force field models based on the pedestrians movements; a radial vector field for static pedestrians, and an elliptical shaped force field for pedestrians with a velocity different from the original model. Both of the vector fields are concentric fields with the pedestrian placed in the middle of the vector field. The circular shaped personal space will give of a radial force field modeled by the following exponential:

$$F_{\alpha\beta-static} = Ae^{Bx+C} \quad (5.1)$$

Where A, B , and C are hyper parameters to create a force field that will allow for smoother encounters than 4.3. As the magnitude of a radial force field is the same for a set distance away from the center, the x in 5.1 represents the distance from the pedestrian to the robot.

As was explored in chapter 2 regarding the evolution of proxemics, the choice of applying an elliptic vector field onto a moving pedestrian to represent repulsive force, was to allow for a larger personal space region in front of a moving pedestrian as an elliptical vector field also allows for a change in shape depending on the pedestrian's speed. If a pedestrian is walking with a relatively slow speed the ellipse can be made smaller and closer to a circle in shape, and if the pedestrian walks faster the semi-major and -minor axis can be extended to create a larger vector field. The faster a pedestrian walks the bigger the ellipse axes grows, but the semi-major axis grows faster than the semi-minor axis elongating the elliptical vector field as shown in figure 5.1.

The elliptic vector field is modelled in much the same way as the radial vector field by using an exponential function, but shifted according to the distance of ellipse that

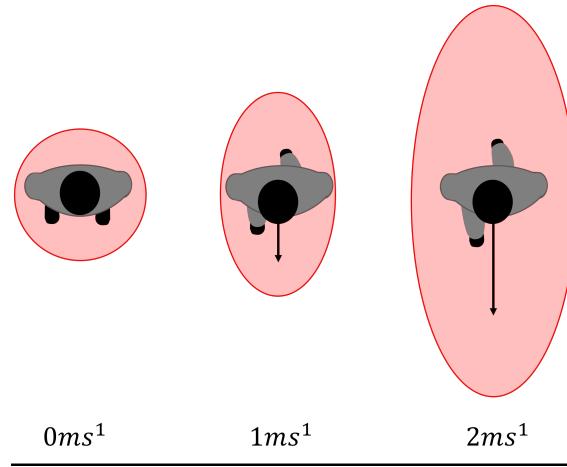


Figure 5.1: ASFM change in force field

The force field shape for the ASFM changes with regards to a pedestrian's velocity. If the pedestrian is static then the force field takes a perfectly circular shape. If the pedestrian moves the force field elongates into the shape of an ellipse. The semi-major axis of the ellipse grows proportionally faster than the semi-minor axis. This creates a space in front of a moving pedestrian that the robot will avoid to allow for comfortable and more natural bypassing of pedestrians.

surrounds the moving pedestrian. The exponential function for force magnitude used on an elliptical vector field is:

$$F_{\alpha\beta-moving} = Ae^{(x-d)B+C} \quad (5.2)$$

Where the hyper parameters A, B, and C have the same values as set for 5.1, and d is the distance from the ellipse centroid to the point on the ellipse closest to the robot. Given the angle between the velocity vector of a moving pedestrian and the distance from the pedestrian to the robot, the distance from the ellipse centroid to the point on the ellipse edge closest to the robot can be calculated using the equation of an ellipse and a trigonometric identity:

$$\text{Ellipse : } \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad \text{Identity : } \tan(\theta) = \frac{y}{x} \quad (5.3)$$

Where a is the semi-major ellipse axis, b is the semi-minor ellipse axis, and θ is the angle between the velocity vector and distance vector. Performing simple rearranging using the

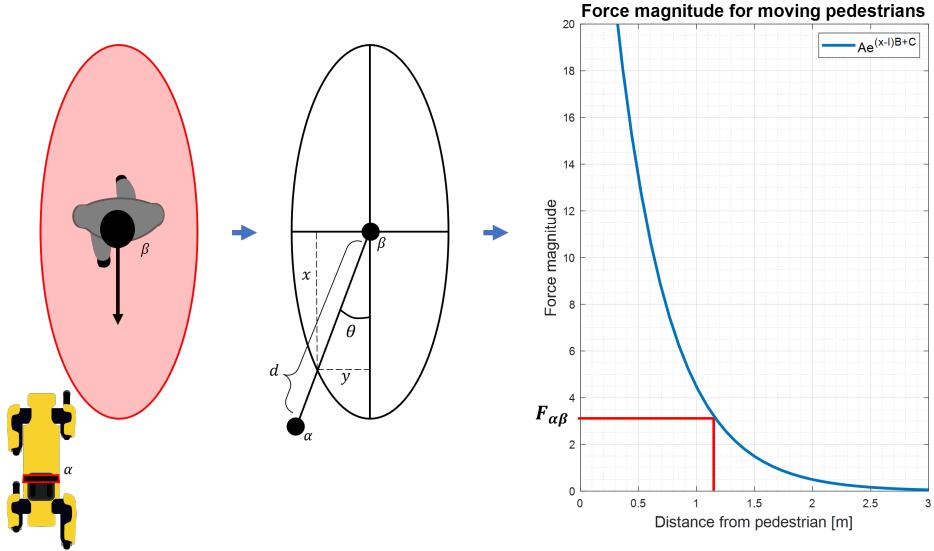


Figure 5.2: Repulsive force measured from a moving pedestrian

The robot calculates the repulsive force from a moving pedestrian from the angle of observation to the pedestrian's direction of motion. From this the distance from the pedestrian to the elliptical force field around it is calculated and used to give the correct magnitude of repulsive force.

substitution $y = x \tan(\theta)$ gives

$$x = \frac{ab}{\sqrt{a^2 \tan^2(\theta) + b^2}} \quad (5.4)$$

Where x now represents the distance along the x-axis from the ellipse centroid to the ellipse point closest to the robot. To find y , the 5.3 ellipse formula is used to isolate y . See figure 5.2 for a visualization of how the robot translates a moving pedestrian into a social repulsive force.

5.1.2 Side-stepping

The second augmentation developed in the ASFM, motivated by shortcoming 2, gives the robot a more natural and human-like response to repulsive forces that push it backwards by converting backwards motion to a side ways motion. This prevents the robot from moving backwards when encountering oncoming pedestrians. A behavioral answer to oncoming

pedestrians such as this, also tend to hold for humans as pedestrians usually do not to walk backwards when waiting for people to pass by them, but instead gets out of the way or simply just waits in place.

The policy of side-stepping is accomplished by prohibiting negative velocity in the x direction and by creating an enforcing effect on the velocity in the y-direction when velocity in the x direction turn negative. The ratio represented by a λ in figure 2 is a hyper parameter to be fine-tuned in testing. The ratio condition to enable the increase in sideways motion is set to a threshold of 2.5 to ensure that only when the backwards force component is 2.5 times greater than the sideways force component should this be enabled.

Algorithm 2 side-stepping

```

if  $x_{velocity}$  is backwards then
    ratio  $\leftarrow \frac{\mathbf{F}_{\alpha\beta}(x)}{\mathbf{F}_{\alpha\beta}(y)}$ 
    if ratio  $\geq 2.5:$ 
         $\mathbf{F}_{\alpha\beta}(y) = \mathbf{F}_{\alpha\beta}(x) * \lambda$ 
    else
        | ratio  $\leftarrow 0$ 
    end

```

If the robot encounters a static pedestrian the policy of stepping to the side or not walking backwards allows for the robot to get close to a pedestrian, without any awkward backwards motion to counteract the force field it violated ever so slightly because the deceleration of the robot cannot follow the instant change in velocity going from slightly positive to negative.

5.1.3 Follow Mode and Crowd Interpretation

The third augmentation developed for the ASFM is the most complex feature added. It allows for the robot to start following a person showing the same directional intention as the robot, after the robot has encountered any repulsive forces. This feature was developed on the motivation of SFM shortcoming 3 to overcome the freezing robot problem of stopping

or freezing up after having encountered a seemingly blocked path and then never progress beyond that point. Following a pedestrian through a crowded area is a very common behavior for humans to do. When pedestrians encounter crowded spaces that they need to navigate through, it can be postulated that humans do one of two things; force their way through the crowd or follow (at least partially) other pedestrians through the crowd walking in the same direction [87]. The robot system is not design friendly for behavior that would gently try to force its way through a crowd as it lacks any means of signaling its intention, but recognizing that another pedestrian's immediate goal coincides with the robot's is possible. Many social navigation solutions focus on leading robots, but it is a novel area to follow other humans with little research done on this area.

Recognizing a pedestrian as having the same immediate goal as the robot can be done by analysing their direction of motion. Once a pedestrian's direction of motion is determined, the angle between the robot system and the pedestrian direction can be calculated, and if it is within a certain range the robot can assume the pedestrian is walking in the same direction as the robot wants to go. For this project an estimated 30° deviation from pedestrian direction of motion to the robot's orientation was set as the threshold to follow someone.

This introduces some challenges as this analysis must run at all times, but there are scenarios where the robot should not follow anyone. For that reason, the follow mode can only start if the robot is experiencing any repulsive forces prior to seeing a pedestrian walk in the desired direction. Another condition that had to be included was the condition of only being able to follow the pedestrian closest to the robot in its field of view. This is necessary to avoid any intrusion of internal space within crowds or not being able to follow the pedestrian as they could disappear out of its vision when going behind someone.

When the follow behavior is initiated the repulsive forces from other pedestrians is lowered to make navigation in crowded and tight spaces possible. This builds on the fact that the personal space that the robot previously respected, hence why it possibly got stuck, can now be intruded upon to some extend as it is imitating the same behavior as the person leading the way. In general it is accepted behavior to get closer to humans

if the space is tight.

The way the robot starts moving towards a pedestrian it has deemed fit to follow, is by setting its objective goal point to the location of the pedestrian. This way the attractive force from the goal point is redirected to the location of the pedestrian. This also only works if the repulsive force from the pedestrian being followed is being ignored.

As described, when a pedestrian follows other pedestrians through a crowded area, they stop following anyone as soon as there is no need for it anymore. Therefore the robot must also be able to judge when to stop following the pedestrian they have started following. The policy implemented for the follow mode set a timer of five seconds to start counting when the robot is following someone and they no longer experience any external repulsive forces - the area is no longer crowded. After five seconds of no external repulsive force experienced, the robot will stop following the pedestrian. See figure 5.3 for a visualization of the decision process the robot makes to follow someone.

When the robot stops following a pedestrian, the repulsive force that was suppressed while following the pedestrian, is brought back into operation. In order not to trigger the follow mode again from the same pedestrian the robot has to log which pedestrians have acted as leaders before. That way, even when all conditions are met for a pedestrian, if the robot has followed that pedestrian before, it will not trigger the follow mode.

Finally if the pedestrian the robot is following goes more than 4 meters away from the robot, the robot will stop following the pedestrian.

5.1.4 Redefining Direction Priority

The SFM's shortcoming 4 was its approach of lowering repulsive forces using the direction dependent weighting function made for a poor practical solution. The less complicated augmentation proposed for the ASFM to get around the weight function, was by annulling any repulsive force coming from pedestrians that are moving away from the robot. This again requires the robot to log the position of detected humans and compare their latest logged position to their current position in order to gauge if the pedestrian is moving away from the robot. This approach is very useful for a first person field of view robot as it does

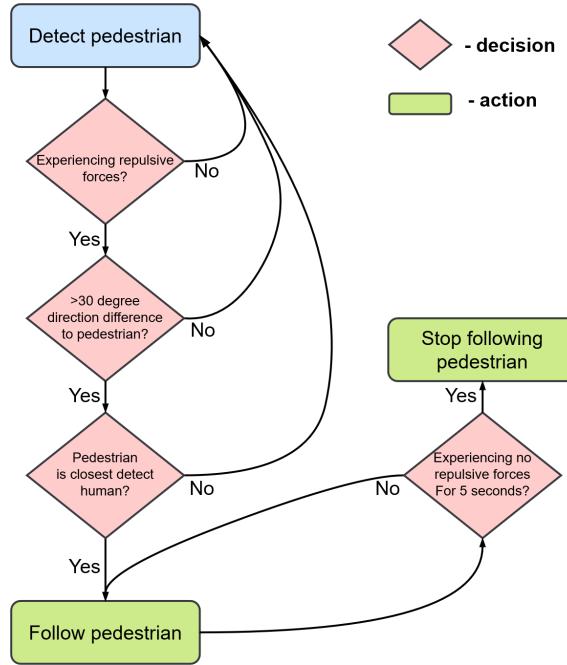


Figure 5.3: ASFM follow mode conditions

The ASFM can only follow a pedestrian if three different conditions are met; are there other repulsive forces present apart from the pedestrian considered? Does the pedestrian have a direction of motion that coincides with the direction of motion of the robots desired path? And is the considered pedestrian the closest pedestrian to the robot? If a pedestrian clears all three conditions the model deems the pedestrian worth following as it will likely guide the robot along a safe path with a smaller chance for social obstruction.

not need to take into consideration the movement of the robot. If the robot is moving and a person is going further away from the robot from one instance to another, then it means the pedestrian is walking faster than the robot and it is a valid assumption to ignore the repulsive forces from the pedestrian.

5.1.5 Human Friendly Velocity Control

With the robot system and camera, the SFM model would often experience the max robot velocity to be too fast for the latency in deceleration when interacting with other pedestrians. Therefore the last proposed augmentation to the ASFM is a reduction in max velocity based on how far away from the robot the closest detected pedestrian is. If the robot is further than 4m away from the nearest pedestrian the maximum velocity allowed

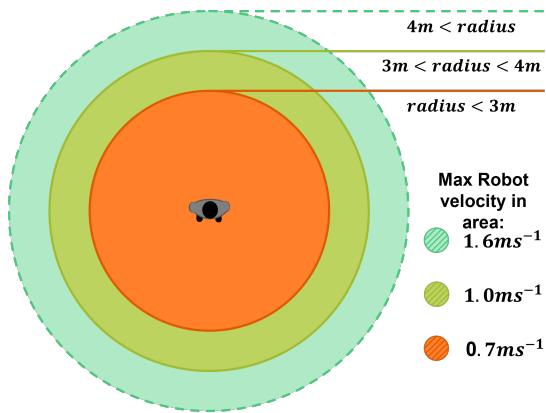


Figure 5.4: the maximum forward velocity allowed depending on how far away the closest pedestrian is.

In order to make the mobile robots motion more interaction friendly for pedestrians and to counteract the relatively slow frame rate of the camera, the robot's maximum velocity is capped relative to the distance from the nearest pedestrian detected.

is $1.6ms^{-1}$ (Spot's max velocity). If the nearest pedestrian is within $3m$ to $4m$ of the robot, then the maximum velocity is capped at $1.0ms^{-1}$. And if the nearest pedestrian is closer than $3m$, then the max velocity is capped at $0.7ms^{-1}$. This max velocity system also creates a more pedestrian friendly behavior when approaching a pedestrian head on. An illustration of the max velocity zone system can be seen in figure 5.4.

5.2 Augmented Social Force Model Implementation

The implementation of the ASFM is more extensive than the SFM implementation so will be divided into several algorithms.

5.2.1 Logging Pedestrian Information

To enable three of the proposed augmentations in the ASFM to work, prior knowledge of the current detected pedestrians are needed. The previous distance from pedestrian to robot, if the pedestrian is coming closer to the robot or not, and if a pedestrian has already been followed. The Zed2i camera model puts a label ID on all humans it detects so the matrix storing all this information can be structured using the different detected

pedestrian ID labels. To log this data a matrix can be created for every camera data input. This matrix will take the form as the example matrix shown in table 5.1. The matrix is updated after every camera input, where the old distances compared to the new distances to see if they are moving closer to the robot (indicated by a -1 under the *To_or_away* column) or have been followed before (marked as 1 under the *Past_leader* column), before they are replaced.

Pedestrian ID	Distance	To_or_away	Past_leader
13	2.1	0	1
14	4.23	-1	0
18	5.6	0	0
...

Table 5.1: Matrix format of the log that the model keeps for every currently detected pedestrian.

The repulsive force implementation of the ASFM can thus be written as seen in algorithm 3

Algorithm 3 Augmented Social Repulsive Force

Constant parameters: Γ, Λ, σ **Function** Social_Force(*Obj_det*):

```

for j in number of pedestrians do
| Create pedestrian data matrix
end

for i in number of pedestrians do
| if pedestrian[i] is already followed then
| | force = 0
| else
| | if pedestrian[i] coming closer then
| | | if pedestrian[i] not moving then
| | | | force =  $Ae^{x*B+C}$  // where x is distance to pedestrian
| | | else
| | | | if criteria to follow pedestrian[i] apply then
| | | | | leader  $\leftarrow$  pedestrianID
| | | | else
| | | | | v  $\leftarrow$  pedestrian velocity, a  $\leftarrow \Gamma\|\mathbf{v}\|$ , b  $\leftarrow \Lambda\|\mathbf{v}\|$ 
| | | | | theta  $\leftarrow$  angle // from pedestrian direction to robot
| | | | | orientation
| | | | |  $x = \frac{ab}{\sqrt{a^2 \tan^2(\theta) + b^2}}$   $\rightarrow$  y =  $\sqrt{b^2(1 - \frac{x^2}{a^2})}$ 
| | | | | rho  $\leftarrow \sqrt{x^2 + y^2}$ 
| | | | | force =  $Ae^{(x-\rho+\sigma)B+C}$ 
| | | | end
| | | end
| | end
| | else
| | | force = 0
| | end
| end

if pedestrian[i] is leader then force = 0
repulsive_force = repulsive_force + force
end

return repulsive_force

```

The full implementation using the new features can be seen in algorithm 4.

Algorithm 4 Augmented Social force Model main function

Parameters: $v_{desired}$, τ , $v_\beta \Delta t$

Function Attractive_Force(*goal*, *position*, \mathbf{v}_α , $\hat{\mathbf{e}}$):

$$\mid \mathbf{F}_\alpha = \frac{1}{\tau}(v_{desired}\hat{\mathbf{e}} - \mathbf{v}_\alpha)$$

return \mathbf{F}_α

Function Main(\mathbf{F}_α , $\mathbf{F}_{\alpha\beta_i}$):

while receiving *robot kinematics & humans detected* **do**

position \leftarrow odometry from *robot kinematics*

if *robot is following a pedestrian* **then**

goal \leftarrow location of pedestrian

else

goal \leftarrow predetermined path checkpoint

end

$\hat{\mathbf{e}} \leftarrow$ unit vector from robot to goal

$\mathbf{F}_\alpha \leftarrow$ Attractive_Force(*goal*, *position*, \mathbf{v}_α)

$\mathbf{F}_{\alpha\beta} \leftarrow$ Social_Force(*detected_humans*)

t_{now} \leftarrow time of instance

$\Delta t = t_{now} - t_{past}$

$\mathbf{v}_{new} \leftarrow \mathbf{v}_{current} + (\mathbf{F}_\alpha - \mathbf{F}_{\alpha\beta})\Delta t$

$\mathbf{v}_{alt} \leftarrow$ Side_Stepping(\mathbf{v}_{new})

$\mathbf{v}_{alt} \leftarrow$ Velocity_Control(\mathbf{v}_{alt})

Publish \mathbf{v}_{alt} to robot

$\mathbf{v}_{current} = \mathbf{v}_{alt}$,

$t_{past} \leftarrow t_{now}$

end

end

Chapter 6

Experiments and Results

6.1 Evaluation of Model Behavior

A crucial step during the development of a social navigation mode is the evaluation of the model. To test and compare SFM implementation and ASFM implementation, the mobile robot was tasked to perform social navigation in real-life scenarios. Five test scenarios were conducted, four planned scenarios each with a different challenge, and one unplanned test scenario.

The experiments performed to test social navigation took place in a narrow 2.1-meter-wide hallway. The robot started at a predetermined spot in the middle of a hallway and has a set final destination 8.5 meters down the hallway. The five testing scenarios were displayed between the start and end point. The setting of a hallway presents many restrictions on a mobile robot that must maneuver around people within a limited space. It is a more difficult environment to navigate compared to a wide road or sidewalk that is less restrictive in nature.

The metrics introduced outlined in chapter 2 are used to evaluate the performance of the robot's social navigation in all test scenarios. For each test scenario the SFM and ASFM implementations were assessed over 20 trials and compared against each other and real-life human behavior.

To evaluate path efficiency and irregularity, trajectory length and average veloc-

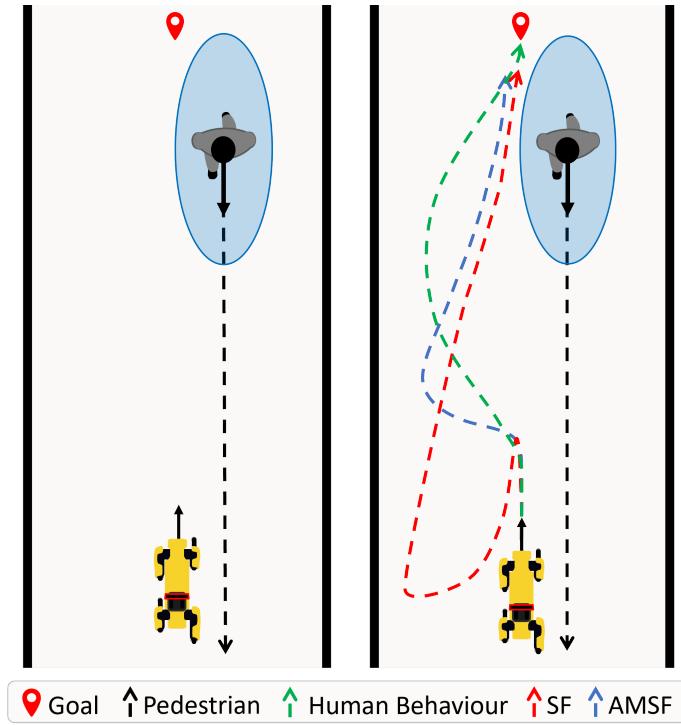


Figure 6.1: Test scenario 1.

For the given scenario, The ASFM compared to the SFM had a more natural behavior of avoiding the oncoming pedestrian, resembling more a humans natural response.

ity were measured. Social comfort and behavioral naturalness are given as an overall evaluation of the robots movement.

6.1.1 Test Scenario 1

Test scenario one presents a pedestrian walking towards the robot. Naturally, the pedestrian keeps to one side of the hallway. The expected human-like reaction in this scenario is to shift away from the oncoming person and move towards the other side of the hallway without slowing down drastically. The purpose of this scenario is to demonstrate the effectiveness of the implemented models when dealing with oncoming pedestrians.

For scenario one, both model implementations arrived at the goal point with a 100% arrival rate. Both models achieved a rate of zero percent, which was expected for the simple scenario. Interestingly, differences in the performance of the two models

were apparent in metrics time to goal, path efficiency by distance travelled, and path irregularity. A visualization of the test scenario along with the average movement pattern of the SFM and the ASFM implementations are illustrated in figure 6.1.

When facing an oncoming pedestrian, the SFM moved in the same direction of the pedestrian and backtracked substantially before the pedestrian came so close that most of the force was pushing it to the left. At times the robot got close enough to the pedestrian while passing by for it to be uncomfortable. On the other hand, the ASFM with its side-stepping behavior was able to step out of the way as soon as it encountered a repulsive force, which resulted in a smoother reaction.

The discrepancies in time to goal, trajectory length, and average velocity between the ASFM and SFM (figure 6.2) highlights the increased path efficiency and a decreased path irregularity in the ASFM compared to the SFM. The behavioral naturalness of the ASFM was also improved compared to the staggering behavior of the SFM. As a result, the social comfort of the ASFM in this scenario was also higher than the SFM.

From figure 6.2 it is clear that on average ASFM cleared the test scenario faster.

The SFM implementation showed a severe path irregularity compared to the human trajectory as it was pushed backwards given the angle it met the oncoming pedestrian at. The ASFM on the other hand demonstrated improved behavior by moving to the side while the pedestrian was walking past the robot. A weakness of both models in this scenario was that the robot attempted to return to its initial trajectory towards the goal and did not move forward at the same rate when being pushed to the side by the oncoming pedestrian.

Since the ASFM demonstrated higher efficiency and minimal path irregularity compared to the SFM, this scenario provides evidence for the superior performance following augmentations in the proposed model for social navigation.

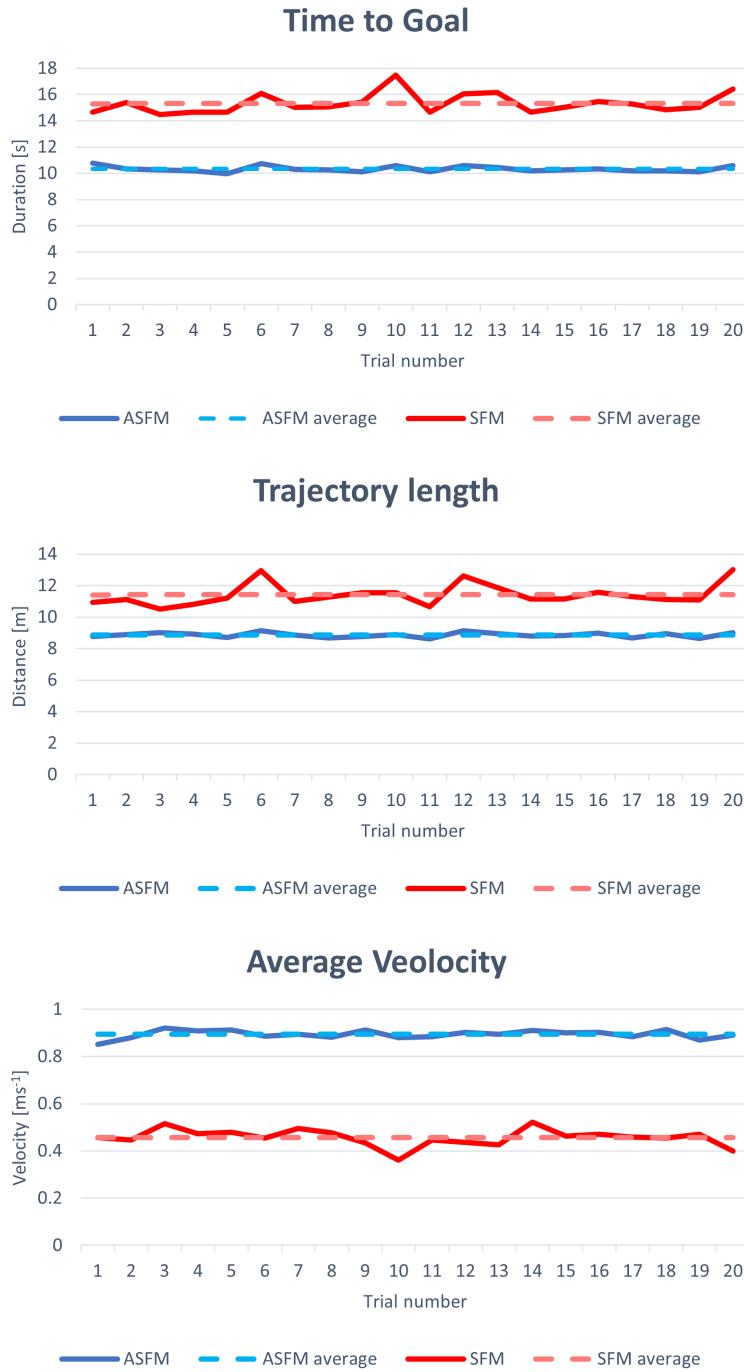


Figure 6.2: Scenario 1 results.

The ASFM proved more efficient with a faster time to goal, a shorter trajectory path, and a higher average velocity.

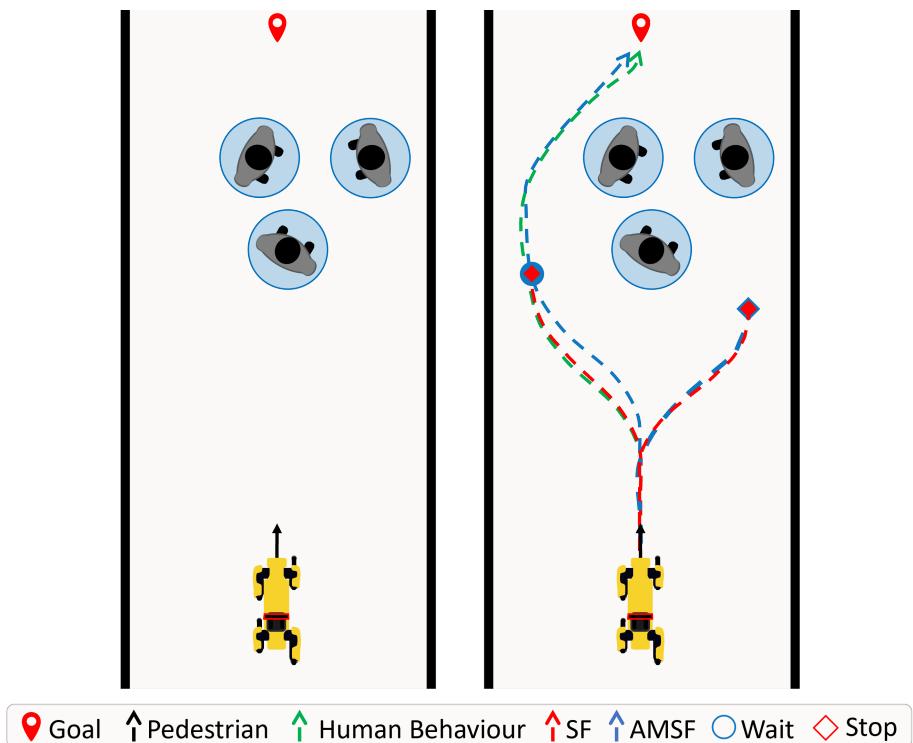


Figure 6.3: Test scenario 2

For the given crowd partially blocking the hallway, both ASFM and SFM were unsuccessful at times failing to interpret the crowd formation. However, the ASFM proved better at maneuvering the narrow space when choosing the right path around the crowd.

6.1.2 Test Scenario 2

In test scenario two, the robot encountered a static crowd of three people. A pedestrian encountering this scenario in a hallway would walk past the three pedestrians with ease by choosing a path with the largest opening. Pedestrians in a crowd formation as presented in this scenario tend not to move for other bypassing pedestrian as they deem the space on the side of the crowd to be large enough. This scenario was designed to test the models on how they interact and adapt to a crowd that is blocking most of the passage way. figure 6.3 shows a visualization of the initial scenario and the average trajectory paths of the models.

Over the 20 test trials conducted with this scenario setup, both models never experienced any collisions, however, the arrival rate for the SFM was at 40% and the ASFM was at 75%. This drop in the ability to reach the end point was due to the inability to interpret this specific type of crowd structure in both models. Both models would have a chance of encountering the closest pedestrians repulsive force field on slightly the right or left side making the robot shift towards the side of impact. If the robot hit the right side the SFM would end up getting stuck never reaching the goal. If the ASFM hit the right side it would try to get as close as possible and then after a three second wait it would backtrack and still be stuck or be pushed to the other side by the cumulative forces of the pedestrians. In trials where the robot hit the left side, the SFM model would be pushed backwards to where it could not get through the gap efficiently. The ASFM also felt the tight space, but would not be pushed backwards and would slowly make its way through the narrow gap.

At times both models lost sight of the people in the crowd and would make its way towards the goal with ease as the restricting repulsive forces would disappear.

Despite the lack of data from the failed trials, the general trends as seen in figure 6.4 show very similar behavior in this scenario. This was expected as there were only static pedestrians and the reaction to radial repulsive forces were similar in both model implementations. It is thus almost impossible to show model supremacy in this scenario. Given the number of failures in this scenario neither model could be deemed 100% successful at

performing social navigation.



Figure 6.4: Time to goal, SFM vs. ASFM

Both the SFM and the ASFM failed at times as seen in the missing entries of the collected results. For the few trials the SFM succeeded the ASFM still proved more efficient with a slightly faster time to goal average, an overall shorter trajectory path average, and a higher average velocity.

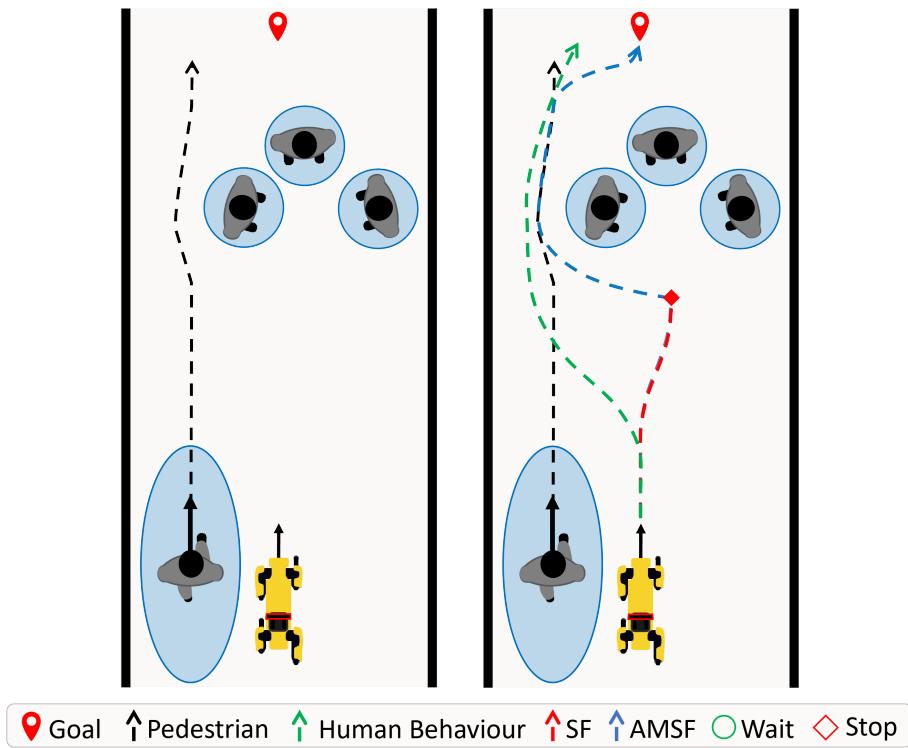


Figure 6.5: Test scenario 3

When considering the path of the pedestrian as shown on the right image, only the ASFM was able to maneuver through the test scenario utilizing its follow ability.

6.1.3 Test Scenario 3

Test scenario three was more complex than the previous two as it was designed to test if the model could follow a pedestrian through an otherwise blocked passage way. The robot encountered a crowd of pedestrians that took up enough space in the hallway and made it impossible for the robot to pass by due to the presence of colliding force fields. Test scenario three start structure and average paths for the two models through the scenario are illustrated in figure 6.5.

Over the 20 conducted test trials for scenario three, the SFM had an arrival rate of 0% as expected since the model could not distinguish or interpret the pedestrian's intention. The ASFM had an arrival rate of 85% since it was able to successfully follow the person through the crowd in the majority of the trials. Neither model caused any collisions.

One explanation for the ASFM's failure at times was faulty human detection with the walking pedestrian not being detected while it showed the same directional intention as the robot. Another explanation for failed attempt to follow was if the pedestrian was not recognized fast enough and the pedestrian no longer had the correct motion direction.

Apart from the camera's faults in the ASFM, the augmentation responsible for executing the follow maneuver was successful and proven useful. The success of social navigation did however lower from this maneuver despite the robot reaching the goal point. The naturalness of a robot not being able to get through a crowd formation waiting in front of the blocking crowd is not natural behavior. On the contrary if the pathway is completely blocked then humans do tend to wait for the pathway to open up without forcing their way through.

In this test scenario it has been shown that the ASFM is indeed capable of navigating through dense crowd using another pedestrian's help for guidance. This gives the ASFM a crucial advantage over the SFM as following other people is a common social navigation behavior.

6.1.4 Test Scenario 4

In test scenario four, the robot encountered two walking pedestrians, an unseen pedestrian crossing the robot's path, and another pedestrian who was already in sight moving slower than the robot. Test scenario four was designed in such a way to test the robot's reaction to both an oncoming pedestrian and someone walking slower than the robot. The reason for having a pedestrian cut across the hallway instead of coming directly towards the robot was to test what would happen if the robot was encountering an unseen pedestrian. The reason for having a slow-moving pedestrian was to see how the robot would overtake a pedestrian.

Across the 20 test trials conducted for this scenario both models showed 100% arrival rate. This scenario also did not cause any collisions despite the unseen pedestrian suddenly cutting across the hallway. The SFM found it difficult to evade in the correct direction and was be pushed back in parallel to the oncoming pedestrian, resulting in

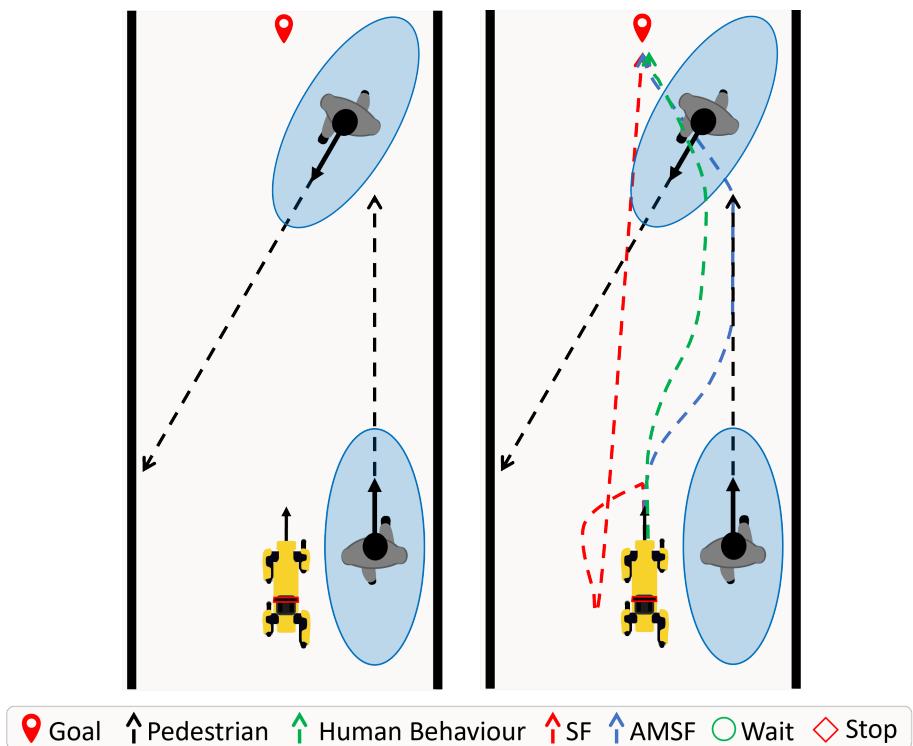


Figure 6.6: Test scenario 4

Encountering the moving pedestrians as seen in this scenario (right image) the movement of the ASFM were more natural than the SFM.

less socially comfortable navigation. The SFM also struggled with interpreting the slow walking pedestrian as an obstacle to avoid and would get too close to the slow pedestrian as it walked by. The ASFM performed better by evading away from the pedestrian and in a direction closer to the opposite direction of the pedestrian's motion. The ASFM model was also able to accurately avoid the slow walking pedestrian as it slowed down correctly interpreting the repulsive force exerted by the pedestrian.

The data recorded for this scenario, seen in figure 6.7 are less indicative of the social navigation success of the two models. The SFM had a lower time to goal, a higher trajectory length and a higher average velocity than the ASFM. For this scenario it means that the robot's social navigation was not comfortable for the pedestrians, since its movements were faster and more abrupt. At times the ASFM would start following the slow walking pedestrian when the other pedestrian cut across, this is to some extend mimics real-life human behavior shown in figure 6.5. This provides an explanation for the large variance in the ASFM's time to goal data since when the robot followed the slow pedestrian it would conform to the same speed.

6.1.5 Scenario 5

Evaluation of the four previous scenarios were designed to target specific aspects of the chosen social navigation models. A fifth pseudo test scenario was conducted to address the limits of test scenario 1-4. The fifth scenario has the robot face pedestrians who have not been given any behavioral guidelines, to present an element of surprise to the robot. This scenario was designed to illustrate how people would interact around the robot and what complications would arise from people exhibiting less pedestrian like behavior.

When pedestrians were given the freedom of motion in front of the robot, the robot often struggled with movements that were too quick for it to interpret. This led to pedestrians having to abruptly stop due to a delayed reaction from the robot.

This test scenario, highlights the challenging nature of ensuring ordinary HRI despite the efforts made towards perfecting the ASFM.



Figure 6.7: Time to goal, SFM vs. ASFM

The data collected in test scenario 4 gave little insight into how the scenario played out with all three metric averages being similar for both models.

6.1.6 Hyper Parameter Tuning

Prior to evaluating the execution of the SFM and the ASFM in the five test scenarios, the hyper parameters of the repulsive forces were fine-tuned to accurately recapitulate human behavior. For the ASFM repulsive force, the equation as described in equation 6.4 of the force field can be written with its modified hyper parameters as

$$F_{\alpha\beta} = 0.04e^{-2.2(x-d)+6}. \quad (6.1)$$

This enabled the robot to be 0.5 meters within range of a static pedestrian before being repelled away. The shape of the exponential curve increases the force value if the robot comes too close. When the robot is following a pedestrian it should decrease its repulsive forces to make crowd navigation easier. The modified repulsive force for a following robot is

$$F_{\alpha\beta} = 0.01e^{-4(x-d)+3} \quad (6.2)$$

The parameters A, B, C are congruent for static and dynamic pedestrians. To elicit an appropriate increase in the elliptical force field for the ASFM, the semi-major and semi-minor axes grew at a rate proportional to the pedestrians velocity. The rate of growth that resulted in the best performance was

$$a = 1.5 + 2||velocity|| \quad (6.3)$$

$$b = 1 + \frac{||velocity||}{10}. \quad (6.4)$$

The modified force fields of a static and moving pedestrian with a speed of $2ms^{-1}$ have been visualized in 3D in figures 6.9 and 6.8. The hyper parameter λ , responsible for amplifying sideways motion in augmenting side stepping was set to 1.8.

Hyper parameter tuning of the SFM used the same numbers for A and B in 4.3 as proposed by Helbing et al. in [88] with the repulsive potential as in equation 4.3 be

$$V_{\alpha\beta}(b) = 0.12e^{-b_{\alpha\beta}/3.9}. \quad (6.5)$$

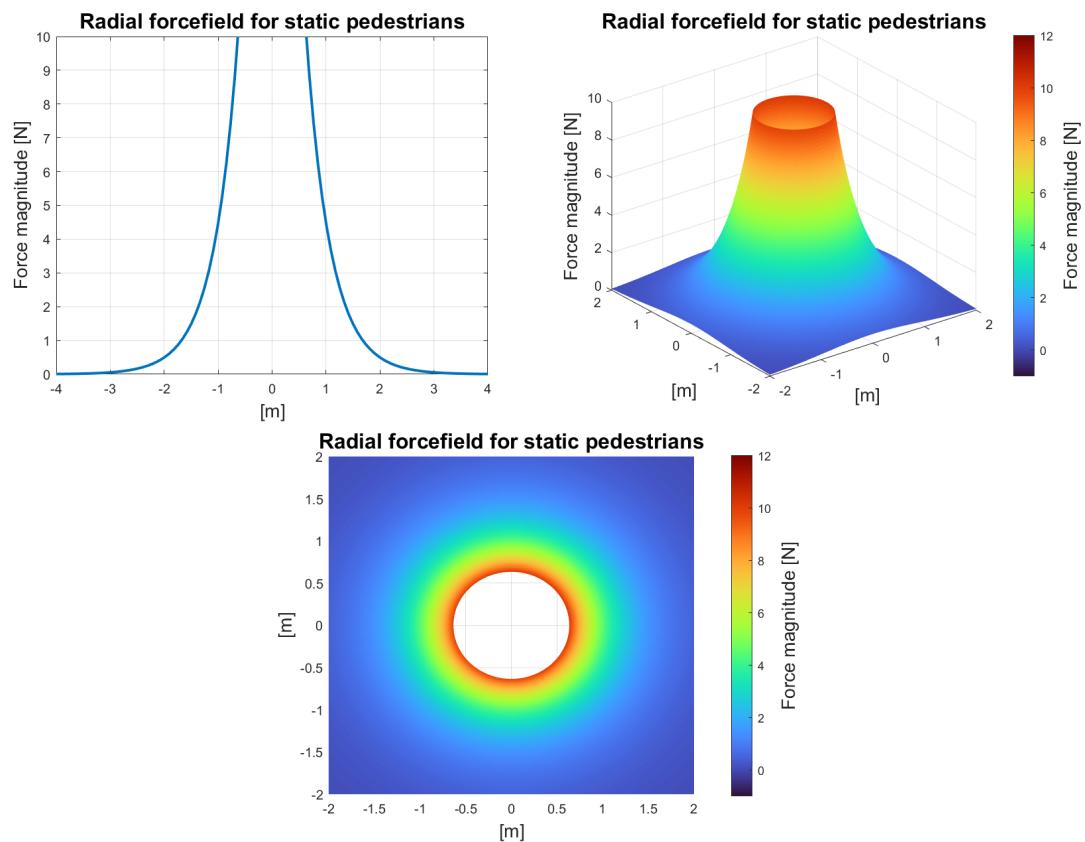


Figure 6.8: The radial force field visualized

The plots show the repulsive force fields from a static pedestrian with the magnitude and shape of the final hyper parameter tuned exponential function.

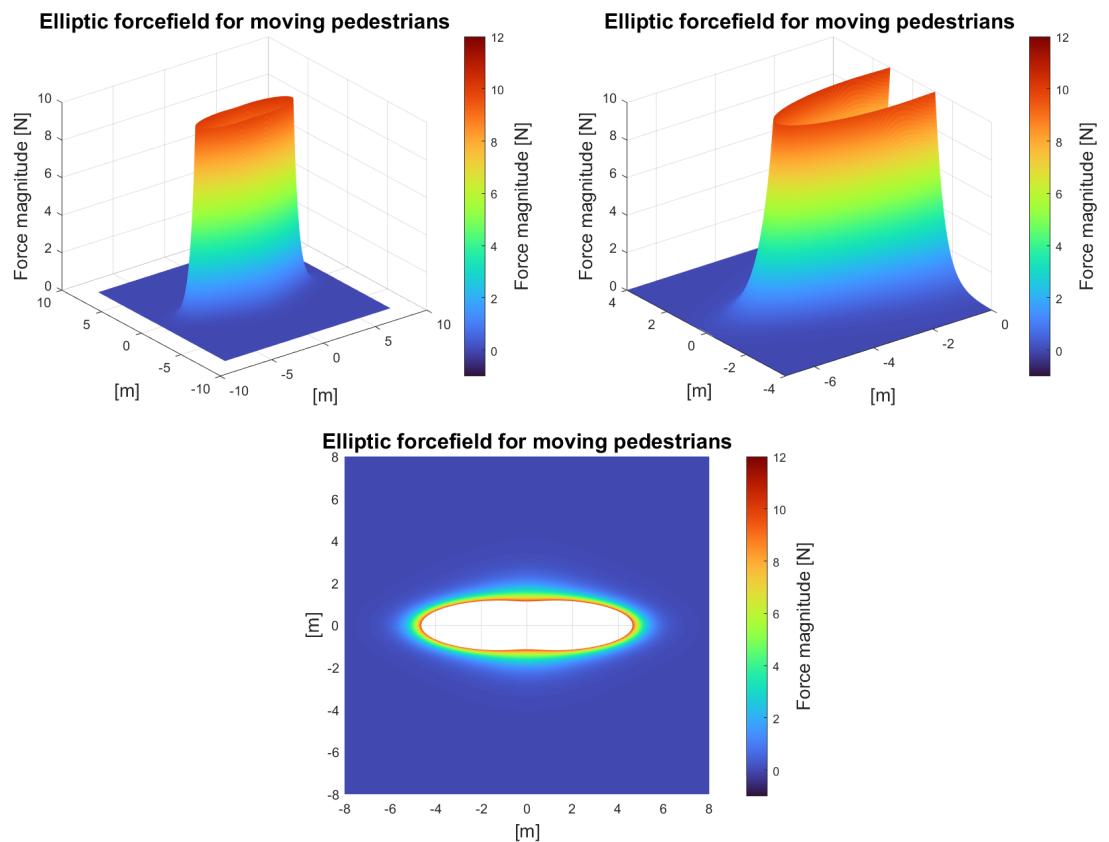


Figure 6.9: The elliptic force field visualized

the plots show the repulsive force field from a moving pedestrian with a speed of 2 ms^{-1} as modelled by the final hyper parameter tuning of the exponential function.

6.2 Discussion

6.2.1 Limitations

A limitation in the robot hardware that became apparent during scenario testing instability of the human detection model. The camera would lose sight of a person if the person was not fully visible or obstructed by an occluding pedestrian. This flaw was not a direct result of the model design since the human detection was developed prior to this project, but it emphasizes the importance of the FOV required for quality visual perception on a mobile robot performing social navigation.

The creation of a compliant model for social navigation has proven to be incredibly complex. The proposed augmentations in the ASFM addressed the most abrupt behavior flaws in the SFM. However, the number of social norms and subliminal signals in social interactions amongst pedestrians are difficult to generalize using a single model based on attractive and repulsive forces.

Another limitation in this project was the choice of model architecture. The SFM as a descriptive model on pedestrian dynamics was chosen for its generalized approach to pedestrian behavior, but other models such as the VO model or models based on IRL were not fully explored.

The design of the four test scenarios shed light on the two model's behavioral characteristics and weaknesses, but also exhibited design limitations. By using humans as test subjects, a limitation in the conformity of the test subject's movement between each trial is unavoidable. Variation in a person's walking speed between each trial diminishes the validity of the recorded data. As seen in the fifth pseudo scenario, the naturalness of the model around people in a scenario without structure was difficult to accurately improve since identifying what it should do for every scenario was almost impossible.

6.2.2 Comments On Scenario Evaluation

Scenario two proved to be the best scenario to tune the hyper parameters that control the strength of the repulsive forces as it clearly presents the limits of a repulsive force field

strength with regards to if the pedestrian could pass by or not.

Scenario three sheds light on an important limitation with the robot's camera system. Since the follow mode relies on detecting people in close proximity, the third scenario showed that if a person was cutting into the camera POV too quickly before its trajectory deviates, the camera would not accurately detect the person in time. This was an important observation to understand the allowed angle to follow someone. The model would improve if the camera angle was widened, but this would compromise the following function in other scenarios, making the intention inference less precise.

The reason for not introducing a wait and subsequent decrease in repulsive forces was to maintain the space within closed crowds, which was off limits according to literature, mentioned in chapter 2.

To better understand the workings of pedestrian behavior the project evaluation could have been more objective by including more test scenarios to possibly shine light on undiscovered weaknesses of the social navigation models.

6.2.3 Comfort

Assessing peoples' comfort was challenging in each test scenario as test subjects grew to become more comfortable with the robot. The test subjects would become accustomed to the robot's behavior and this paralyzed their reaction as they understood that the robot would never hit them. Test subjects would even express abnormally high comfort towards the robot as the interest in the robot outweighed the uncertainty it posed.

Another difficulty of measuring pedestrian comfort towards the robot system used in this project is the subconscious nature of social navigation in humans. The lack of human expression in a robot represents a loss of social cues that humans need to perform social navigation. The lost social cues make the human unsure of what the robots intentions are. As such a justification for what the correct behavior for a robot to exhibit in a given scenarios could only be an educated postulation.

6.2.4 Social Force Model in Real Life

The major take-away from the evaluation of both models was the limitations of the SFM. Despite the immense effort spent on tailoring the model to fit the robot system, the SFM is by nature a descriptive model for pedestrian dynamics in scenarios where every agent follows the same motivation. As soon as human nature goes beyond the general motivations of a pedestrian as proposed in the SFM, the model no longer works. This was demonstrated in the implementation of the SFM as the robot did not receive the same cooperative responses from encountered pedestrians as if the pedestrians also operated on the principles of attractive and repulsive forces.

The model is also highly susceptible to limits in environment knowledge. The camera setup with its limited vision hinders the SFM without any of memory ability that would enable the model to remember the position of a lost pedestrian for future reference. This would keep the robot from getting too close to pedestrians lost out of sight yet still very close. The camera setup also does not know if the person detected has been detected before. When a pedestrian is out of the camera's view, the model would treat them as a new person when they are detected again at a later time. A memory function would potentially solve this as well by enabling it to figure out that where the lost pedestrian stood, a new pedestrian now stands within a short enough time duration so must be the same person.

Chapter 7

Conclusion and Future work

7.1 Conclusion

Social navigation is navigation through space under the influence of human activities within that space. This project successfully enabled a mobile robot to perform social navigation and achieved the goal of maneuvering through a pedestrian environment comfortably and safely for pedestrians. This was accomplished by understanding the necessary factors involved in social navigation such as path planning, necessary pedestrian proxemics, model evaluation, and determining an approach to implement the navigational model in mobile robots.

The well-established SFM based on attractive and repulsive force fields was the chosen approach to describe pedestrian dynamics. The SFM was successfully implemented on the Spot robot system. The robot system consisted of a quadruped mobile robot, equipped with a stereo camera and an edge GPU which performed human pose detection for social navigation inference. The behavior of the SFM was analyzed to identify the limits of its navigational capabilities. The shortcomings of the SFM include unawareness of surrounding pedestrian formation, inefficient movement when interacting with pedestrians, and limited sensory input due to its incompatibility with the robot.

To improve upon the SFM, an augmented navigational model was designed to overcome its limitations. To increase efficiency of movement, this project designed augmenta-

tions that redefined the repulsive forces, the method for direction priority, and pedestrian path avoidance, as well as giving the model the ability to follow a pedestrian to guide it through a dense crowd.

This study demonstrates the successful implementation of the ASFM on the Spot robot system. Both the SFM and ASFM were tested, and their performance was compared in four different scenarios set in the environment of a hallway to test the robot's interaction with dynamic pedestrians and behavior in crowd formations. Both model implementations demonstrated the ability to navigate through a social environment, but the comfort and success rate of the ASFM were higher as the robot displayed behavior similar to an actual pedestrian.

The adaptation of the SFM along with the augmentations developed in this project enables a quadruped robot to safely navigate through public environments while ensuring the comfort of pedestrians.

7.2 Future Work

Better Crowd Interpretation. As mentioned in the limitations, the ASFM generated in this project was not able to interpret crowd formation efficiently, thus making the model unable to identify the optimal path around a crowd. A possible solution to this problem would be to use the orientation of crowds to group certain pedestrians together and create a common repulsive force field for the crowd as a whole.

Memorizing Past Pedestrian Position. Another improvement that would improve the social navigation capabilities of any model implemented on the robot system, would be to incorporate a memory function. The model would remember the position of pedestrians when they leave the camera's FOV and continue to assume their position within close proximity to the robot. This would minimize the behavior of getting too close to pedestrians around the robot.

An Additional Rear-View Camera. To further strengthen the social navigation capabilities of the robot system, a second camera could be mounted to face the opposite direction of the robot’s motion. This would allow the robot to visualize and react to the dynamic environment behind itself and prevent getting in the way of bypassing pedestrians.

Display of confusion when stuck. A feature that would improve the social navigation of the robot system would be to incorporate an ability to signal the humans that it is stuck. Spot’s body can horizontally rotate clockwise and counter-clockwise. A movement like this could be seen as the robot being confused to why it is not able to move forward and signal the surrounding pedestrians that they should give space for the robot.

Bibliography

- [1] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, “Minerva: a second-generation museum tour-guide robot,” in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, vol. 3, 1999, pp. 1999–2005 vol.3.
- [2] D. F. D. H. G. L. D. S. W. S. S. T. Wolfram Burgard, Armin B. Cremers, “The interactive museum tour-guide robot,” in *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1998, pp. 11–18.
- [3] D. Vasquez, B. Okal, and K. O. Arras, “Inverse reinforcement learning algorithms and features for robot navigation in crowds: An experimental comparison,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 1341–1346.
- [4] C. I. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman, A. Steinfeld, and J. Oh, “Core challenges of social robot navigation: A survey,” *CoRR*, vol. abs/2103.05668, 2021. [Online]. Available: <https://arxiv.org/abs/2103.05668>
- [5] E. Bassoli and L. Vincenzi, “Parameter calibration of a social force model for the crowd-induced vibrations of footbridges,” *Frontiers in Built Environment*, vol. 7, 2021. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fbuil.2021.656799>
- [6] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, “Human-aware robot navigation: A survey,” *Robotics and Autonomous Systems*, vol. 61, no. 12,

- pp. 1726–1743, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889013001048>
- [7] D. Dugas, J. Nieto, R. Siegwart, and J. J. Chung, “Ian: Multi-behavior navigation planning for robots in real, crowded environments,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 11 368–11 375.
- [8] F. Zanlungo, T. Ikeda, and T. Kanda, “Social force model with explicit collision prediction,” *EPL (Europhysics Letters)*, vol. 93, no. 6, p. 68005, mar 2011. [Online]. Available: <https://doi.org/10.1209/0295-5075/93/68005>
- [9] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, “Planning-based prediction for pedestrians,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 3931–3936.
- [10] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, “Learning motion patterns of people for compliant robot motion,” *The International Journal of Robotics Research*, vol. 24, no. 1, pp. 31–48, 2005.
- [11] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, “A bayesian nonparametric approach to modeling motion patterns,” *Autonomous Robots*, vol. 31, no. 4, pp. 383–400, 2011.
- [12] K. Li, M. Shan, K. Narula, S. Worrall, and E. Nebot, “Socially aware crowd navigation with multimodal pedestrian trajectory prediction for autonomous vehicles,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.
- [13] P. Trautman, J. Ma, R. M. Murray, and A. Krause, “Robot navigation in dense human crowds: Statistical models and experimental studies of human–robot cooperation,” *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 335–356, 2015.

- [14] B. Kim and J. Pineau, “Socially adaptive path planning in human environments using inverse reinforcement learning,” *International Journal of Social Robotics*, vol. 8, no. 1, pp. 51–66, 2016.
- [15] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, “Socially compliant mobile robot navigation via inverse reinforcement learning,” *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1289–1307, 2016.
- [16] D. Vasquez, B. Okal, and K. O. Arras, “Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 1341–1346.
- [17] G. F. Cooper, “The computational complexity of probabilistic inference using bayesian belief networks,” *Artificial Intelligence*, vol. 42, no. 2, pp. 393–405, 1990. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/000437029090060D>
- [18] C. I. Mavrogiannis, V. Blukis, and R. A. Knepper, “Socially competent navigation planning by deep learning of multi-agent path topologies,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 6817–6824.
- [19] E. Stein, J. S. Birman, J. N. Mather *et al.*, *Braids, Links, and Mapping Class Groups*. Princeton University Press, 1974, no. 82.
- [20] C. I. Mavrogiannis and R. A. Knepper, “Multi-agent path topology in support of socially competent navigation planning,” *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 338–356, 2019.
- [21] M. A. Berger, “Topological invariants in braid theory,” *Letters in Mathematical Physics*, vol. 55, no. 3, pp. 181–192, 2001.
- [22] C. I. Mavrogiannis, W. B. Thomason, and R. A. Knepper, “Social momentum: A framework for legible navigation in dynamic multi-agent environments,” in *Proceed-*

- ings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 361–369.
- [23] C. I. Mavrogiannis and R. A. Knepper, “Multi-agent trajectory prediction and generation with topological invariants enforced by hamiltonian dynamics,” in *International Workshop on the Algorithmic Foundations of Robotics*. Springer, 2018, pp. 744–761.
- [24] A. Turnwald and D. Wollherr, “Human-like motion planning based on game theoretic decision making,” *International Journal of Social Robotics*, vol. 11, no. 1, pp. 151–170, 2019.
- [25] S. Russell, “Learning agents for uncertain environments,” in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 101–103.
- [26] P. Henry, C. Vollmer, B. Ferris, and D. Fox, “Learning to navigate through crowded environments,” in *2010 IEEE international conference on robotics and automation*. IEEE, 2010, pp. 981–986.
- [27] Y. F. Chen, M. Everett, M. Liu, and J. P. How, “Socially aware motion planning with deep reinforcement learning,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1343–1350.
- [28] T. Fan, X. Cheng, J. Pan, P. Long, W. Liu, R. Yang, and D. Manocha, “Getting robots unfrozen and unlost in dense pedestrian crowds,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1178–1185, 2019.
- [29] Y. F. Chen, M. Liu, M. Everett, and J. P. How, “Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning,” in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 285–292.
- [30] M. Everett, Y. F. Chen, and J. P. How, “Motion planning among dynamic, decision-making agents with deep reinforcement learning,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3052–3059.

- [31] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2255–2264.
- [32] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, “Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 424–14 432.
- [33] A. Vemula, K. Muelling, and J. Oh, “Social attention: Modeling attention in human crowds,” in *2018 IEEE international Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4601–4607.
- [34] D. Zhao and J. Oh, “Noticing motion patterns: A temporal cnn with a novel convolution operator for human trajectory prediction,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 628–634, 2020.
- [35] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, “Intention-aware online pomdp planning for autonomous driving in a crowd,” in *2015 ieee international conference on robotics and automation (icra)*. IEEE, 2015, pp. 454–460.
- [36] T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, and D. Rus, “Intention-aware motion planning,” in *Algorithmic foundations of robotics X*. Springer, 2013, pp. 475–491.
- [37] C. Choi, S. Malla, A. Patil, and J. H. Choi, “Drogon: A trajectory prediction model based on intention-conditioned behavior reasoning,” *arXiv preprint arXiv:1908.00024*, 2019.
- [38] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, “Activity forecasting,” in *European conference on computer vision*. Springer, 2012, pp. 201–214.
- [39] J. Liang, L. Jiang, K. Murphy, T. Yu, and A. Hauptmann, “The garden of forking paths: Towards multi-future trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 508–10 518.

- [40] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, “Precog: Prediction conditioned on goals in visual multi-agent settings,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2821–2830.
- [41] T. Yagi, K. Mangalam, R. Yonetani, and Y. Sato, “Future person localization in first-person videos,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7593–7602.
- [42] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, “Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 085–12 094.
- [43] E. T. Hall, “A system for the notation of proxemic behavior,” *American Anthropologist*, vol. 65, no. 5, pp. 1003–1026, 1963. [Online]. Available: <http://www.jstor.org/stable/668580>
- [44] E. Hall, “The hidden dimension, originally published 1966,” 1990.
- [45] R. Kirby, *Social robot navigation*. Carnegie Mellon University, 2010.
- [46] L. A. Hayduk, “The shape of personal space: An experimental investigation.” *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, vol. 13, no. 1, p. 87, 1981.
- [47] D. Helbing and P. Molnár, “Social force model for pedestrian dynamics,” *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, may 1995. [Online]. Available: <https://doi.org/10.1103%2Fphysreve.51.4282>
- [48] M. Gérin-Lajoie, C. L. Richards, J. Fung, and B. J. McFadyen, “Characteristics of personal space during obstacle circumvention in physical and virtual environments,” *Gait & posture*, vol. 27, no. 2, pp. 239–247, 2008.
- [49] L. A. Hayduk, “Personal space: Understanding the simplex model,” *Journal of Non-verbal Behavior*, vol. 18, no. 3, pp. 245–260, 1994.

- [50] T. Amaoka, H. Laga, S. Saito, and M. Nakajima, “Personal space modeling for human-computer interaction,” in *International Conference on Entertainment Computing*. Springer, 2009, pp. 60–72.
- [51] R. Mead and M. J. Mataric, “A probabilistic framework for autonomous proxemic control in situated and mobile human-robot interaction,” in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 193–194.
- [52] G. Csibra and G. Gergely, “‘obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans,” *Acta psychologica*, vol. 124, no. 1, pp. 60–78, 2007.
- [53] C. L. Baker, R. Saxe, and J. B. Tenenbaum, “Action understanding as inverse planning,” *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.
- [54] E. Wiese, A. Wykowska, J. Zwickerl, and H. J. Müller, “I see what you mean: How attentional selection is shaped by ascribing intentions to others,” 2012.
- [55] D. Carton, W. Olszowy, and D. Wollherr, “Measuring the effectiveness of readability for mobile robot locomotion,” *International Journal of Social Robotics*, vol. 8, no. 5, pp. 721–741, 2016.
- [56] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, “Legibility and predictability of robot motion,” in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.
- [57] T. Kruse, P. Basili, S. Glasauer, and A. Kirsch, “Legible robot navigation in the proximity of moving humans,” in *2012 IEEE workshop on advanced robotics and its social impacts (ARSO)*. IEEE, 2012, pp. 83–88.
- [58] D. Szafir, B. Mutlu, and T. Fong, “Communication of intent in assistive free flyers,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 358–365.

- [59] C. Lichtenthäler, T. Lorenzy, and A. Kirsch, “Influence of legibility on perceived safety in a virtual human-robot path crossing task,” in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 676–681.
- [60] M. Saerbeck and C. Bartneck, “Perception of affect elicited by robot motion,” in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 53–60.
- [61] H. Knight and R. Simmons, “Expressive motion with x, y and theta: Laban effort features for mobile robots,” in *The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, 2014, pp. 267–273.
- [62] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*. CUP Archive, 1990, vol. 7.
- [63] F. Setti, C. Russell, C. Bassetti, and M. Cristani, “F-formation detection: Individualizing free-standing conversational groups in images,” *PLoS one*, vol. 10, no. 5, p. e0123783, 2015.
- [64] M. Vázquez, A. Steinfeld, and S. E. Hudson, “Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 3010–3017.
- [65] D. Helbing, “A fluid dynamic model for the movement of pedestrians,” *arXiv preprint cond-mat/9805213*, 1998.
- [66] A. Garrell and A. Sanfeliu, “Cooperative social robots to accompany groups of people,” *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1675–1701, 2012.
- [67] N. Noceti and F. Odone, “Humans in groups: The importance of contextual information for understanding collective activities,” *Pattern Recognition*, vol. 47, no. 11, pp. 3535–3551, 2014.

- [68] S. D. Khan, G. Vizzari, S. Bandini, and S. Basalamah, “Detection of social groups in pedestrian crowds using computer vision,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2015, pp. 249–260.
- [69] L. Bazzani, M. Cristani, and V. Murino, “Decentralized particle filter for joint individual-group tracking,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1886–1893.
- [70] G. Gennari and G. D. Hager, “Probabilistic data association methods in visual tracking of groups,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II.
- [71] M.-C. Chang, N. Krahnstoever, and W. Ge, “Probabilistic group-level motion analysis and scenario recognition,” in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 747–754.
- [72] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn, “Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker,” in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2011, pp. 120–127.
- [73] R. Mazzon, F. Poiesi, and A. Cavallaro, “Detection and tracking of groups in crowd,” in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2013, pp. 202–207.
- [74] B. Lau, K. O. Arras, and W. Burgard, “Multi-model hypothesis group tracking and group size estimation,” *International Journal of Social Robotics*, vol. 2, no. 1, pp. 19–30, 2010.
- [75] M. Mucientes and W. Burgard, “Multiple hypothesis tracking of clusters of people,” in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 692–697.

- [76] M. Luber and K. O. Arras, “Multi-hypothesis social grouping and tracking for mobile robots.” in *Robotics: Science and Systems*, 2013.
- [77] C.-E. Tsai and J. Oh, “A generative approach for socially compliant navigation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2160–2166.
- [78] J. Oh, T. M. Howard, M. R. Walter, D. Barber, M. Zhu, S. Park, A. Suppe, L. Navarro-Serment, F. Duvallet, A. Boulaïas *et al.*, “Integrated intelligence for human-robot teams,” in *International Symposium on Experimental Robotics*. Springer, 2016, pp. 309–322.
- [79] C. Mavrogiannis, A. M. Hutchinson, J. Macdonald, P. Alves-Oliveira, and R. A. Knepper, “Effects of distinct robot navigation strategies on human behavior in a crowded environment,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 421–430.
- [80] T. Fraichard, “A short paper about motion safety,” in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 1140–1145.
- [81] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 261–268.
- [82] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.
- [83] P. Fiorini and Z. Shiller, “Motion planning in dynamic environments using velocity obstacles,” *The international journal of robotics research*, vol. 17, no. 7, pp. 760–772, 1998.
- [84] J. Alonso-Mora, A. Breitenmoser, M. Rufli, P. Beardsley, and R. Siegwart, *Optimal Reciprocal Collision Avoidance for Multiple Non-Holonomic Robots*. Berlin,

- Heidelberg: Springer Berlin Heidelberg, 2013, pp. 203–216. [Online]. Available: https://doi.org/10.1007/978-3-642-32723-0_15
- [85] J. v. d. Berg, S. J. Guy, M. Lin, and D. Manocha, “Reciprocal n-body collision avoidance,” in *Robotics research*. Springer, 2011, pp. 3–19.
- [86] Stereolabs. Zed 2i - industrial ai stereo camera. [Online]. Available: <https://www.stereolabs.com/zed-2i/>
- [87] D. Dugas, J. Nieto, R. Siegwart, and J. J. Chung, “Ian: Multi-behavior navigation planning for robots in real, crowded environments,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 368–11 375.
- [88] A. Johansson, D. Helbing, and P. Shukla, “Specification of a microscopic pedestrian model by evolutionary adjustment to video tracking data,” *arXiv preprint arXiv:0810.4587*, 2008.

Appendix A

Access to Project Resources

A.1 Access to model code

The SFM implementation and ASFM implementation are available at: https://github.com/SebastianAgi/social_navigation.git

A.2 Project Video Demonstration

A demonstration video of the SFM and ASFM implementation can be seen at: https://github.com/SebastianAgi/social_navigation.git