Title	人とロボットの社会的インタラクションにおける対人距離を学習するプロクシミクスの研究
Author(s)	Patompak, Pakpoom
Citation	
Issue Date	2019-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/16172
Rights	
Description	Supervisor:丁　洛榮，情報科学研究科，博士

Learning Proxemics for Identifying Human Private Space in Human-Robot Social Interaction

Pakpoom PATOMPAK

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Learning Proxemics for Identifying Human Private Space in Human-Robot Social Interaction

Pakpoom PATOMPAK

Supervisor: Professor Chong Nak-Young

School of Information Science
Japan Advanced Institute of Science and Technology

September 2019

Abstract

Mobile robots are tended to provide more and more service in the shared environment with humans. Human-robot interaction (HRI) is a critical component to allow a robot to operate with humans in the proper direction. To design the robot system to operate with humans natural and acceptable, robots should have the ability to perceive, understand and act in a manner that conforms to the social convention like move to the right side of corridor or keep human personal or private space during an interaction , which is the fundamental key to human-robot symbiosis. Notably for a navigation task that robots should move to provide the services in a different location, robots should manoeuvre themselves without harm or damage the surrounding environment which includes humans. Although robots can generate safe navigation, sometimes humans feel not safe with the robot motion. The main reason come from the lacking of trust to the technology which occurs from the unfamiliar of the robot's appearance or less experience with the robot. Therefore, the robot navigation task should not consider only safe behaviour but should increase attention to generate social behaviour which enables the robot to behave more naturally and acceptable to operate with humans.

For human-human interaction, the personal area is the one instance social convention that humans consider when interact with others. This interaction area of humans consists of two areas. First is the quality interaction area, where humans can be engaged in high-quality interactions with others. Second is the area of privacy where humans do not want to interfere with others speech or action. The size of these two areas usually depends on various social information such as their motion, personal traits, and acquaintanceship. The same concept applies to the case of human-robot interaction, especially when the robot is required to exhibit a certain level of social competence. Therefore, the challenge is how to formalize or estimate the personal area from various human social information.

In this dissertation, we proposed a new robot navigation strategy to socially interact with humans reflecting upon the social information between the robot and each person. The proposed model aims to enable the robot to estimate or delineate the personal area of each person by using their social information and it is possible to update this personal area based on their feedback. The results of our method enable the robot to estimate the personal area and update it until it appropriates to each person. This adaptive personal area assists the path planner to generate the path that does not intrude into the area of privacy but keeps distance to give a quality interaction.

The proposed model uses an asymmetric Gaussian function to estimate each personal area where a fuzzy inference system is used to design the required parameters. The fuzzy membership functions are optimized to give the robot the ability to navigate autonomously in the quality interaction area using a reinforcement learning algorithm. It was verified through simulations and experiments with a real robot that the proposed strategy can generate a suitable personal area of each person that allowing the robot to maintain the quality of interaction with each person while keeping their private personal distance.

Keywords: Proxemics, Social Interaction, Social Force Model, Fuzzy Inference System, Reinforcement Learning

Acknowledgments

Foremost, I would like to express my gratitude to my supervisor Prof. Dr. Nak Young-Chong for his tremendous guidance and support. I appreciate his broad and deep knowledge and patience whenever we discuss. He provided me with the chance to open my eyes and ears on the robotic research. He also helps me for financial support in the last year study in Japan. His guidance helped me in all the time of research. Without his supervision, there seems to be no end in sight to my PhD course.

I also sincerely like to extend my heartfelt gratitude to my vice supervisor, Asst. Prof. Dr. Sungmoon Jeong whose lot of insightful question and comment on my presentation, manuscript and my research results which help me to improve my skills and knowledge.

I also sincerely like to thank Asst. Prof. Dr. Itthisek Nilkhamhang, my thesis co-advisor from Sirindhorn International Institute of Technology, Thailand. He gave me much motivation and gave me chances to develop my skills.

Besides my advisors and co-advisors, I would like to thank my Thai professors: Dr. Kanok-vate Tungpimolrut, my thesis co-advisor from the National Electronics and Computer Technology Center (NECTEC). Assoc. Prof. Dr. Waree Kongprawechnon and Assoc. Prof. Dr. Toshiaki Kondo for their encouragement, insightful comments, and hard questions.

I greatly appreciate my classmates and my fellow doctoral students in JAIST Dual Degree program for all the fun we had while we studied. I am also grateful for my friends in Chong's Lab at JAIST to answer my simulation, coding, and my research problem and the sleepless nights working together.

Last but not least, I would like to appreciate my family members who supported me spiritually throughout my life especially my mother who always ask and take care of my health while I was working. She always understands and supports me everything in all situation, consoles me when I feel despond. All of the words that mean thank you is not enough to describe my feeling.

This research is financially supported by the European Commission and The Ministry of International Affairs and Communications of Japan, Japan Advanced Institute of Science and

Technology, Sirindhorn International Institute of Technology (SIIT), and Thammasat University (TU).

Table of Contents

This dissertation was prepared according to the curriculum for the Collaborative Education Program organized by Japan Advanced Institute of Science and Technology and Sirindhron International Institute of Science and Technology, Thammasat University.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Importance and Challenges

Mobile robots are trended to provide more and more service in a shared environment with humans, e.g., house, office or co-worker space. The application of mobile robots has ranged from a co-worker robot in the industries to domestic services robot that assists humans in their home, for example, the robot that takes care of the elderly person or handicapped person in the house. For a task that requires robots to move and provide services in a different location, safe navigation is one of the essential functions that the robot needs to concern. The robot needs to generate a path that does not harm or damage the surrounding environment which includes humans. However, even when the robot is moving safely, sometimes human have unsafe feeling from the robot motion because of they do not understand in the technology [7]. These lacking trust feeling may occur from the lack of experience to the technology or unfamiliar of the robot appearance. Therefore, the navigation of the robot should not consider only on safe behaviour but should increase attention to generate socially competent behaviour like moving naturally or considering the human comfort space. This concept makes the robot to behave more naturally and acceptable for humans to feel safe and comfortable to operate with the robot in the human-robot shared environment.

Roboticist should consider two majors constraints to design robot navigation in the shared environment. First is the safety constant which prevent the situation that can damage the robot and surrounding environment [8]. The second is an instance of social constraint such as which is helpful to avoid situations potentially annoying or making discomfort to humans. Early robotics research had mainly focused only on the first constraint. For example, [8, 9] use dynamic window approach which consists of velocity reachable within a short time interval. This method determines the stop area before the robot colliding with obstacles. This stop area might make the robot to perform instant stop behaviour which might interrupt with human feeling during the operation. Therefore, recent developments of mobile robot navigation have integrated early research with social science and psychology studies to meet both safety and socially constraints. These research are grouped into the robot field called human-aware navigation, which is one of the crucial challenges for human-robot symbiosis.

There are different approaches that human-aware navigation research follows, for example, the approach to make a robot move or behave more naturally, the approach to design the robot navigate according to the social norm, or design the robot to approach or interact with humans by minimizing their annoyance, stress or discomfort feeling. All of these approaches have a common theme to make the robot acceptance to humans. Hence, the challenge is how can the roboticist design the robot behaviour to natural and acceptable operate to all humans in a shared environment.

To make robots navigate or interact naturally with humans, robots should have the ability to understand each social convention of humans. In human-human interaction, humans perform different behaviour to a different person which depend on their social information. This phenomenon is described in the famous theory of social science called *Proxemics theory* [10]. The theory describes how human use surrounding space and effect that population density has on behaviour, communication and social interaction. In other words, proxemics theory describes different interpersonal space that the human keep from others depends on social information like cultures, personal traits, age or relationship. These interpersonal spaces can represent the area of privacy or personal area that the human does not want to share with others. The same concept can be applied and integrated into the case of human-robot interaction and human-aware navigation, especially when the robot is required to exhibit a certain level of social competence. However, it is still a challenging problem to formalize this social science and psychological concept into a mathematical model to delineate and estimate individual personal area.

Lindner summarized that the area of privacy or personal area of a human could be delineated and estimated based on the geometric and potential field [11]. The geometric model has a clear boundary to represent a sharp transition between different zones of interaction area while potential field or cost function is superior to solutions defining forbidden zone around humans [5, 12–15]. The results of these estimated methods are used to assist the path planning algorithm in generating the path that does not intrude into the human's comfort area or private area. However, the existed method have not considered human's social information or consider just only one information, for example, a motion of a human, which causes the incorrect estimation in size or boundary of the estimated personal area. This result makes the generated path interrupts or intrudes into the area of privacy of the humans which make them discomfort to operate with the robot.

Even though the understanding of the relationship between the personal area of the human and the social information allows the robot to estimate the personal area, but it still has uncertainties that originate from humans and the surrounding environment that the roboticist may overlook when design method to estimate the personal area. These uncertainties comes from a difference in cultures or lifestyle of each person, which makes the robot estimates inaccurate personal area. Therefore, learning ability is another essential ability that roboticist should consider equal to the understanding ability in designing the robot behaviour.

In the humans learning process, humans try to solve the uncertainties problem based on their experience. Humans improve their performance by obtaining the experience and knowledge from the surrounding environment's feedback signal. This experience must be obtained by interact and observe the results. Then, humans use these results as the base to make better decisions to improve performance when facing the same situation. A framework that using the experience to improve the performance of the agent is similar to one of the machine learning, *reinforcement learning* (RL). The concept of RL is to reinforce the decision that has led to a good outcome according to the experience by increasing the chance to perform the same decision again. The same concept can be applied to the robot as the learning ability to solve the uncertainty problem of environment [16].

Therefore, this dissertation proposed a new robot navigation strategy to socially interact with humans reflecting upon the social information between the robot and each person like genders, perception range or acquaintanceship. The proposed method aims to enable the robot to estimate or delineate the personal area of each person by using their social information and update the estimated area according to their feedback. Therefore, the robot can interact with humans by not intrude into their private area but also in the range that humans can receive good interaction quality.

The proposed method based on the potential field concept which gives the different degrees at different locations [17] by using the fuzzy inference system (FIS) to map the social information to important parameters for estimate the personal area of each person. However, with the preliminary setting of the mapping process, the aberration of the estimated area can occur due to the uncertainty of humans. Here, the machine learning algorithms like reinforcement learning which is the learning algorithm that learning from the experience and similar to the human learning process is applied to adjust the parameters that can make the estimated personal area more accurate and appropriate to the human.

1.2 Research Motivation

A literature review of many human-aware navigation research has suggested that the model of human's private area or personal area is useful to enable the robot to operate or navigate in the shared environment with acceptance from humans. The concept is that the robot estimates the area and uses it for the base information in path planning algorithm. However, existed work have two critical problems.

1. The existed privacy or personal area estimation methods have not used social information or use only one social information to estimated the area. The reason is that social information is difficult to formulate into mathematics or accurate value which means it is difficult to estimate the correct personal area. This effect on the generated path that may interrupt human comfort.

2. The existed estimated methods have no ability to learn and adjust the incorrectly estimated personal area that may occur from the uncertainties like the difference of cultures or lifestyle of each person. This makes the estimated area is not appropriate to the person.

Our motivation for working with the estimation of the personal area of each person comes from the requirement of the robot to naturally and acceptably operate in the shared environment with humans. Our premise is that once the robot understands the relationship between the personal area and individual social information, and can receive the human response or feedback, the robot will be able to estimate and update the individual personal area then behave to interact with the humans in the appropriate distance which does not violate human's comfortable feeling.

1.3 Research Objective

Our ultimate goal of this work is to model an adaptable personal area that use to assist the path planning of the robot. To reach that ultimate goal, a few sub-goals are set for this dissertation.

- To determine the variance parameters of asymmetric Gaussian function that uses to estimate the personal area of each person by using their social information like genders, perception range and acquaintanceship.
- To modify or update the mapping technique's parameters by using reinforcement learning that enables the robot to adapt parameters according to the humans' response during the operation.
- To study the efficacy of popular reinforcement learning techniques for parameters adaptation problem that uses in human's social space model.
- To demonstrate the proposed method the real-world human-robot interaction.

Finally, the robot should be able to estimate each person's personal area and be able to adapt it to individual's preferences in the shared environment.

1.4 Thesis outline

This thesis organizes as follows. Chapter 2 shows a review of works related to our research. It presents the summary of the social model in human-robot interaction which used to estimate the personal area of the human , and reinforcement learning which used for parameter adaptation. This information is necessary for understanding the proposed method in this thesis.

Chapter 3 discusses the personal space estimation for robot navigation. In this chapter, the mathematical model of the human's social space or personal space is described. It also covers the process to map social information or factors to human's personal space, and how to formalize our social model to the RL frame work.

Chapter 4 discusses the detail of reinforcement learning that we have used to deal with the parameters adaptation for human's social space estimation. The results of different reinforcement learning algorithm are shown to compare their efficacy.

Chapter 5 shows the experiment results that implement with the humanoid robot *Pepper*. The results prove that our proposed model can enable the robot to estimate the social area and interact with humans in the appropriate area.

The last chapter gives a concluding remark and direction for future research. The organization of this dissertation can be illustrated as Figure 1.1

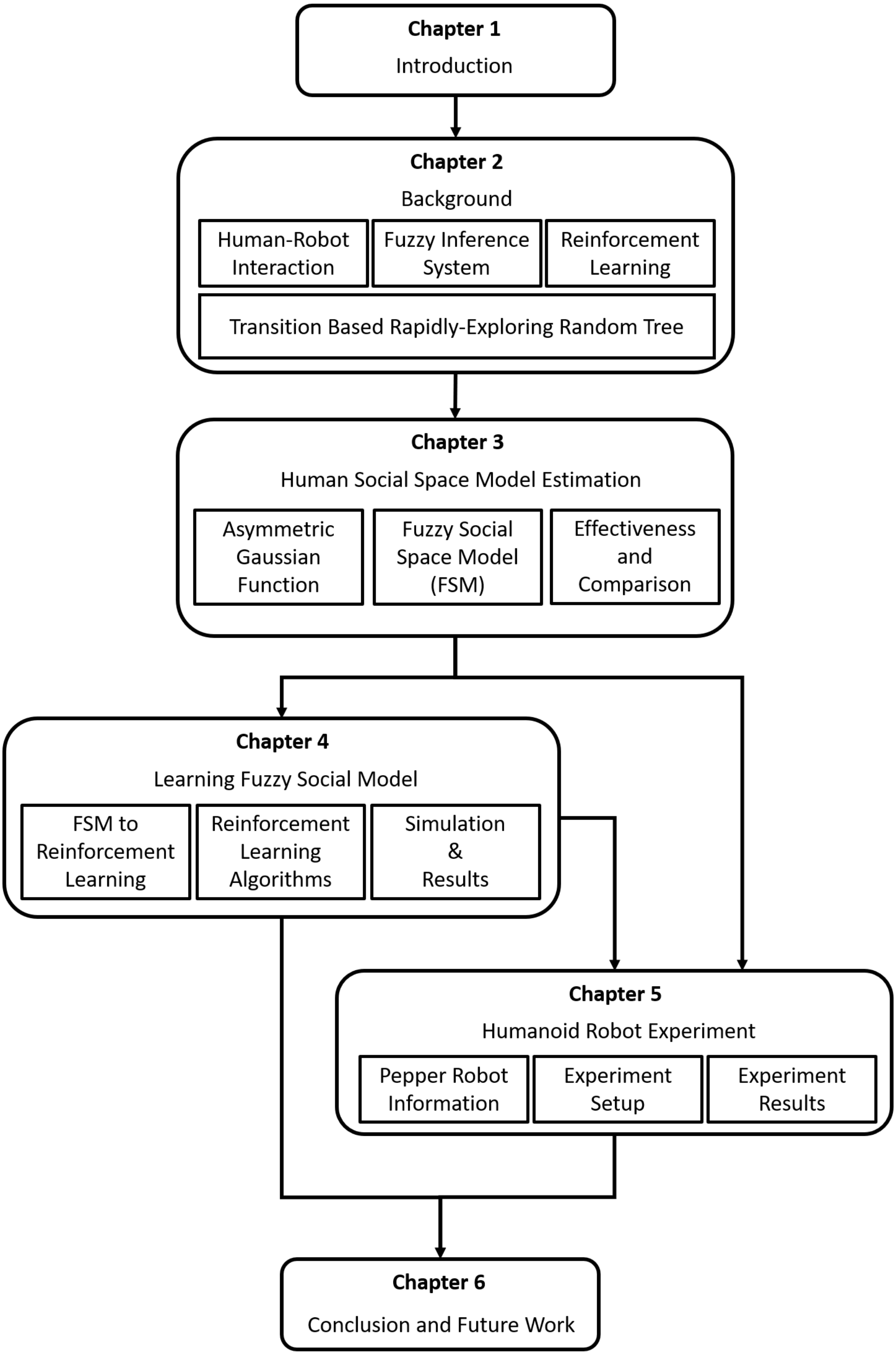


Figure 1.1 The organization of this dissertation

Chapter 2

Background

This section has attempted to provide a summary of the literature relating to our work. This chapter began by briefly introduces the concept of human-robot interaction (HRI) which include the knowledge of social science that useful to human-aware navigation, and the application in robot navigation in the human shared environment. Then provide the information about reinforcement learning that will be used to determine the efficacy of its techniques to parameter adaptation.

The first section has endeavored to grasp the definition of the human's area of privacy in social science and exemplified the studies to support its definition. Then the studies that relate to robot application like human-aware navigation are present. The second section has provided the necessary information about reinforcement learning (RL) and shown the beauty of the variety of its applications that can be used in any application.

2.1 Human-Robot Interaction

Human-Robot Interaction (HRI) is a field of study to understanding, designing, and evaluating robotic systems for use by or with humans [18]. Interaction can be several forms such as speaking to each other, operating in the same area, walking companion or guiding to the destination location [12, 15]. However, to design the robot system to operate naturally and acceptable in the shared environment, the roboticist should understand the social information from humans' behaviors. Therefore, the knowledge and comprehension of social science and psychology are vital to model and design human-robot interaction. The goal of this section is to present some definition of social science study that useful for human-robot interaction and discuss challenge problems that are likely to shape the human-robot interaction field in the near future.

2.1.1 Privacy and Proxemics in Social Science

The key to formalize human personal space model is to understand and accommodate human behavior. Therefore, the knowledge of social science and psychology is a vital aspect which allows the robot to better understand the behavior of the humans. When robots operate in a shared environment, the area of privacy is the crucial key for naturalness, sociability, and acceptance. Privacy was defined in human-human interaction study by Jonathan Herring [19]. They defined privacy as the ability of an individual or group to separate themselves and select to share some of their information to whom they allow. The boundaries and content of what is considered private differ between cultures and individuals.

There is a lot of research that exemplifies the study of privacy of the living things. For instance, Westin [20] mentioned that most animals seek privacy either as individuals or in small groups. In this study, he reported three areas of privacy observed among animals which included: personal distance between animals, social distances between groups, and fighting distances at which an intruder cause conflicts. At the same time, animals often gather in large groups. They seem to live in a tension between privacy and sociability.

Zeeger studied human privacy in childhood [21]. He found that 58 of 100 of three to five-year-olds said they had a special place at the daycare center which belonged only to them. Newell et al. investigated the reason why humans required privacy. By the survey, they found that most of the participants in different cultures believed that emotion like grief, fatigue or attention were the main useful sets associated with seeking privacy [22].

Another study about the spacing of human was introduced by Edward T. Hall [10]. He introduced the theory of Proxemics, which describe how humans use space, and the effect that population density has on behavior. He emphasized the impact of the use of space on interpersonal communication. According to his study, Proxemics is valuable in organizing the surrounding space to interact with others. These organized spaces depend on the type of interaction and social information between individual. Therefore, the human interaction area could be organized as follows:

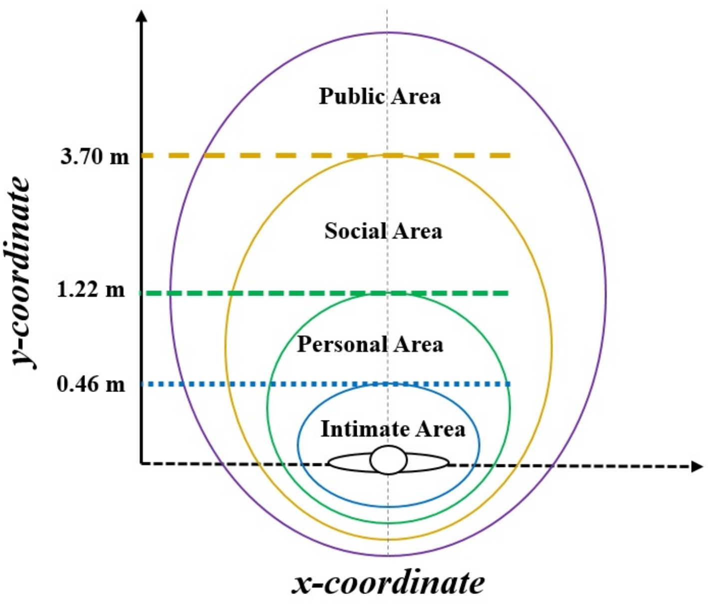


Figure 2.1 Human interaction area according to "Proxemics" theory which introduced by Edward T. Hall

- **Intimate Area** is an area for intimate contact like whispering, touching or hugging with very close relationship person like wife and husband or mom and children. This area has a distance less than 0.46 meters with respect to the human's center.
- **Personal Area** is the zone for people who have close relationships. In this space, humans feel discomfort if unfamiliar being enters this area. This area has a distance greater than 0.46 meters but less than 1.22 meters.
- **Social Area** is the space that humans use to contact with new acquaintances. The distance of this area is between 1.22 and 3.70 meters.
- **Public Area** is the space that often used to interact with strangers or to give public speeches. The distance of this area is more than 3.70 meters.

The organized space for human interaction is shown in Figure 2.1. On the one hand, humans use these organized space concepts to approach others humans. For example, humans try to get near to the close friend or the member in the family to get a better quality of interaction; however, they keep the distance or space from the strangers to increase comfortable feeling. Consequently, it is evident that closeness is paramount for good interaction, but an area of privacy should also be respected.

Furthermore, protecting one's privacy is an essential prerequisite for forming long-term, stable relationships. This concept can be applied to develop the social robot. For example, in the approaching to the human problem in [23] or the problem of path planing in crowded environment in [24]. Human-aware navigation is the topic of research that assists the robot to navigate in the human-robot shared environment. The next section will provide the studies in human-aware navigation that relate to the area of privacy of the human.

2.1.2 Human-Aware Navigation

There are different goals that human-aware navigation research follows. Most of the research attempt to minimize annoyance, stress, and discomfort, so that the robot can interact more comfortably with humans [12, 25, 26]. Other approaches focus on the robot behaving more naturally and behaving according to the social norm. All of these goals have a common theme that attempts to make the robot acceptance to humans. However, the method may vary. Therefore, the following definition of naturalness, sociability, and comfort can be used to classify the research reviewed.

Naturalness

Naturalness is the low-level behaviors pattern of the robot that are similar to humans. This group of research attempts to imitate nature behavior, such as human motion, to recreate the robot's behavior. Most research in this group works well with low-level behaviors like shapes and velocities where a continuous measure can be applied between human-robot behavior.

Natural motion is another goal in human-aware navigation that mimics human behavior to the robot navigation for human acceptance. The assumption of natural behavior research is if a robot behaves more similarity to humans, the interaction between them becomes easier and more intuitive for humans [7].

One aspect of natural motion is smoothness. This aspect refers to both the geometric path and the velocity of the robot. For example, [27] presented that a principle of energy optimization influences human motion. They summarized that the behavior of to approach the group of humans like the speed of movement should depend on the distances between the robot and humans. The robot should slow down when getting closer to not scare anybody.

The motion relative to other agents, like humans or robots, is one of naturalness research.

Figure 2.2 The experiment to investigate the natural motion of person-following in hallway [1].

Gockley et al. [1] did the experiment to investigated two different approaches of person-following such as direction following and path following, to see which approach is more natural motion. These two different approaches have been rated by the participant in the pilot study, as shown in figure 2.2. The results show that humans feel more acceptable when the robot navigate with sharing the same direction by following rather than using same path as humans.

Another exaple is the behavior design of approaching a group of human and maintaining formation [28]. They suggested that the robot should maintain a certain distance to the closest person, and it should face to the middle of the group. Another aspect is to investigate on how the different kinds of nonverbal cues were used to catch the attention of humans. Saulnier et al. [29] designed the path planning based on the cost-map function to the robot arm. This robot arm tried to pick up an object and send to the human according to the desired path. This study shows that the navigation behavior can serve as the messages for nonverbal communication. In addition, this is another natural language that must be considered for avoiding misunderstanding.

Gracia et al. [24] and Tamura et al. [30] used Social Force Model (SFM) as a means to guide the robot to navigate through a group of humans in a natural way. Social Force Model (SFM) represents moving agents like a robot or human as a mass under virtual force. Thus, the robot can move to its goal and avoid the obstacles by virtual repelling force from them. This model can be used as the input for robot motion control.

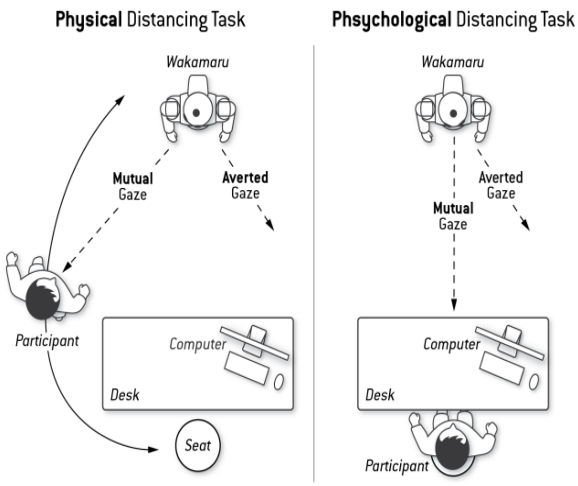


Figure 2.3 The experiment for physical distance task and psychological task that have set to explore that social norm is effected to these distances. These could be used to design proxemics behavior for the robot [2].

The recently challenge is to make the robot navigate into densely crowded area which is difficult to make the robot behave in natural way. For example, the strategy of making a robot exhibit human-like behavior in the highly populated environment [30, 31].

Sociability

Sociability adheres to explicit high-level culture conventions. This group of research is concerned with the different abilities of the robot compared to humans and how not all of these abilities are suitable to transfer from humans to the robot. Therefore, it is considered that the robot are able to make the same high-level decisions like humans. Protocols consider for sociability are constraints imposed by society. For example, the rule to walk on the right-hand side in corridors, or to approach the human concerning the social relationship information.

Human-aware navigation can be improved by adding behavior that considers social protocols for behavior in a particular situation. In navigation, there are rules such as standing in queues, excusing oneself when one has to traverse a personal zone to reach a goal, and so on. Consequently, the robot should understand social rules or social signals to behave correctly to the human in human interactions. Amount of research studies improved approach direction initiate explicit interaction [2, 23, 28, 32–35]. They suggested that violation of social rules or social signal can also cause the discomfort of the humans.

In [2], the experiment is constructed to explore whether the proxemics model can explain how people physical and psychological distance themselves from the robot. This also guideline how to use proxemics behavior for the robot. The experiment was set two test with two different distances. The participant asked to approach the robot to do some task and measure physical distance. Then the robot asked their personal distance to measure the psychological distance as shown in figure 2.3. The results show that the person who did not like the robot maintained the physical distance with the robot when the robot moved its gaze, and also disclosed less personal information to the robot.

Another example interesting experiment is in [33]. The experiment was set to investigate the direction that the robot should use to approach. The robot is controlled to approach the participants in a different direction. The results show that most of the participants did not prefer the robot to approach from the front direction. They preferred the robot to approach at the side especially the right side of them. The extended experiment from this paper includes the behavior that the robot hand the can of soft drink to the human [35]. The robot handing over human' hand position had the most influence on determining from where the robot approach.

All of these research suggested that violation of social rules or social signal can also cause the discomfort of the humans.

Comfort

Comfort is the absence of annoyance and stress for humans when interacting with robots. The comfort is different from safety. Even when the robot moves safely in human's zone, humans may still feel unsafe because of the lack of trust in technology, due to being unfamiliar with the appearance or the robot type. Therefore, the research on comfort attempts to not only make the robot move safely but also manoeuver to make human feel more relaxed.

When the robot is moving toward its destination, it can cause discomfort to humans by moving too close, too fast, too slow or getting in the way. This section presents the research that points out the causes of discomfort reactions felt by humans and how to alter the robot's behavior to reduce this discomfort. Most of the given literature, on comfort requirement, stresses the importance to the distance a robot needs to keep from humans. This distance does not serve collision avoidance but prevents the feeling of discomfort.

Edward T. Hall proposed the concept of virtual personal space around a person which others

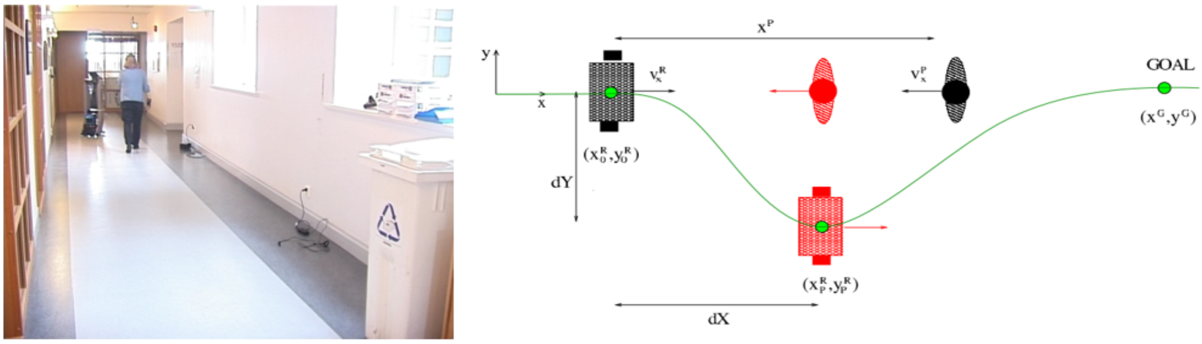
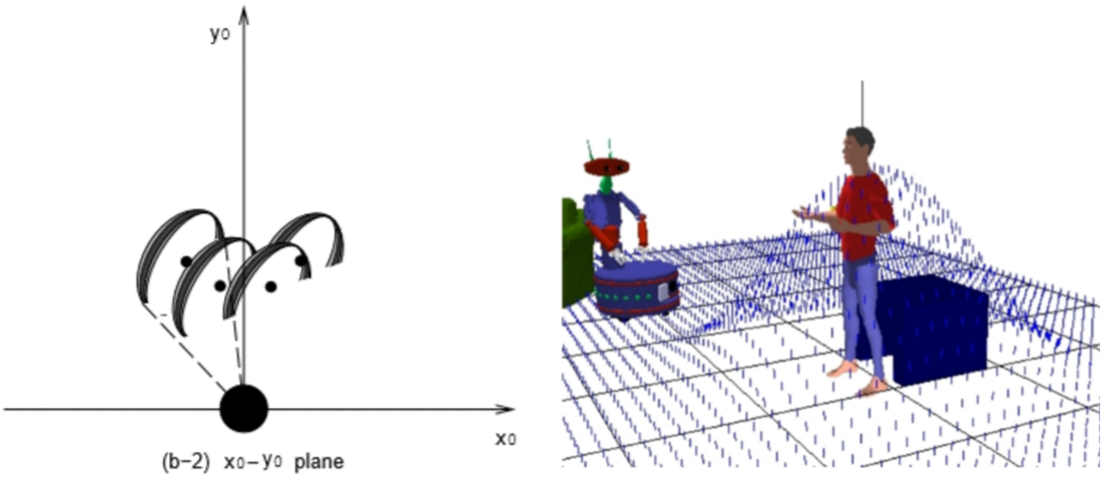


Figure 2.4 The experiment located in the hallway. Pacchierotti et al. desired the robot to move to the left side of humans by keep the large distance as possible to give the human more comfort [3].

should respect called *Proxemics*. He found differences in interaction space that humans chose for human-human interaction depending on the relationship and intention. The idea is that when interacting with other agents, humans feel annoyed or show signs of discomfort when others get too close or too far away. This general idea of Proxemics can be applied to the appropriate space chosen by the robot for any explicit or implicit interaction with humans. Paccheierotti [3, 25] presented the studies of the robot navigation in a hallway with humans walking in opposite directions, as shown as figure 2.4. In their studies, they applied a control strategy as a reaction to humans coming the other way. The robot deviated a larger lateral distance making participants feel better. However, on a few occasions, a large lateral distance was evaluated as unnatural. Butter and Agah [32] presented several experiments where a robot approached a standing person. They found that different types of the robot, such as vacuum robots or humanoid robots, caused different levels of discomfort even at the same distance. Takayama and Pantofaru [23] extracted the robot design for approaching distance and gaze depending on social contexts. They summarized that the familiarity with robots and attitude towards them should be taken into account for personal space selection.

To take the aspect of comfort beyond the definition of "a distance to maintain," Martinson [36, 37] takes into account the noise generated by the robot's motion itself and presents an approach to generate a hiding path while moving around a person. A model of human awareness to navigate the robot in a way that reduces noise discomfort is present in [26]. Another way to be comforting to others is not to disturb them unless it is necessary.

For example, Tipaldi et al. [38] considers this aspect by programming the robot to operate in the area which does not impediment humans while performing task like cleaning the home.



(a) (b)

Figure 2.5 Human personal space can be model based on geometric 2.5a and cost function like potential field concept 2.5b. For example, the ellipse function was used to determine the personal space of human in line scenario [4], while Sisbot et al. use the potential filed concept as the cost function with depend on human posture [5].

The strategy is the robot tried to avoid the area that human stay by using a "spatial affordance map" which presented the location of human from the probabilities of human activity in each area of time interval that come from observation. The results show that the robot used the map to make make a decision about whether its activity was likely to occur in shared environment.

2.1.3 Human Social Model

Human-aware navigation of mobile robot should consider two constraints. The first is the task constraints which include minimizing the distance traveled toward a goal, avoiding obstacles and keeping a safe distance from them [4, 13]. This task constraint is considered to be a major significance in every research of robot navigation. An additional constraint is the social constraints that include the social convention, such as comfort, naturalness, sociability [15, 39], [5]. The challenge is how to formalize both constraints into mathematics model. Therefore, in this section, we show the research about human social interaction space model that is used as the base for the robot navigation.

Using the concept of Proxemics, Lindner summarized that the interaction area could be delineated based on the geometric and potential field [11], as shown in figure 2.5. The geometric model has a clear boundary to represent a sharp transition between different zones of interaction area [12], for example, explicit model to represent personal space in nearest histograms for local navigation in [13, 40], or using the ellipse function to determine the humans' personal space in queue scenario. This help the robot to approach or avoid the person in queue [4], as shown in figure 2.5a.

Another method to describe interpersonal space is to define the cost function or potential field [5, 14, 15, 41, 42]. The cost function is superior to solutions defining forbidden zone around humans [39] because in the limited space the cost function can be useful and necessary for the robot to move past humans. This cost function and potential field concept were used in [15] to prevent the robot from getting into the human's forbidden area. Hansen et al. employed Rapidly Exploring Random Tree (RRT) in the social map which provides the degree of comfortable of humans and gained the response from the robot dynamics and human motion prediction [14]. A model of the level of comfort in humans' field-of-view and posture was used as the cost to guide a human aware motion planner [5], as shown in figure 2.5b. The potential field was used as the cost function to assist the robot in determining the position to approach humans [34].

Even though the human's interpersonal model is the key to estimate the human's comfort space, the ability to re-estimate the model while operating with humans is also important. The important reason is while operating, the change of environment, human's emotion, human's behaviour or their social factors make the estimated model methods less efficient. There are many research deal with these uncertainties. For example, Luber introduced the adaptive method to adjust the interpersonal space of humans while walking in the hallway [43]. The approach used unsupervised learning to produce a set of Relative Motion Prototypes (RMP). The results showed that the generated paths with RMP are similar to the human path which can change due to the time of operation. In [44, 45] presented the adaptive model in a different approach. They presented the model for a group of human while they do the activity like throwing a ball or group talking. They model the humans' comfort zone by employed kernel-based regression with the position, orientation and velocity of humans. Their method outperforms the traditional cost function model and can be used in real time.

2.2 Fuzzy Inference System

Fuzzy inference system (FIS) is the computation technique following an approach that is considered to be somewhat similar to both human reasoning and decision-making process [46]. This thesis uses FIS to choose the appropriate Gaussian parameters value that used to estimating human personal space from social factors. Therefore, this part will explain the concept of FIS that will be used in the research.

In most of the decision making process, the quality of the decision is depend on the capability of addressing uncertainty and imprecise information. For example, the relationships of the robot to the human are frequently described in term of 'familiar to the human' rather than ' 80 percent of the total time with the robot'. In addition, the human social factors which are describe in chapter 3 are defined based on the linguistic variable. Therefore, imprecision between social factors and the value of parameters can be easy to determine.

In this thesis, FIS is incorporated to personal space model estimation to map the social factors like genders, level of the relationship, or perception distance of the human to determine the value of the parameters that will be used to determine the human interpersonal area. The FIS process includes four parts, fuzzy membership function (MFs), fuzzy rules, Inference method, and a defuzzification method which will be described as follows:

2.2.1 Fuzzy Membership Functions

In FIS, a input variable's value can be turned into a fuzzy value by using the membership function (MFs) which describe the degree of the input into linguistic term. These MFs are design based on the experiment data that presents the relationship between one domain to another domain.

Let $\mathcal{U}$ represents the universe or all possible value of input, x is the elements in $\mathcal{U}$. A set $\mathcal{A}$ is a fuzzy subset of $\mathcal{U}$. The element x which belong to a set $\mathcal{A}$ and has the degree between 1 and 0, is called MF values $\mathcal{A}(x)$ in a fuzzy. A fuzzy set $\mathcal{A}$ is marked by an MF $\mu_{\mathcal{A}}$. This membership function links the elements of the universe $\mathcal{U}$ to their corresponding membership value $\mathcal{A}(x)$.

$$\mathcal{A}(x) = \mu_{\mathcal{A}}(x) \in [0, 1], x \in \mathcal{U} \tag{2.1}$$

The μ can be designed by a variety of mathematics shapes depending on how the experimenter or expert connects one domain values to other belief values.

2.2.2 Fuzzy Rules

The relation of domain expert knowledge and belief knowledge are collected and used to construct a fuzzy rules which is usually expressed as a set of "IF-THEN" rules. The antecedent of a fuzzy rules is a combination of fuzzy propositions, which is in the form of "x is $\mathcal{A}$". The result is calculated by the degree to which the antecedent is satisfied.

The FIS describe the connection between input variable and output variable by using the linguistic rules in the form " IF $variable_{input}$ IS $fuzzy_{set}$ THEN $variable_{output}$ IS $fuzzy_{set}$ ".

2.2.3 Inference Method

There are three different method in inference process: Mamdani, Larsen, and Takagi-Sugeno. Mamdani is the famous and stable method which is chosen to used in many of research. The method calculates the fuzzy output of each description parameters based on sub-minimum composition. The general form of multidimensional multiple fuzzy reasoning models is defined by:

$$\begin{array}{cccccc} \mathcal{A}_{11}, & \mathcal{A}_{11}, & ..., & \mathcal{A}_{1n}, & \rightarrow & \mathcal{A}_1 \\ \mathcal{A}_{21}, & \mathcal{A}_{21}, & ..., & \mathcal{A}_{2n}, & \rightarrow & \mathcal{A}_2 \\ \vdots & \vdots & ..., & \vdots & \vdots & \\ \mathcal{A}_{m1}, & \mathcal{A}_{21}, & ..., & \mathcal{A}_{mn}, & \rightarrow & \mathcal{A}_m \\ \mathcal{A}_1^*, & \mathcal{A}_2^*, & ..., & \mathcal{A}_n^*, & \rightarrow & \mathcal{B}^* \end{array} \tag{2.2}$$

where $\mathcal{A}_{ij}$ and $\mathcal{A}_j^*$ are the fuzzy subset of input universe of discourse $\mathcal{U}_j$; $\mathcal{A}_{ij}$ represents the jth input of the ith fuzzy rule in a fuzzy inference model; $\mathcal{A}_j^*$ represents the jth input of an actual antecedent; $\mathcal{B}_{ij}$ and $\mathcal{B}_j^*$ are the fuzzy subsets of output universe of discourse $\mathcal{V}$; $\mathcal{B}_i$ represents the jth output of the ith fuzzy rule; $\mathcal{B}_i^*$ represents the composite output of an actual antecedent (i = 1,2,...,m; j = 1,2,...,n); m is the number of fuzzy rules for a fuzzy inference model; n is the number of antecedent inputs of an 'IF-THEN' fuzzy rule. The inference process is written as:

$$\begin{aligned}
\mathcal{A}_1(x) &= min\mathcal{A}_{11}(x_1), \mathcal{A}_{12}(x_2), ..., \mathcal{A}_{1n}(x_n) \\
\mathcal{A}_2(x) &= min\mathcal{A}_{21}(x_1), \mathcal{A}_{22}(x_2), ..., \mathcal{A}_{2n}(x_n) \\
&\vdots \\
\mathcal{A}_m(x) &= min\mathcal{A}_{m1}(x_1), \mathcal{A}_{m2}(x_2), ..., \mathcal{A}_{mn}(x_n) \\
\mathcal{A}^*(x) &= min\mathcal{A}^*_1(x_1), \mathcal{A}^*_2(x_2), ..., \mathcal{A}^*_n(x_n)
\end{aligned} \tag{2.3}$$

$$\begin{aligned}
\mathcal{B}^*_1(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_1(x) \wedge \mathcal{B}_1(y)] \\
\mathcal{B}^*_2(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_2(x) \wedge \mathcal{B}_2(y)] \\
&\vdots \\
\mathcal{B}^*_m(y) &= \vee_{x\in\mathcal{U}}[\mathcal{A}^*(x) \wedge \mathcal{A}_m(x) \wedge \mathcal{B}_m(y)] \\
\mathcal{B}^*(y) &= \mathcal{B}^*_1(y) \vee \mathcal{B}^*_2(y) \vee ... \vee \mathcal{B}^*_m(y)
\end{aligned}$$

where x_j is the input value (j = 1, 2, ..., n) and $\mathcal{B}^*_i(y)$ is the intermediate result of each 'IF-THEN' rule (j = 1, 2, ..., n). The operators $\wedge$ and $\vee$ take the minimum and maximum values of the membership functions, respectively; $\mathcal{B}^*(y)$ represents a composite fuzzy set of output decision preferences.

2.2.4 Defuzzification Method

To transformed the result of fuzzy rule which is also fuzzy subset degree, the fuzzy value should be transformed back to the crisps values or the real value through the defuzzification process. The accuracy of the centroid method in defuzzification process is the key to transform the fuzzy value to real output. The formula is given as:

$$y_{final} = \frac{\int_{\mathcal{V}} \mathcal{B}*(y)ydy}{\int_{\mathcal{V}} \mathcal{B}*(y)dy} \tag{2.4}$$

where y_{final} is a final output of fuzzy inference system.

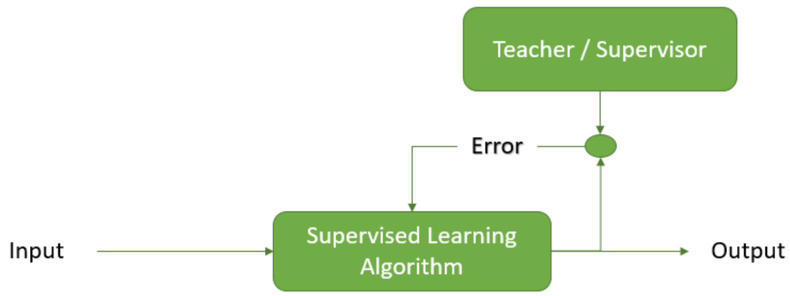
2.3 Reinforcement Learning

Learning is one of the abilities that substantially equivalent to understanding and adapting to the environment. These abilities should include into the human-awareness navigation. Machine learning, one of the sub-fields of artificial intelligence, is the method that enables the robot to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models. Machine learning tasks are typically classified into several broad categories: supervised learning, unsupervised learning and reinforcement learning (RL).

Supervised learning is the method that learns from the example that given by a "teacher". Its goal is to learn a general rule that maps the labeled inputs to labeled outputs. These can be seen mostly in classification problems [47]. Classification problems are the problems that ask the algorithm to predict or identify the input data as the member of a particular class or group. For example, in a training data set of animal images, that would mean each photo was pre-labeled as cat, koala or turtle. The algorithm is then evaluated by how accurately it can correctly classify new images of other koalas and turtles [48]. This type of learning method is suited to the problem that input and output data sets are known and need the algorithm to sort the data.

On the other hand, sometimes the data set is not easy to label or classify, therefore, unsupervised learning is used to answer the which criteria or model that can be used to classify the data set. The well-known method in unsupervised learning is deep learning model [49]. This model has handled a data set without explicit instructions on what to do with it. The training data set is a collection of examples without a specific desired outcome or correct answers. The neural network then attempts to automatically find structure in the data by extracting useful features and analyzing its structure. The unsupervised learning model can organize the data in different ways such as clustering, anomaly detection, association or autoencoders. Therefore, unsupervised learning is suited to the problem that labeled data is too difficult to get. Therefore, unsupervised learning rein to find patterns that can produce high-quality results.

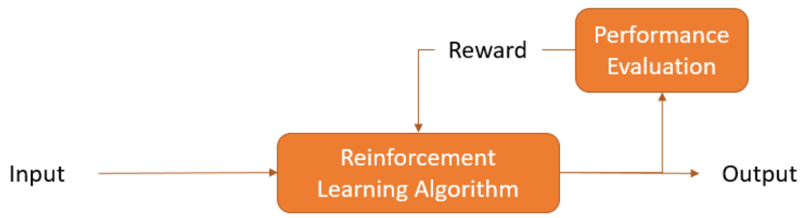
Another learning which is not in supervised or unsupervised type is the reinforcement learning (RL). RL is the process that agents or robots need to learn what to do and how to map situations and actions so that they can gain the maximum reward. The agent will try to discover by itself to choose the action that yields the reward in each situation. This action affects not only the immediate reward but also the next situation and all sub-sequence reward. RL imitate how humans and animals learn. Therefore, the machine tries a bunch of different things and is rewarded when it does something well. The structure of these three machine learning can be



(a) Supervised Learning

tb

(b) Unsupervised Learning



(c) Reinforcement Learning

Figure 2.6 The structure of tree type of machine learning. Supervised learning has the teacher or supervisor to help evaluate the output to the algorithm 2.6a while unsupervised learning does not need one 2.6b. Reinforcement learning is the learning algorithm that learn from the experience. It does not need the supervisor to evaluate but evaluate it self by response or feedback from the system in term of reward or punishment 2.6c.

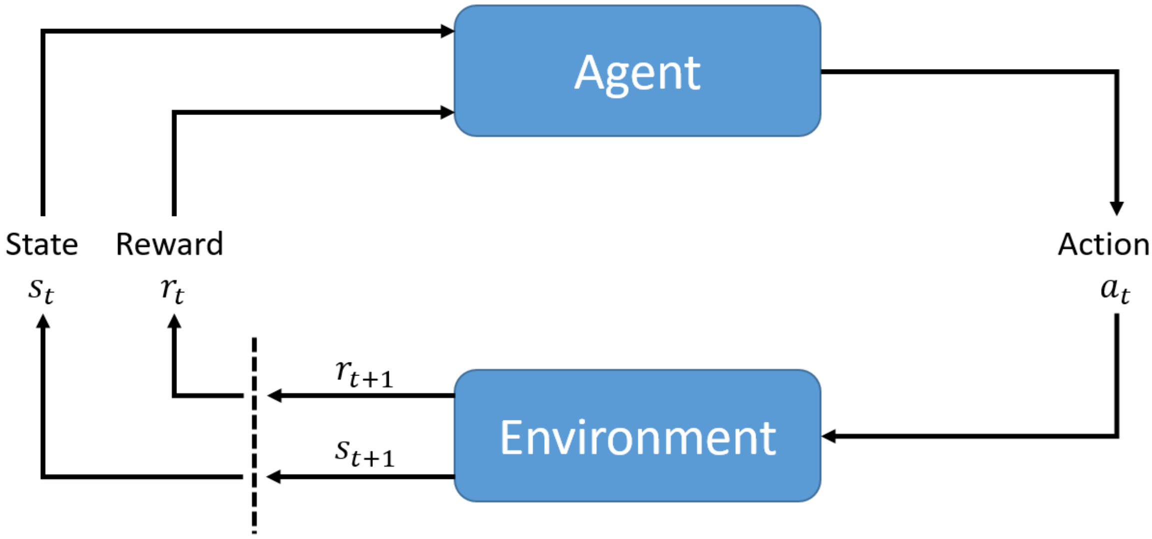


Figure 2.7 The agent-environment interaction in reinforcement learning.

summarized as in the figure 2.6

In this research, robots should understand humans' information and should learn to adapt their performance according to humans' feedback. In this case, Reinforcement learning (RL), which is one of the machine learning techniques, played the role to give the robot able to learn from its experience throughout the humans' interaction. Thus, robots are able to improve their performance without explicit programming from roboticist. Before proceeding to examples of reinforcement learning research, it is vital to understand the element of reinforcement learning. To illustrate the RL concept, the process of agent-environment interaction can be shown in Figure 2.7.

The agent and the environment interact at each of a sequence of time steps. At each time step t, the agent receives some presentation of the environment's *state* $s_t \in S$, where S is the set of possible states or situations, and on that basis selects an *action*, $a_t \in A(s_t)$, where $A(s_t)$ is the set of actions available in state s_t. One time step later, in part as a consequence of its action, the agent receives a numerical *reward*, $r_{t+1} \in R$, and finds itself in a new state, s_{t+1}.

2.3.1 Markov Decision Process

The general concept of RL can be described with the Markov Decision Process (MDP) framework. An MDP is a tuple $\langle S, A, T, R, \gamma \rangle$, of a set of states, actions, transitional probabilities, reward, and discount factor.

- S is a set of states
- A is a set of actions
- T is a probability function taht describe the transition over states.
- R is a reward function representing the expected amount of feedback, given current state s_t, action a_t and next state s_{t+1}.
- γ is a discount factor that keeps the expectation finite in the case of an MDP without terminal states, where $\gamma \in [0,1]$

The concept of MDPs is that the agent chooses the action a_t in the state s_t and waiting for the feedback or reward r and the change of situation or next state s_{t+1} from the environment. For the environment that can be modeled as a MDP, the goal of the agent in the envirment is to maximize the expected reward over time. The most common criteria are:

- Finite-horizon model where the agent tries to maximize the sum of rewards for the following M step:

  $$\mathbb{E}\left\{\sum_{k=0}^{M} r_{t+k} | s_t\right\} \tag{2.5}$$

  The aim is to determine the best action, considering there are only M steps to collect rewards.

- Infinite-horizon discounted reward model where the agent tries to maximize the reward at the long-run but favoring short-term action:

  $$\mathbb{E}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t\right\}, \gamma \in [0,1] \tag{2.6}$$

  The discount factor γ represent the important of interest of the agent. A γ close to 1 gives long-term action similar importance to short-term action, but if γ close to 0, the short-term action is more important.

- Average reward model where the agent tries to find the actions that maximize the average reward on the long-run:

  $$\lim_{M\to\infty} \mathbb{E}\left\{\frac{1}{M}\sum_{k=0}^{M} r_{t+k} | s_t\right\} \tag{2.7}$$

This model makes no distinction between policies which take reward in the initial phase from others that shoot for the long-run reward.

2.3.2 Function to Improve Behavior in RL

The agent is expected to progressive collect more rewards which in M step, therefore, actively learning by reinforcement. Each state is followed by an action, which leads to another state and corresponding reward. The objective function to collect more reward or maximized the reward can be formulated as the state value or action-value function which are described as the follows:

- **State Value Function**: The state value function can be defined as the expected sum of rewards following the distribution of actions given states, called policy(π). The state value can be defined as:

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}\left[G_t | s_t\right] \tag{2.8}$$

where G_t is the cumulative reward that can be written as a sum and adding the discount factor to make future reward less important and make sum finite in the continuous problem. The discount reward G_t can be defined as:

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{2.9}$$

The state value function can expressed recursively with the Bellman expectation equation as:

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}\left[r_t + \gamma V^{\pi}(s_{t+1}) | s_t\right] \tag{2.10}$$

- **Action Value Function**: The action value function can be decomposed similarly as the state action value function, as shown as:

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{\pi}\left[G_t | s_t | a_t\right] \tag{2.11}$$

and to the recursive form obtained with the Bellman expectation equation can be defined as:

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{\pi}\left[r_t + \gamma Q^{\pi}(s_{t+1}, a_{t+1}) | s_t, a_t\right] \tag{2.12}$$

To solve the RL problem, the optimal policy that achieves a great amount of reward in the long-term should be defined. There are multiple policies to solve the same problem, some better than others but there is always at least one policy better than all the others. This is called an optimal policy denote by π^*, and it will have an optimal state value function V^* which defined as:

$$V^*(s_t) = max_\pi V^\pi(s_t) \tag{2.13}$$

for all $s \in S$. An optimal policy also has an optimal action value function, denote by Q^* and defined as:

$$Q^*(s_t, a_t) = max_\pi Q^\pi(s_t, a_t) \tag{2.14}$$

for all $s \in S$ and $a \in A$ This Q^* function gives the expected return for taking action a_t in the state s_t and thereafter following the optimal policy π^*.

2.3.3 Reinforcement Learning Related Works

Reinforcement learning is slowly gaining some stance in the agent simulation field. Its varied use in the area shows its real versatility. It can solve many learning problems and implement for learning of the high-level decision. In following works, reinforcement learning techniques are used to solve some learning problem but approaching differently. The beautiful of RL is that its framework can adapt to whatever it is we desire to model, as long as, its elements are defined for such a purpose.

Cuayahuitl et al. [50] present an approach to include the adaptive behavior of route instructions. They proposed a two-stage approach to learn a hierarchy of wayfinding strategies using hierarchical reinforcement learning. Their experiments were based on an indoor navigation scenario for a building that is complex to navigate. Their results showed adaptation to the type of user, and the structure of the spatial environment, plus the learning speed was better than the baseline approaches they used.

Using RL to train a virtual character to move participants to a specified location was introduced by Kastanis and Slater [51]. The states for the agent were four distances from the avatar to the participant. This states based on the *Proxemics* theory. The agent has six actions that involved working forward, backward, stay and wave. The reward function was based on the response of the human. If the person moved towards the target, the agent got a positive reward and in all other case got the punishment. The results showed that the agent did learn the rules that when the agent moves too close to the participants, they will tend to move backward.

Social Learning for the population of agent coordinate on social optimal outcome in the context of general-sum gateways proposed by Hao and Leung [52]. RL is used as a learning strategy instead of evolutionary learning. The results showed that the agents were able to achieve stable coordination on socially optimal outcomes and suitable for both the settings of the symmetric and asymmetric game.

Gil et al. [53] presented a calibration method for a framework based in multi-agent RL. The agents learned to control its instantaneous velocity vector individually in scenarios with collision and frictions forces. The results indicated similarity in the learned dynamics of the agent with the real pedestrians.

The different way to use reinforcement learning can be seen in [54]. The model of the human social system was model by RL to constructing normative agent. The case study was focused on using this architecture to predict trends in smoking cessation resulting from a smoke-free campus initiative. The agents learn from their interactions with other agents, their judgment is their reward and socially correct behavior is learned.

As we can see from these works, RL has been present in many areas for some years. Its potential has already explored, but its improvement has been proved to give satisfying results. Our work presents another way of using RL to parameter adaptation of the method for estimate human personal space and studying the efficacy of RL algorithms to determine the appropriate RL algorithm for the real-world task.

2.4 Transition Based Rapidly-Exploring Random Tree

The Transition-based RRT (T-RRT) algorithm is a sampling-based approach to a cost space path planning which has been extended from the Rapidly Random Tree (RRT). It takes advantage of two approaches. The first approach is the exploration strength of the RRT algorithm to grow random trees toward the unexplored area. The second approach is the feature to accept or reject the potential state of stochastic optimization methods. It has been applied to various cost-space path planning problem, for example, [55–58] and computation structural biology [56, 59].

T-RRT integrates a stochastic transition test to explore the low-cost space by accepting or rejecting a new candidate configuration based on the cost variation. The algorithm of T-RRT has been shown in Algorithm 1.

The **transitionTest** is used to accept or reject the planning path between two configurations

Algorithm 1 Transition-Based Rapidly Random Tree

Input: the cost-space $\boldsymbol{C}$

Parameters: the tree $\mathcal{T}$

Initialize: $\mathcal{T} \leftarrow$ initTree($node_{init}$)

while not stoppingCriteria($\mathcal{T}$) **do**
    $node_{rand} \leftarrow$ sampleRandomConfiguration($\boldsymbol{C}$)
    $node_{near} \leftarrow$ findNearestNeighbor($\mathcal{T}$, $node_{rand}$)
    if refinementControl($\mathcal{T}$, $node_{near}$, $node_{rand}$) **then**
        $node_{new} \leftarrow$ extend($node_{near}$, $node_{rand}$)
        if $node_{new} \neq$ null and transitionTest($\mathcal{T}$, cost($node_{near}$), cost($node_{new}$)) **then**
            addNewNode($\mathcal{T}$, $node_{new}$)
            addNewEdge($\mathcal{T}$, $node_{new}$, $node_{near}$)
        end if
    end if
end while

from their cost. The level of selectivity of this function is controlled by the adaptive parameter, called temperature T. With low temperature, the T-RRT algorithm will limit the expansion to gentle slopes of cost space. Conversely, with high temperature, the T-RRT algorithm enable the planning path to climb steep slopes. In T-RRT, the temperature is dynamically tuned during the search process, based on the current number of consecutive rejection $nFail$ and on the maximal number of consecutive rejection $nFail_{max}$. For example, after each accepted uphill move, T is decreased to avoid over-exploration on the high cost-space. More precisely, T is divided by the temperature adjustment factor $alpha$, and $nFail$ is reset to 0. On the other hand, after each rejection uphill move, different actions are possible depending on the value of $nFail$. If $nFail$ is less than $nFail_{max}$, the temperature remains unchanged, but $nFail$ is incremented by 1. If $nFail$ is greater than $nFail_{max}$, T is increased to facilitate exploration and avoid being trapped in local minima. More precisely, T is multiplied by the temperature adjustment factor α, and $nFail$ is reset to 0. The reansitionTest function can be seen in Algorithm 2.

The objective of the **refinementControl** function which has shown in Algorithm 3, is to limit this refinement and facilitate tree expansion toward unexplored regions of space. The idea is to reject an expansion that would lead to more refinement. If the number of refinement nodes already present in the tree is higher than a certain ratio ρ of the total number of nodes, a refinement node is defined as a node whose distance to its parent is less than the expansion step-size δ. Another benefit of refinement control is to reduce the computation cost of nearest-neighbor search by limit the number of nodes in the tree.

Algorithm 2 transitionTest(tree $\mathcal{T}$, cost($node_i$) c_i, cost($node_j$) c_j)

Input: the maximum cost c_{max}

the current temperature T

the temperature adjustment factor α_T

the current number of consecutive rejction $nFail$

the maximum number of consecutive rejction $nFail_{max}$

Output: *true* if the transition is accepted, *false* otherwise

if $c_j > c_{max}$ **then**
  return *False*
else if $c_j \leq c_i$ **then**
  return *True*
else if rand(0,1) < exp(- (c_j - c_i) / T) **then**
  $T \leftarrow T/\alpha_T$
  $nFail \leftarrow 0$
  return *True*
else
  if $nFail > nFail_{max}$ **then**
    $T \leftarrow T \cdot \alpha_T$
    $nFail \leftarrow 0$
  else
    $nFail \leftarrow nFail + 1$
  end if
  return *False*
end if

Algorithm 3 refinementControl(tree $\mathcal{T}$, $node_{near}$, $node_{rand}$)

Input: the extesion step-size δ_T

the refinement ratio ρ_T

Output: *true* if the refinement is not too high, *false* otherwise

if dist($node_{near}$, $node_{rand}$) < δ_T and nbRefineNodes($\mathcal{T}$) > $\rho_T \cdot$ nbNodes($\mathcal{T}$) **then**

  return *False*

else

  return *True*

end if

Chapter 3

Human Personal Space Model Estimation

This chapter presents the proposed method that applies social information of each human to model individuals' personal space. Many of previous work model human's personal space basing on one social information like human's appearance or human's body width. However, these parameters are almost similar and assume to be the common parameters to all humans in the environment which make the robot to estimate similar personal space which has the same size to everyone. This affects the performance of approaching or avoiding humans. The generated path which based on the personal space cannot determine the optimal path length and optimal path cost like discomfort feeling.

Therefore, this research proposed the method that estimates the individual's personal space base on their social information such as genders, perception range and relationship level between humans and the robot. The following part of this chapter moves on to describe in greater detail the estimation of the human personal space model. First, the overall process of our proposed method is described to understand the whole process for human-aware navigation. Then the basic social model that uses an asymmetric Gaussian function to generate human personal space with cost is introduced to use as the cost map for robot navigation. After that, human's states and social factors will be described to design the parameters for the Gaussian function. These parameters affect the size and shape of the frontal and lateral spaces of each human. Finally, the comparison results of path length that generated from the personal space with common parameters and our proposed method, fuzzy socialmodel, will be shown to see the effectiveness of using social factors to model human personal space.

3.1 Overall Process of Human Aware Navigation

In human aware navigation framework, the robot has to operate in a shared environment with humans and should consider two constraints: safety, and social constraint. The safety constraint is the constraint that the robot should consider to prevent itself from the harmful situation. The social constraint is the constraint that helps the robot to avoid situations that potentially make discomfort to humans. Both constraints can be defined as the cost of navigation that enables the robot to operate more natural and acceptable.

The overall view of the human aware framework can be illustrated as in Figure 3.1. Assume that the robot knows the environments such as obstacles or free space which is the static environment. From this static environment, the robot can generate the static cost-map which is the safety constraint for navigating in the environment. For the environment that humans exist, the robot considers the social constraint which navigates the robot capable of generating a socially competent path from the perceiving social information such as the human state or human social factors. The robot can generate the path based on this combination of both constraints. This generated path enable the robot to operate more natural and acceptable for humans in the environment.

However, to formalize the social constraint from the social information is the big challenge for the roboticist. Therefore, this research proposed the personal space estimation method that determines the social constraints or social cost-map basing on the humans' social factors.

3.2 Asymmetric Gaussian Model

personal space model of the human is derived according to the interactive human spaces between person and person in the environment. These spaces depend on the human state and social signal, i.e., position, motion, relationship, and so on. The asymmetric Gaussian distribution function is used to integrate that information into a mathematical model and used as the cost for path planning process. The personal space of human group is derived based on this function of each person.

$$F(x, y) = \sum_{i=1}^{n} f_i(x, y) \tag{3.1}$$

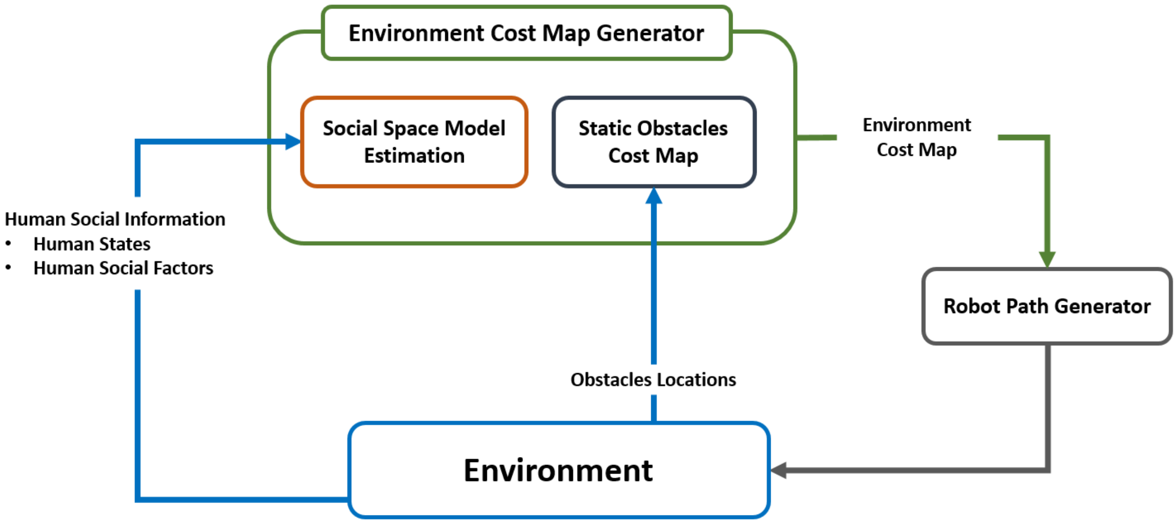


Figure 3.1 The overall process of the proposed human personal space model estimation

where n is the total number of persons, f_i is the repulsive force originating from the i-th person which can be expressed by the asymmetric Gaussian distribution function.

As shown in Figure 3.4, a person p_i located at coordinate (x_i, y_i), and its direction is θ_i. Social cost surrounding the person p_i can be defined by using two dimensional Gaussian function as:

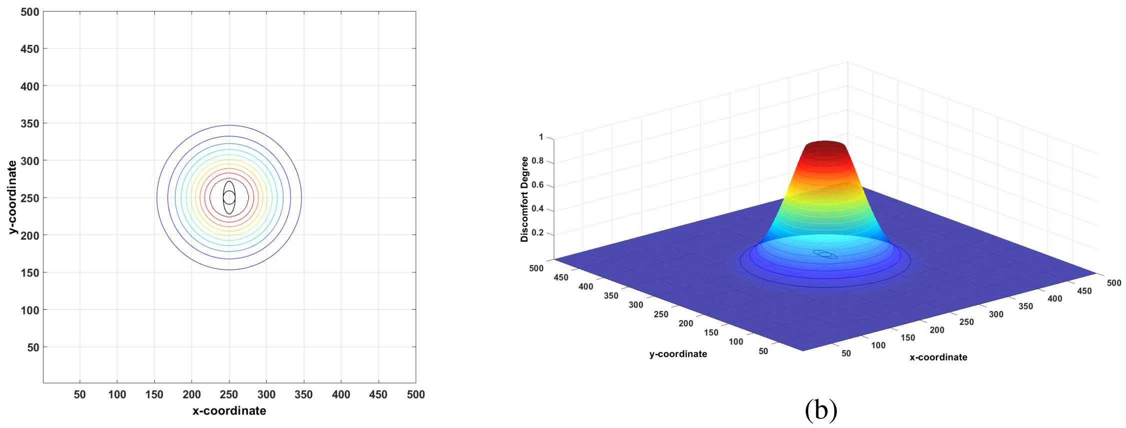
$$f_i(x, y) = A * \exp\left(-\left(\beta_{fr} - \beta_{la}\right)\right) \tag{3.2}$$

which has its maximum at the human's center (x_i, y_i) and decreases when faraway from center. A is the magnitude of the repulsive force which can be determined by a person's physique. β_{fr} and β_{la} are factor of personal space in front and side of human respectively, which can be defined as:

$$\beta_{fr} = \frac{(d * \cos(\theta - \theta_i))^2}{2 * \sigma_f^2} \tag{3.3}$$

$$\beta_{la} = \frac{(d * \sin(\theta - \theta_i))^2}{2 * \sigma_s^2} \tag{3.4}$$

where θ_i is human direction. σ_f and σ_s are variances which can be modified according to human states and social factors of humans relate to the robot. Figure 3.2 shows the example social map that determined by the social model function above.



(a)

(b)

Figure 3.2 The personal space model from asymmetric Gaussian function.

3.3 Fuzzy Social Model

Social interaction space is the spatial space relation between humans described by Hall [10]. The space around the person can be divided into four areas according to the distance to the person. These areas can be described as behavior and relationship between a person and others which depend on comfortable of the individual's sentiment. To model social interaction space, human's social information are integrated to model human's social area. There are many social factors that can be used for social space modelling. Many of previous method applied only specific one or two social factors which are not enough information and produced the imprecise social space. This research apply more variety of social factors such as, position, motion relationship, genders and perception range to generate efficient human's social space model.

The structure of the proposed estimation method can be visualize as Figure 3.3. During the human-robot interaction, the robot perceives the human states and social factors, such as positions, velocities, orientation, human-robot relationship level and human genders to design of human's personal space. These human states and social factors are used to formulate the personal space in frontal and lateral of the human respectively. Humans usually decide to avoid the area in the front of others depends on the speed of them. This concept can be used to define the frontal personal space of humans. Social signals and cues are used to be the social factors to determine the lateral space of the humans. These social factors are the crisp set of input data which gathered for the fuzzy inference system. These crisp set are converted to a fuzzy set using fuzzy linguistic variables, fuzzy terms, and membership functions. Afterward, an inference is

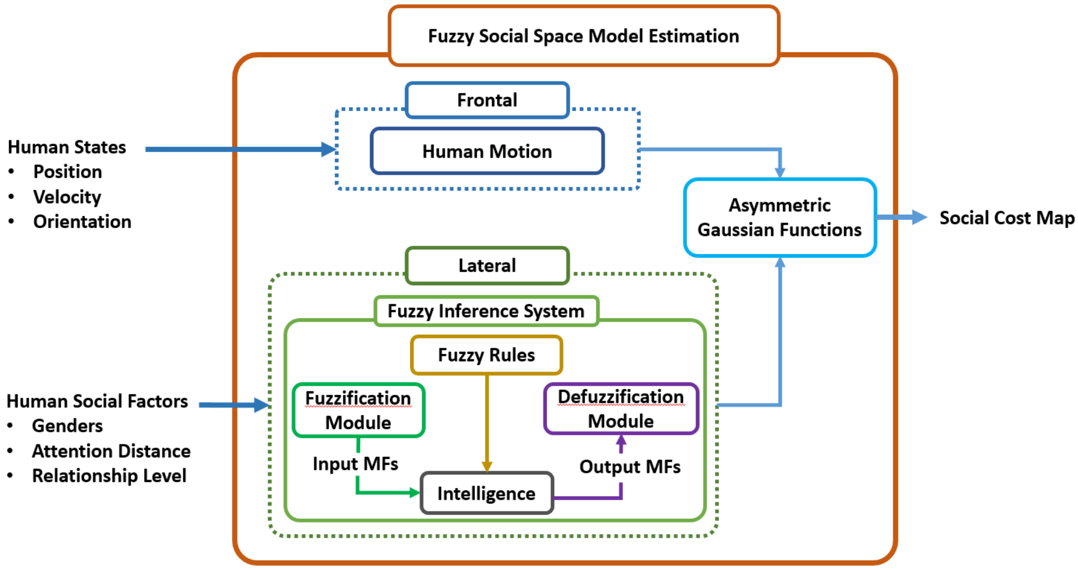


Figure 3.3 The overall process of the fuzzy socialmodel for human social model estimation

based on a set of fuzzy rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the output membership function, in the defuzzification module. The output from the fuzzy inference system is the parameter to determine the personal space model of the human which can be calculated by the asymmetric Gaussian function. Based on this human's social model, the robot can estimate the social cost map that includes humans' personal space. This social cost map will be combined with the static cost map. This combination cost map will be used in the path planning to generate the socially navigation paths to perform social interactions with humans.

3.3.1 Human States

The robot must be aware of the human position and movement in the shared work space. In reality, to avoid the collision with other persons in the environment, the human decides to move to avoid the area in the front of others depends on their speed. From this concept, the robot should avoid the frontal area of humans rely on their speed to make them more comfortable when working in the shared environment.

Let $H = (p_1, p_2, ..., p_n)$ be the number of people detected by the robot by the robot sensing and perception capacity. Human states of a person p_i is represented as $p_i = (x_i, y_i, \dot{x}_i, \dot{y}_i, \theta_i)$ where the location of the human is the coordination (x_i, y_i), $\dot{x}_i$ and $\dot{y}_i$ are the velocity on x and y direction respectively, and θ_i represent the orientation of the human reference to the world frame. This

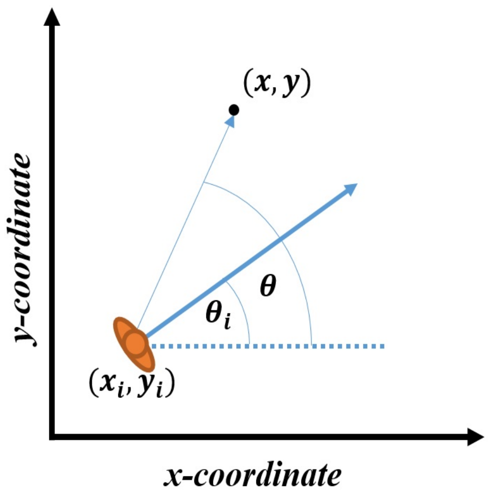


Figure 3.4 Human state which consists of position, orientation and velocity reference to the world frame

human's states can be shown in Fig.3.4.

The magnitude of human motion can be computed by:

$$v_i = \sqrt{\dot{x}_i^2 + \dot{y}_i^2} \tag{3.5}$$

The value of variance (σ_{fi}) of the personal space model in the frontal and rear of human can be calculated as:

$$\sigma_{fi} = \begin{cases} \sigma_{f0}, & \cos(\theta - \theta_i) \leq 0 \\ \sigma_{f0} / \left(1 + f_f v_i\right), & \text{otherwise} \end{cases} \tag{3.6}$$

where σ_{f0} is chosen according to personal space of a person. θ is the angle of the environment position (x, y). f_f is the normalization factor. $\cos(\theta_r - \theta_i)$ can be evaluated the area where in the front of or the rear of the person. Therefore, the personal space in the front of the human will be larger than the rear.

3.3.2 Social Signals & Cue For Lateral Personal Space Estimation

Human interaction and behaviors are involved and govern by not only intended goal, physical constraints, and other humans in the environment, but also social characteristics. For front and rear of the human, the personal space of the human is depended on the state of the person as describe in the section above. However, for the lateral space of the human, the personal space should base on the social signal and social cues between the person and others.

Social signals and cues are considering as the factors to model human lateral personal space. Social cues are consisted of five categories which are physical appearance, gesture and posture, face and eyes, voice, and environmental space, while Human social signals are defined as a communicative, informative signal that provides information about social facts such as social interactions, a social emotion, social relationship.

In our research, we are considering the factors from human social signals and cues, such as the human's gender, the perception range of the human and the level of relationship between the human and the robot, to design the lateral space according to the Proxemics concept that introduced by Hall [10]. Therefore, the reason and related work about these social factor are decribed as follows.

Human Genders

Genders of humans are the social cue that can use to design behaviors or interactions. Various research support that humans perform difference action depending on the different genders of interacting people [60]. Research of gender differences has been quite substantial. There are several studies demonstrate the interaction of humans pair to the interaction distances. For example, Hayduk's review counts seven studied supporting that men choose greater distances than women [60]. In [61], found that female pairings preserved the smallest distance, followed by male-female pairing, and male-male pairing was most distant.

Genders of humans also influence proxemic behavior between women and men could be dominance and submission [62]. It might be that females, as well as males, maintain greater distances from males because men are labeled as more dominant, and their personal space is more respected than the personal space of women. Studies confirm that persons approaching a man stay more distant than the person who approaches a woman.

However, when humans are interacting with non-human agents such as robots and avatars in an immersive virtual reality setting, the proxemic behavior of females and males is opposite to the studies of the human-human interaction [23]. For instance, The experiment in virtual reality setting shows the result that females were less comfortable moving close to virtual agents which controlled by software, than avartar that controlled by a person, whereas males did not differentiate between both type of controllers [63]

Therefore, in this research, genders are used to design the human personal space according to the studies of human and non-human agent that mentioned above. From the above studies, females have larger personal space than males because they have less comfortable to interact with non-human agents. Therefore, for the robot to approaching or navigating around humans, the robot should consider this information of humans to protect itself from interrupting human's comfort zone.

Perception Range

The perception of the person with human's sensing organ like eyes or ears can be used to identify the locations, movement or appearance of other existences in the shared environment. The quality of human perception also depends on the distance between the person and the others [64]. For example, when the person in the room with other two people, the person can provide the information of the nearest person more than the further person because the near distance enables the person to perceive clear information more than the far distance.

For human's emotion, humans tend to form very different mental representation on the different perception distance. As the experiment in [65], they implement by separate two group of participants to administer a pain decision task involving two different sizes of face stimuli painfully. These sizes manipulate the retinal size and perception distance. They observe the empathy reaction from the participants. The result shows that the empathy reaction was absent in the group of participants that receive the smaller stimuli that corresponding to the face stimuli that observed from the far distance. The experiment results show the effect of perception distance of the humans to their emotion.

The concept of perception distance can be seen in [5]. Sisbot et al. introduced a human aware navigation framework that uses the ability of humans' perception combine with safety concept to model the human's comfortable space. The concept that humans generally feel more comfortable when the robot is in their perception range, in their experiment is in the human's field of view. The results make the robot to be mostly in human's field of view during the motion.

In this research, the concept of the relation between perception distance and human emotion is used to estimate the human personal space. During the robot operation, if the robot moves closer to the human, the personal space or the area of privacy of the human should be smaller because humans perceive more quality information of the robot that makes them more comfortable. In another word, the personal space of the human will get more significantly when the robot moves far away due to less quality of information make less comfortable.

Level of Relationships

Human behavior is mostly based on social relationships, which can be in the form of a dyadic relationship, where there is a complex peer relationship among different individuals. [10] already pointed out that how humans feel to others with a different relationship. Many studies of human-human interaction confirm that the level of relationship affects the proxemic behavior of human. For example, [60] confirmed that human familiar with the other person leads to smaller interpersonal distance. The signal to like other person is used to study the interaction distance between human and human [66]. Their study shows that humans use close interaction distance when they signal to like other people. They also show that humans who are familiar with other interact more proximally.

For human-robot interaction, the level of relationship can be assumed as the experience between the human and the robot. The experiment in [67] presented that humans who have prior experience with robots also tend to approach closer to it in a subsequent interaction. [23] also investigate to prove the hypothesis that experience in robotics will decrease the personal space that people maintain around the robot.

Therefore, this research takes into account of human's experience of the robot to estimate human personal space. The designed social model estimation enable the robot to estimate the social model of the familiar person smaller than the stranger person.

3.3.3 Mapping Social Factors to personal space Estimation

This research incorporates a fuzzy logic system into an estimation of social interaction space framework. The fuzzy logic performs as the mapping tool to map between social factors and the variance of the lateral social interaction space. In our research, social factors, such as the gender of the human, the human perception range and the human-robot relationship level, have been considered as a linguistic variable and used as the inputs of fuzzy inference systems. The output of fuzzy logic is the value of variance for lateral social interaction space. These output sets are set according to '*Proxemics*' theory; Near personal area (NPA), Far personal area FPA, Social area (SA) and Public area (PA). According to the description of social factors and social interaction space, the fuzzy sets of inputs has been determined as follows:

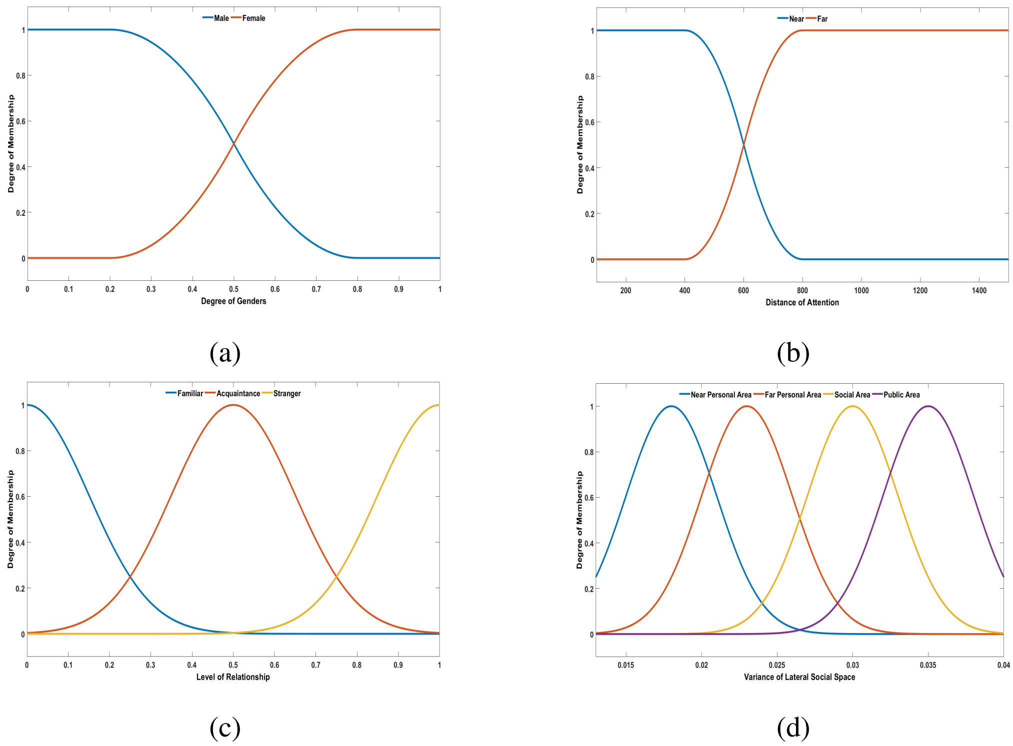


Figure 3.5 Input MFs; Degree Genders(a), Perception Range(b), Level of Relationship(c). The output MFs; the variance of lateral personal space(d).

- **Genders** is the genders of the human that the robot can estimate from detection method. The input MF of gender is defined as a sigmoid function subject to male (M) and female (Fe). Let r_g be the estimation value of gender input, a_g the steepness of the distribution of genders, and c_r the inflection point. Then the MFs of the genders is given as follows:

$$\Gamma_1(r_g; a_g, c_g) = 1/\left(1 + exp\left(-a_g * (r_g - c_g)\right)\right) \tag{3.7}$$

- **Perception Range** is the relative distance between the human and the robot which can be divided into two sets such as near ($Near$) or far (Far). It is represented by a sigmoid function. Let r_d be the input of the relative distance, a_d the steepness of the distribution of relative distance, and c_d the inflection point. Then the MFs of the relative distance is given as follows:

$$\Gamma_2(r_d; a_d, c_d) = 1/(1 + exp(-a_d * (r_d - c_d))) \tag{3.8}$$

- **Relationship Level** describes the personal knowledge or experience with the robot which can be set by three Gaussian functions, familiar (Fam), acquaintance (Acq), and stranger (Str). Let r_i be the relationship degree that the robot perceives from people. Therefore,

Table 3.1 Designing the social interaction space using fuzzy rules

Input			Output
Gender (Γ_1)	Perception Range(Γ_2)	Relationship Level.(Γ_3)	Interaction Area $\mathcal{N}(\mu, s^2)$
M	Near	Fam	NPA
M	Near	Acq	FPA
M	Near	Str	SA
M	Far	Fam	NPA
M	Far	Acq	FPA
M	Far	Str	PA
Fe	Near	Fam	FPA
Fe	Near	Acq	SA
Fe	Near	Str	SA
Fe	Far	Fam	FPA
Fe	Far	Acq	PA
Fe	Far	Str	PA

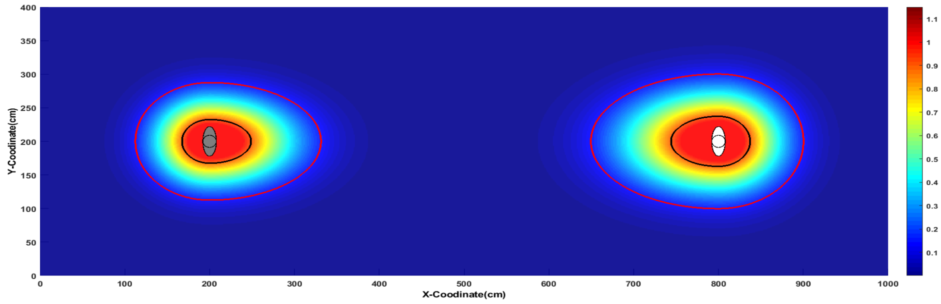
the relationship degree MFs are given as follows:

$$\Gamma_3(r_i) = \begin{cases} \mathcal{N}\left(\mu_{Fam}, s^2_{Fam}\right) & \textit{if Fam} \\ \mathcal{N}\left(\mu_{Acq}, s^2_{Acq}\right) & \textit{if Acq} \\ \mathcal{N}\left(\mu_{Str}, s^2_{Str}\right) & \textit{if Str} \end{cases} \tag{3.9}$$

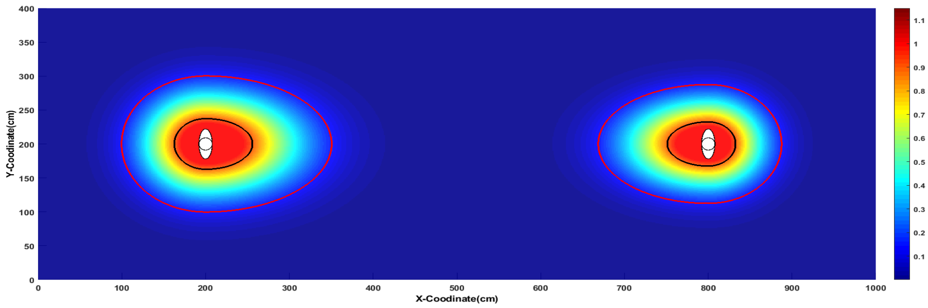
For the output of the fuzzy logic, there are several ranges in the human interaction area according to the theory of Proxemics [10]. The distance of human interpersonal space inspires us to estimate the social interaction space of the human. Therefore, the concept of different parameters in determining the different social model for each person is chosen related to these interpersonal space concept. The fuzzy sets of outputs has been determined as follows

- **Variance of lateral space** can be separated into four sets. Therefore, four Gaussian functions are used to represent a change of variance(σ_{si}) in each social interaction area which is defined as:

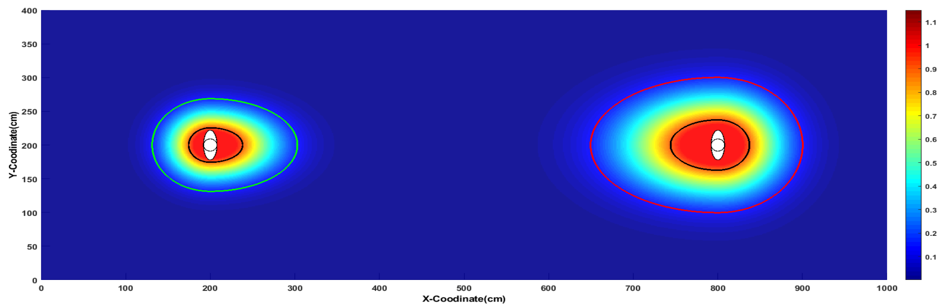
$$\sigma_{si} = \mathcal{N}\left(\mu, s^2\right) = \begin{cases} \mathcal{N}\left(\mu_{PA}, s^2_{PA}\right) & \textit{if PA} \\ \mathcal{N}\left(\mu_{SA}, s^2_{SA}\right) & \textit{if SA} \\ \mathcal{N}\left(\mu_{FPA}, s^2_{FPA}\right) & \textit{if FPA} \\ \mathcal{N}\left(\mu_{NPA}, s^2_{NPA}\right) & \textit{if NPA} \end{cases} \tag{3.10}$$



(a)



(b)



(c)

Figure 3.6 The designed human personal space depends on different genders (a), perception range (b) and relationship level (c)

The input and output membership functions can be illustrated in figure 3.5. For fuzzy rules, we construct a rules base containing fuzzy rules on relationship between social factors and the lateral variance according to Proxemics theory. A detailed description of the proposed fuzzy rule is shown in Table 3.1. Combining the above-mentioned social factors, β_{si} can be defined as follows:

$$\beta_{si} = \frac{(d * \sin(\theta - \theta_i))^2}{2 * \mathcal{N}\left(\mu, s^2\right)^2} \tag{3.11}$$

This means that, to prevent the robot from intruding onto the human social interaction space, the robot is required to delineate the dynamic boundary of interaction areas based on the human

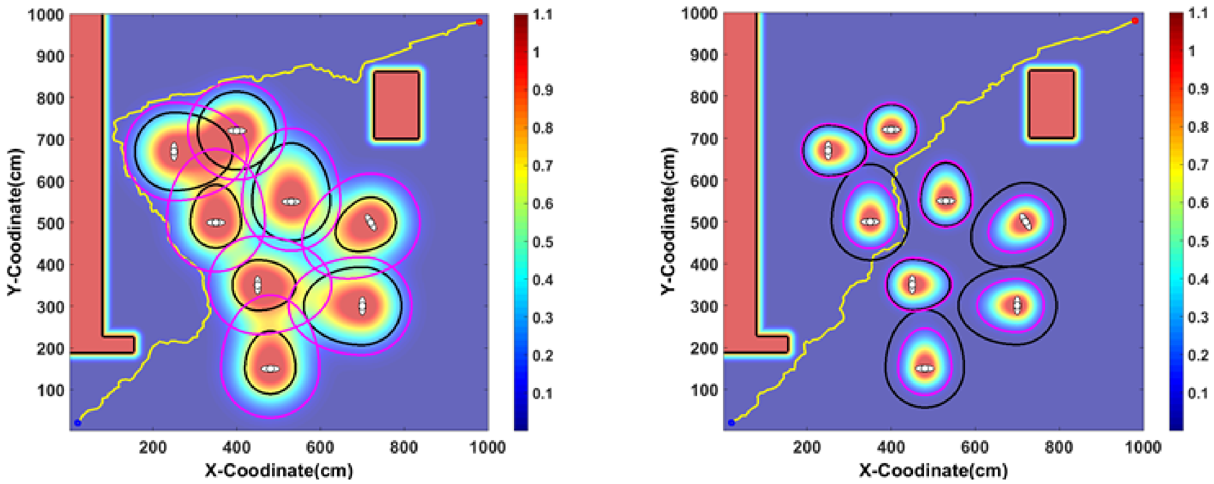


Figure 3.7 The cases of estimate personal space with common parameters. The ground truth personal space regions are presented by black line. The purple line represent the estimate personal space with commonparmeter.

social factors.

The proposed social model of the human which using social factors can be shown in figure 3.6. For the same level of relationship, both humans are stranger (red line), and the perception range between humans and the robot are the same, male (grey body) provides smaller personal space than female (white body). For the same genders of the humans, and the level of relationship between humans and the robot are the same, the robot position is set at x = 200 and y = 50. Therefore the result of the personal space of near-person is larger than faraway human. For the last social factor, by the same genders and perception range, the boundary of the familiar person (green line) is smaller than the stranger person (red line)

3.4 Effectiveness Comparison

This section shows the influence of the proposed social model to the path planning like Transition-based Rapidly Random Tree(T-RRT) which is the cost-map based path planning algorithm. T-RRT has abilities to explores and finds the optimal path cost in the large space. The scenario to simulate is that the robot tries to generate a path from the starting point at the bottom left to the ending point at the top right without colliding and intruding into humans personal space. The robot will collect the path cost from human ground truth due to the position of the generated path.

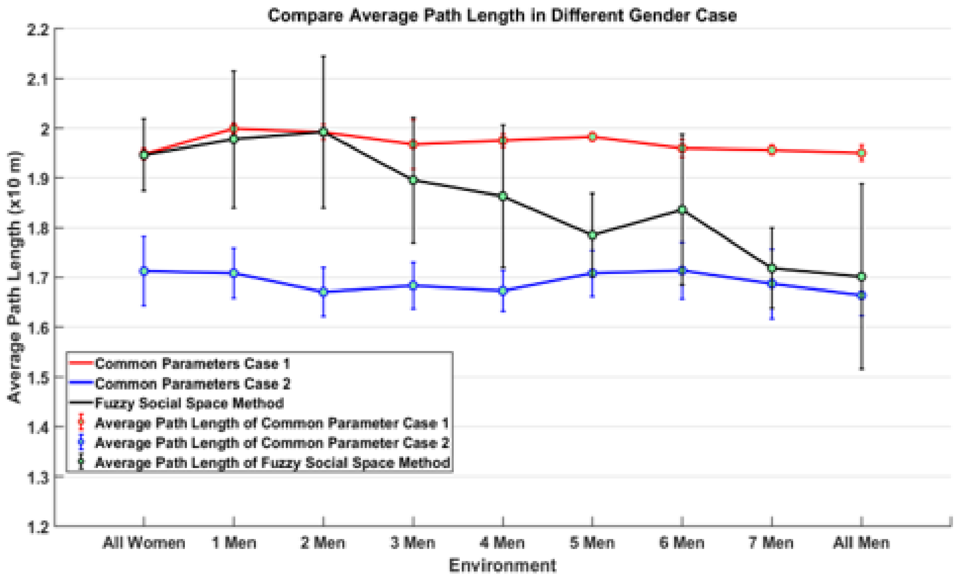
The results show the impact of the proposed method compares to the common parameters method. Here the comparison results will compare the proposed method and common parameters method in different cases.

The first case is T-RRT generate path based on estimate personal space that determined by the common parameters. This estimate personal space is larger than ground truth personal space. The second case is the estimate personal space is smaller than the ground truth personal space. This two cases can be visualized in Figure 3.7

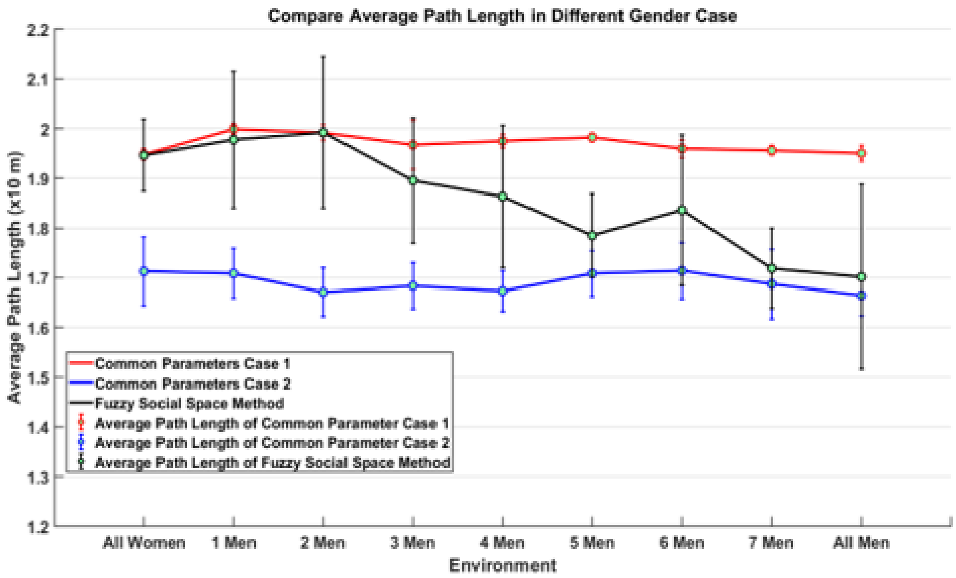
In the environment, there are humans with different genders, different level of relationship. Therefore, the behavior of T-RRT based on each social factor is compared with common parameters.

The results show that the estimate personal space with common parameters gives similar path length and unacceptable degree even the number of familiar person or number of males is changed. With the small estimate personal space, T-RRT can provide optimal path length. However, the results of total unacceptable degree that T-RRT collect is very high. This means that the robot can generate short path but the path intrude into human's private space and make humans discomfort. However, with the large estimate personal space, the collect unacceptable degree is decrease and approach to zero, but the generated path is long due to less free space. This means the generated path avoid to intrude into human personal space but it may be far from the human and difficult to interact with. The results of common parameters can be seen in Figure 3.8 and 3.9.

By using the proposed fuzzy personal space model estimation, T-RRT performs adaptively to the change of environment. For example, the change of the number of the familiar person in the environment. With all stranger in environment, the estimate personal space is big to avoid interrupt human's comfort. So, path length is long. However, when human get more familiar to the robot or the number of the familiar person increase. The path length is decreasing. Therefore, the fuzzy personal space model can enable the robot to estimate the individual's personal space that can provide more space to generate short path length while considering human's unacceptable degree, as shown in Figure 3.9 black line. The results of change of number of genders is similar, as shown as Figure 3.9.

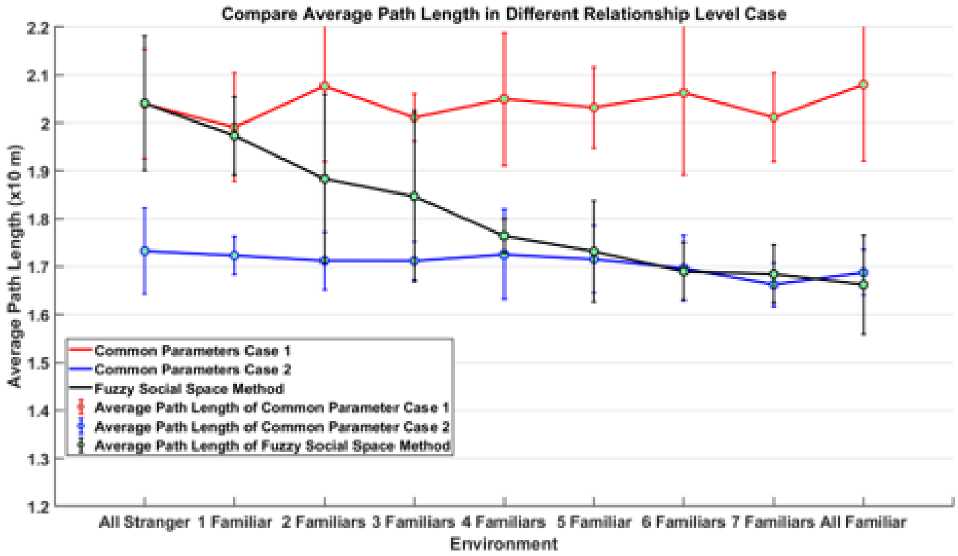


(a)

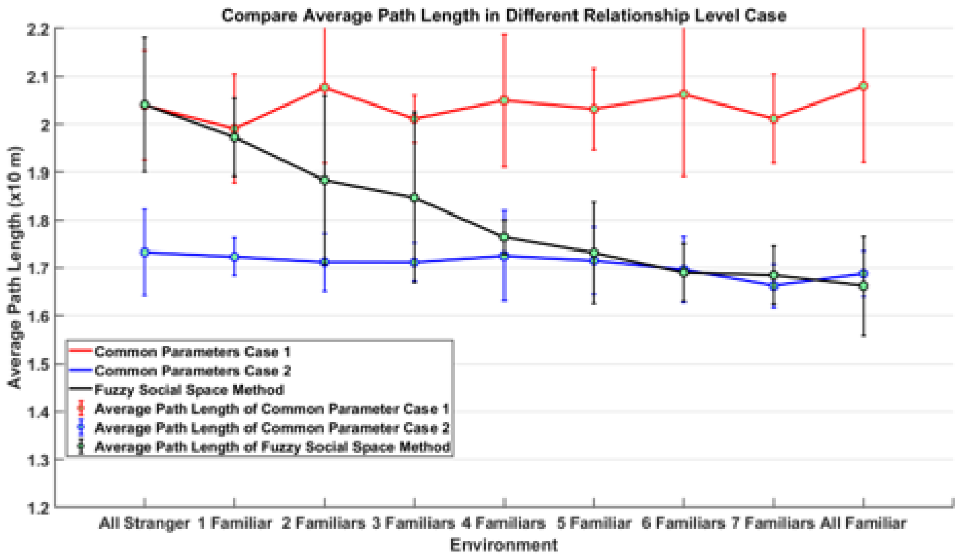


(b)

Figure 3.8 The case of the number of men changes from no men to all men in the environment



(a)



(b)

Figure 3.9 The case of the number of familiar person changes from no familiar person to all familiar person in the environment

3.5 Summary

In this chapter, the designed human personal space model is presented. This human model provides the size and cost that can be used to assist the navigation algorithm. The shape of the social model is not symmetric between frontal, rear and lateral. Asymmetric Gaussian function plays the role to model human personal space. However, the parameters of the function depend on human factors.

The concept that human avoids to intruding into the frontal area depends on their speed. Therefore, human state, which includes position, orientation, and velocity, is used to determine the parameter for model frontal personal space. The lateral space is designed based on social factors. These factors consist of social signals and cues. Social signals are defined as a communicative, informative signal or cue that, either directly or indirectly, provides information about social facts such as social emotion or social relationship. Social cues are from humans such as physical appearance, gesture, and posture. Many of previous works used only one social information and used it as the common parameters for asymmetric Gaussian function.

The proposed method consider three social factors of each humans to model individual's personal space. However, these social factors are difficult to detect and vary depending on the variety of conditions. Therefore, a fuzzy logic approach is used to quantify these parameters.

The proposed model can be used to assist navigation algorithm like T-RRT which is cost-based navigation algorithm. The results show that our proposed model assist T-RRT to generate the optimal path length and optimal unacceptable degree of human.

Nevertheless, the designed membership functions in fuzzy inference system may not suit or satisfy to every group of humans. This problem cause the accuracy of human social model, as shown in figure 3.10. The problem can cause the robot to intrude into human privacy area or maneuver out of the interaction range. To solve this problem, the learning ability is important for the robot to learn and adjust it parameters via the the response of humans during the interaction.

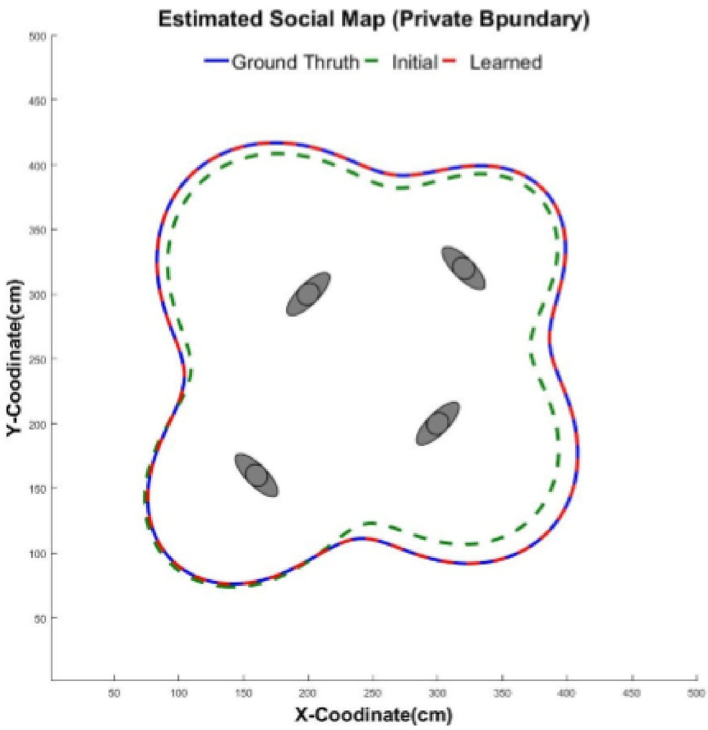


Figure 3.10 Comparison of private area boundaries. Realistic personal space boundary (*blue solid line*), estimated social boundary with primary setting (*green dash line*)

Chapter 4

Learning Fuzzy Social Model

From the previous chapter, the fuzzy social model uses the fuzzy inference system to maps the crisp values of detected human's social factors to the variance values of the asymmetric Gaussian function that relates to the proxemics theory. The output of the fuzzy social model is used as the cost function in the path planning algorithm. This generated path is used to guide the robot to interact with or avoid humans in a shared space. However, the design MFs, in the fuzzy inference system, are designed from the knowledge-based which may not satisfy every group of humans which have different social information. This incorrect design MFs cause the robot to behave in the way that disturbs human feeling like intruding into the privacy area or moving far away from the interaction range of humans.

Therefore, from the incorrect design mapping tool's parameter problem, the method to modify parameters should be applied to recreate personal space. Carefully calibrate model parameters is one of the useful methods which enables the user to set the appropriate parameters for themselves. However, it requires the technical knowledge to modify which is difficult for the normal user. Therefore, machine learning techniques are the appropriate way to give the learning ability to the robot. The robot can learn and modify the parameters automatically. Here, reinforcement learning is chosen to apply to solve incorrect pre-design MFs automatically because it automatically learn from the experience of interacting with humans and does not require any database.

This chapter contributes the study of the reinforcement's algorithms that integrate to fuzzy inference system to adapt the parameter of MFs. By using this learning fuzzy social model, the robot is able to estimate and modify the estimate individual's personal space according to an individual's response.

The comparison of RL's algorithms in term of learning time is presented to see which algorithm is suited to the problem. The results also show that the RL algorithms can modify the MFs and convert the estimate social model to similar to the real human's social model. This learning fuzzy social model allows the robot to gain the maximum reward which means it avoids to intruding the private area of human and in the interaction quality range.

4.1 Fuzzy Social Model Estimation To Reinforcement Learning

Our research aims to modify the MFs that used to estimate the human social model, more accurately by learning while interacting with humans in a shared environment. This learning fuzzy social model enables the robot to correctly estimate human personal space that help the robot to avoid intruding into human's area of privacy and also keep distance for good quality of interaction. This research applied reinforcement learning as the tools to modify the MFs.

With the fuzzy social model estimation, the robot can estimate the human personal space by estimating from the pre-designed MFs. The robot can generate the path to approach or avoid the human based on this social model. However, with pre-designed MFs, the robot may estimate human personal space smaller or larger than realistic which effect to the human's feeling and emotion due to the movement of the robot. The robot receives this humans' response detection method like the verbal/non-verbal reaction or some detection method like face detection or emotion detection method [68]. This response information can be used as the reward or punishment for RL algorithms. The RL algorithms will modify the MFs by increasing or decreasing MFs value until the robot receive the maximum reward from humans. Figure 4.1 shows the overall process of learning fuzzy social model estimation. To express this fuzzy social model estimation into reinforcement learning framework, the element of reinforcement learning such as states, actions and reward should be defined.

4.1.1 States-Space

This work focus on modifying the pre-design MFs, especially the level of relationship MFs, to maximize the reward that given by humans. Three Gaussian functions are used to design the

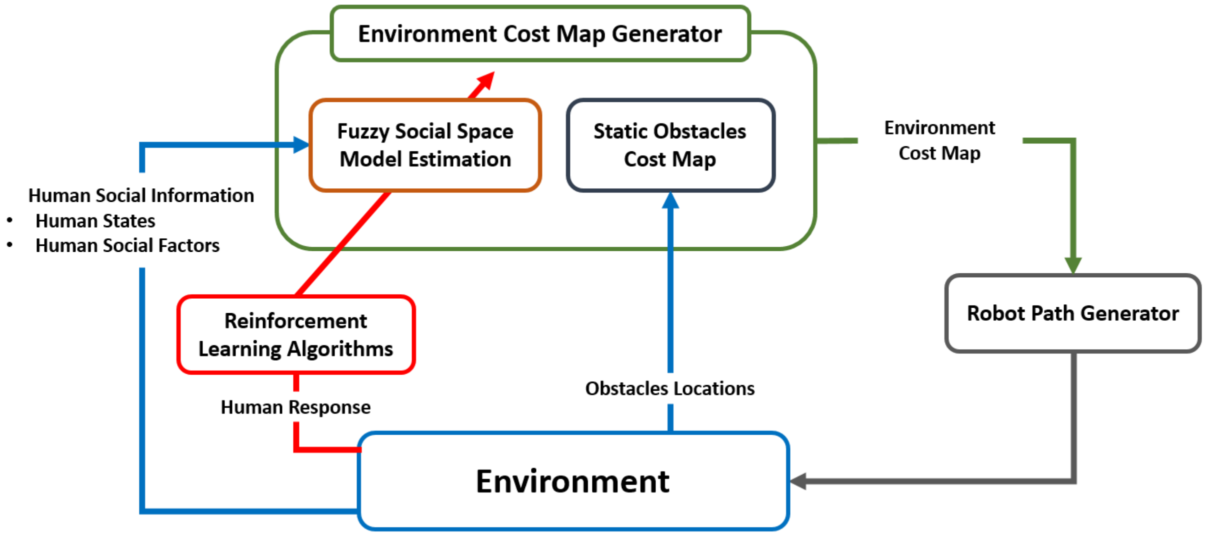


Figure 4.1 The overall process of the proposed human personal space model estimation

level of relationship MFs. The importance values of the Gaussian function are mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$. Therefore, states $s \in S$ for our problem which consists of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ can be defined as:

$$s = \left[\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right] \tag{4.1}$$

where $\boldsymbol{\mu} = [\mu_{Fam}, \mu_{Acq}, \mu_{Str}]$ and $\boldsymbol{\sigma}^2 = [\sigma^2_{Fam}, \sigma^2_{Acq}, \sigma^2_{Str}]$. However, we make the states more simple by make $\boldsymbol{\sigma}^2$ as the constant and modify only $\boldsymbol{\mu}$. Therefore, the state of our problem can be modified as:

$$s = \left[\mu_{Fam}, \mu_{Acq}, \mu_{Str}\right] \tag{4.2}$$

4.1.2 Action-Space

The actions $a \in A$ is the set of how to adjust MFs. In this work, means of MFs are modified by increasing, decreasing or staying at the same value. Therefore, the actions $\boldsymbol{a}$ can be defined as:

$$a = \left[a_{Fam}, a_{Acq}, a_{Str}\right] \tag{4.3}$$

where a_{Fam}, a_{Acq} and a_{Str} can be set as decreasing or increasing or do noting to the mean values $\boldsymbol{\sigma}^2$.

4.1.3 Reward Function

The reward function $R(s_t, a_t, s_{t+1})$ is defined as the response from environment where performing the action a_t in the st ate s_t that lead to the state s_{t+1}. In this work after the robot first estimate the personal space of the human, the robot will move to approach each human by using the estimate personal space. However, because of the error of design, the estimated personal space may smaller or larger than actual human's personal space. Therefore, the reward is the response or feedback from humans. This response can be collected from the human's feeling which can be correctly detected by manually evaluation or emotion recognition process such as facial movement, heart rate, blood pressure, etc. However, in this thesis, we assume that the human's feeling depends on the distance from human centers.

Here, the reward is designed from two different areas that used to evaluate our social model estimation. First, the quality interaction area where humans can be engaged in high-quality interaction with the robot. At this area, interaction degree (ID) is designed to be the evaluation function that describes the easiness of interaction. Second, the private area where humans do not want to interfere with the robot speech or action. This area provides the degree of discomfort feeling or an unacceptable degree (UD) of humans that affected from the robot behavior. The ID and UD are increasing and decreasing respectively depending on the distance between the robot and the humans. Another factor in designing the reward is the difference of estimate path length and optical path length $\triangle l$. This factor related to the generated path that the robot generate. If the difference of path length is too large, that means robot generate path far from the human, and if the difference of path length is too small, that means robot generate path close to the humans. This different path length prevents the robot from generating path far away from humans.

The ratio ID and UD is used to determine the reward. Reinforcement learning techniques try to modify parameters of mapping tool in fuzzy social model to minimize interaction degree and minimize unacceptable degree and different path length. Therefore, the estimate personal space will let the robot to interact or approach human in the area of interaction quality area but out side the human's area of privacy.

The work aim to maximize the ID and minimize the UD and different path length $\triangle l$ while the robot is operating. Therefore, the designed reward function $R(s_t, a_t, s_{t+1})$ at state s_t for human-robot interaction can be defined as:

$$R(s_t, a_t, s_{t+1}) = \frac{k_1 * ID(s_t)}{k_2 * UD(s_t) + k_3 * \triangle l(s_t)} \tag{4.4}$$

where k_1, k_2 and k_3 are weight for each degree. This reward correspond to the ID, UD and $\triangle l$ at state s_t. To this end, the aim of our work is to determine the state that give the maximum reward:

$$s = \arg\max_{s} R \tag{4.5}$$

4.2 Reinforcement Learning Algorithms

In this section, RL's algorithms that used to integrate with our fuzzy social model are described. RL algorithms are applied to determine how to adjust the MFs. This section will start with the famous and basic RL algorithm, Q-learning which is the model-free based reinforcement learning. Then the average reward learning (R-learning) is described. This R-learning is similar to Q-learning but uses the average reward instead of the discount factor. The third is the policy gradient method, Actor-Critic, which use the loss function to adjust the learning parameter. Finally, the advanced RL method that combined with the deep learning technique and experience replay buffer. This method will use the buffer memory to store the previous leaning parameters and use it to update learning parameters.

4.2.1 Q-Learning

Q-learning is a popular and widely used technique of RL which learns to optimize the state-action value (Q-value). The state-action value represents the quality of actions in the successor state and makes the task of choosing an actor easier. This means that the robot will evaluate the quality of the actions in the state. When the robot comes back to the same state, the robot will select the appropriate action based on this quality value. The Q-learning pseudo code is shown as Algorithm 4

In this thesis, the tabular Q-learning is used to determine the action to modify MFs. The tabular Q-learning is using each cell of n states and m actions table to store the value of quality of each pair state-action. The initial values in each cell of Q-table usually set as zero. This table is called Q-table which depicted by Figure 4.2.

Algorithm 4 Q-Learning

Initialize: state-action value $Q(s,a)$

Loop Process

Choose action a in s using behavior policy (e.g. ϵ-greedy)

Take action a, observe R, next state s'

$\delta \leftarrow R + max_{a'} Q(s',a') - Q(s,a)$

$Q(s,a) \leftarrow Q(s,a) + \alpha\delta$

$s \leftarrow s'$

Actions

States	a_1	a_2	...	a_{m-1}	a_m
s_1	Q(1,1)	Q(1,2)	...	Q(1,m-1)	Q(1,m)
s_2	Q(1,1)	Q(2,2)	...	Q(1,m-1)	Q(1,m)
...	...	...	...	...	...
s_{n-1}	Q(n-1,1)	Q(n-1,2)	...	Q(n-1,m-1)	Q(n-1,m)
s_n	Q(n,1)	Q(n,2)	...	Q(n,m-1)	Q(n,m)

Figure 4.2 Q-Table of n states and m actions

By following the Algorithm 4, the beginning of the state s_t is set as the location of the pre-designed MFs means $\boldsymbol{\mu}$. The action a_t in this state s_t is selected from the values in Q-table by ϵ-greedy method which is the method to make the agent choose the random action with the probability ϵ and choose that action that has maximum Q-value for otherwise. After the action a_t is taken, the robot will observe the reward r_t from R function which corresponding to ID, UD and $\triangle l$ when MFs as the state s_t, and observe the next state s_{t+1}.

The Q-learning method aims to compute the optimal value $Q^*(s_t,a_t)$ for all state-action pairs, which can be done iteratively by collecting samples. Q-table is updated for every sample using the following update rules:

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \alpha\left[r_t + \gamma \max_{a'} Q_t(s_t,a') - Q_t(s_t,a_t)\right] \tag{4.6}$$

where α is the learning rate that allow for determination of the significance of new information in relation to the existing value. The motivation of this update equation is to calculate the difference between the predicted Q-value $r_t + \gamma \max_{a'} Q_t(s_t,a')$ and its current value $Q_t(s_t,a_t)$ for every sample.

Algorithm 5 R-Learning

Initialize: average reward ρ and state-action value $Q(s,a)$;

LOOP Process

Choose action a in s_t using behavior policy (e.g. ϵ-greedy)

Take action a, observe r_t, next state s_{t+1}

$\delta \leftarrow r_t$ - ρ + $max_{a'}Q(s_{t=1}, a') - Q(s_t, a_t)$

$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\delta$

if $Q(s_t, a_t) = max_{b'}Q(s_{t+1}, b')$ **then**

    $\rho \leftarrow \rho + \beta\delta$

end if

$s \leftarrow s'$

Once a certain amount of iterations is reached, the selected action process will select the action that gives the maximum reward. It means that after a certain amount of time, most of the quality value of state-action pairs are determined. Therefore, in the state, there is the action that can gain more quality to reach the state that has a maximum reward.

4.2.2 R-Learning

The R-learning algorithm, proposed by Schwartz [69], is the adaptation of Q-learning to the average criterion. The average criterion mostly applies to the continuous problem. In a continuous setting, there are no terminal state or some task also no start state, and no special time step which mean cannot define the task as the episode and the interaction between agent and environment goes on and on and forever without termination state. Therefore, this R-learning, which is the average reward RL algorithm, is chosen as an interesting algorithm to solve our parameter adaptation problem. The goal of R-learning is to learn an action a whose average reward ρ is as close as possible to the maximal average reward ρ^*.

In this thesis, the tabular R-learning is also used to determine the action to modify MFs. R-learning algorithm is similar to Q-learning but use and adaptive shifting values to approximate the optimal average reward to avoid the unboundedness of Q-value in state action pairs, as shown in Algorithm 5

By following the Algorithm 5, the beginning of the state s_t is set as the location of the pre-designed MFs means $\boldsymbol{\mu}$. The the action is selected from the values in Q-table by ϵ-greedy method. After the action a is taken, the robot will observe the reward r_t from R function and observe the next state s_{t+1}.

The R-learning method also aims to compute the optimal Q value $Q^*(s_t, a_t)$ for all state-action pairs, which can be done iteratively by collecting samples. However, instead of using discount factor for discounting the weight of the pass reward, R-learning using the average reward to update the Q value in Q-table. Therefore, Q value is updated for every sample using the following update rules:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\delta \tag{4.7}$$

$$\delta = r_t - \rho + \max_{a'} Q_t(s_t, a') - Q_t(s_t, a_t) \tag{4.8}$$

where ρ is the shifting value that approximately to the averga reward. This ρ will update every time that the action selection algorithm use greedy stratergy to select the action. The update of this shifting value can be determine as:

$$\rho = \rho + \beta\delta \tag{4.9}$$

where β is the update rate of the shifting value.

4.2.3 Actor-Critic

Actor-critic methods are RL methods that have a separate memories to represent the policy independent of the value function explicitly. The policy structure is known as the actor, while the estimated value function is known as the critic. Critic is used to evaluate the action that sellected by the actor. The critic must critique whatever policy is currently being followed by the actor. The critique takes scalar signal of an error to update all learning in both actor and critic, as suggested by Figure 4.3.

Actor-critic methods are extended the idea of comparison methods into reinforcement learning problem. Typically, the critic is a state-value function. After each action selection, the critic, which is a state-value function, evaluates the new state to determine whether things have gone better or worse than expected. That evaluation is the temporal-difference (TD) error:

$$\delta_t = r_t + \gamma\hat{v}(s_{t+1}) - \hat{v}(s_t) \tag{4.10}$$

where $\hat{v}$ is the value function implemented by the critic. This TD error can be used to evaluate the action just selected, the action a_t taken in state s_t. If the TD error is positive, it suggests

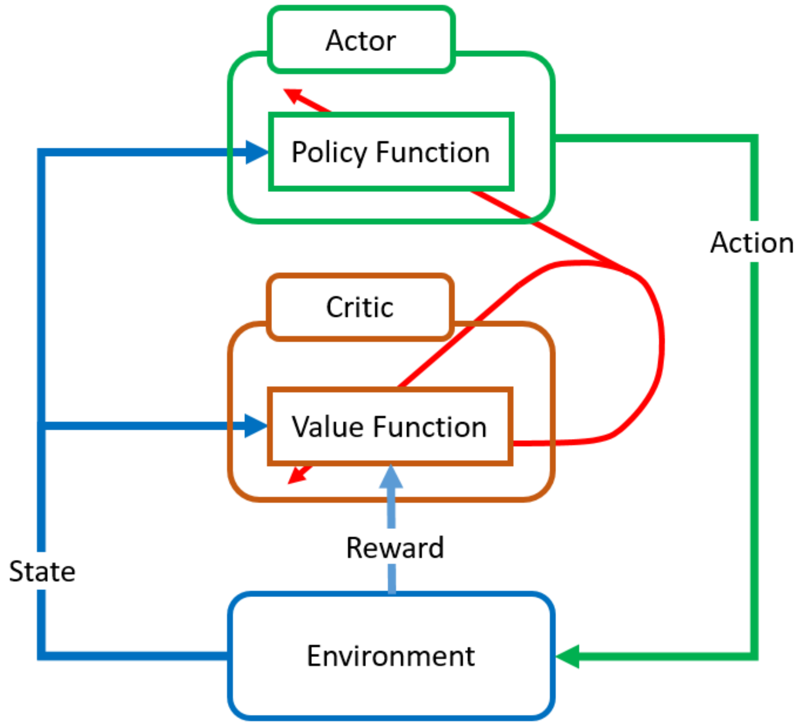


Figure 4.3 The actor-critic architecture

that the tendency to select a_t should be strengthened for the future, whereas if the TD error is negative, it suggests the tendency should be weakened.

The function approximation for both actor and critic which used in this thesis are two neural networks. According to algorithm 6, the action a_t are selected from the policy function which generate from the neural network $\pi(a_t|s_t,\boldsymbol{\theta})$. This indicating the tendency to select each action a_t when in each state s_t with respect to the weight $\boldsymbol{\theta}$. Then, the agent observe the reward r_t and the next state s_{t+1} by taking the action a_t. After that, the state values $\hat{v}(s_t,\boldsymbol{w})$ and $\hat{v}(s_{t+1},\boldsymbol{w})$ of state s_t and s_{t+1} are defined by critic process with neural network weight $\boldsymbol{w}$. These state values are used to determine the TD-error in equation 4.10. Because AC have two neural network that must be trained, it means that two set of weights ($\boldsymbol{\theta}$ for the actor and $\boldsymbol{w}$ for critic) must be optimized separately by using back propagation process of neural network:

$$\boldsymbol{w} = \boldsymbol{w} + \alpha_{\boldsymbol{w}}\delta \nabla_{\boldsymbol{w}} \hat{v}(s_t,\boldsymbol{w}) \tag{4.11}$$

$$\boldsymbol{\theta} = \boldsymbol{\theta} + \alpha_{\boldsymbol{\theta}}\delta \nabla_{\boldsymbol{\theta}} \ln \pi(a_t|s_t,\boldsymbol{\theta}) \tag{4.12}$$

where $\alpha_{\boldsymbol{\theta}}$ and $\alpha_{\boldsymbol{w}}$ are the learning rate. The process will be repeat until the action from the actor provide optimal TD-error from critic.

Algorithm 6 Actor-Critic

Input: a differentiable policy parameterization $\pi(a|s, \boldsymbol{\theta})$

a differentiable state-value parameterization $\hat{v}(s, \boldsymbol{w})$

Parameters: step size $\alpha_{\boldsymbol{\theta}} > 0$, $\alpha_{\boldsymbol{w}}$

Initialize: policy parameter $\boldsymbol{\theta}$ and state-value weights $\boldsymbol{w}$

Loop Process:

Choose action a_t from the function $\pi(a|s, \boldsymbol{\theta})$ using behavior policy (e.g. ϵ-greedy)

Take action a_t, observe r_t, next state s_{t+1}

$\delta \leftarrow r_t + \gamma\ \hat{v}(s_{t+1}, \boldsymbol{w})$ - $\hat{v}(s_t, \boldsymbol{w})$

$\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha_{\boldsymbol{w}} \delta \nabla_{\boldsymbol{w}}\ \hat{v}(s_t, \boldsymbol{w})$

$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha_{\boldsymbol{\theta}} \delta \nabla_{\boldsymbol{\theta}} \ln \pi(a_t|s_t, \boldsymbol{\theta})$

$s_t \leftarrow s_{t+1}$

4.2.4 Deep Reinforcement Learning

In a modern RL frame work, function approximation is the key for estimate the vale function or policy of reinforcement learning problem. In Deep Reinforcement Learning (DRL), Deep Neural Network is used as the function approximation for value function and policy. The advantage of DRL is that it can handle high dimensional of state-action pair space which is appropriate for our parameters adaptation problem.

Q-Learning (Section 4.2.1) with tabular scheme was suitable for small state-action space application. However, for large state-action space, it was not satisfied in term of learning. It was shown that DNN could be integrated into Q-learning and perform as the human level. For example, on Atari games, it used only pixels of a raw image. [70].

In Deep Q-learning (DQN), the Q-function is parameterized with a neural network with weighs θ. The network is trained by minimizing a loss function at every iteration i, given by

$$L_i(\theta_i^Q) = \mathbb{E}_{s\sim\rho_\pi(\cdot),a\sim\pi(\cdot)}(y_i - Q(s, a, \theta)) \tag{4.13}$$

where the target at iteration y_i can be defined as:

$$y_i = \mathbb{E}_{s_{t+1}\sim\varepsilon}\left[r_t + \gamma \max_{a'} Q(s_{t+1}, a', \theta)|s, a\right] \tag{4.14}$$

$\pi(\cdot)$ is the behaviour policy, $\rho_p i(\cdot)$ is the distribution of states under policy $\pi(\cdot)$ and ε refers to the environment.

Algorithm 7 Deep Q-Learning (DQN)

Initialize: replay memory D

state-action value function with random weight $Q(s, a, \theta)$

Loop Process:

```
Choose action $a_t$ from function $Q(s,a)$ using behavior policy (e.g. $\epsilon$-greedy)
Take action $a_t$, observe $r_t$, next state $s_{t+1}$
Store transition $(s_t, a_t, r_t, s_{t+1})$ in replay buffer $D$
Sample random transitions $(s_j, a_j, r_j, s_{j+1})$ from $D$
Calculate target $y_i$
if $s_{j+1}$ is terminal then
    $y_j = r_j$
else
    $y_j = r_j + \gamma \max_{a'} Q(s_{j+1}, a', \theta)$
end if
Train the Q-network on $(y_j - Q(s_j, a_j, \theta))^2$ using Equation (4.15)
$s_t \leftarrow s_{t+1}$
```

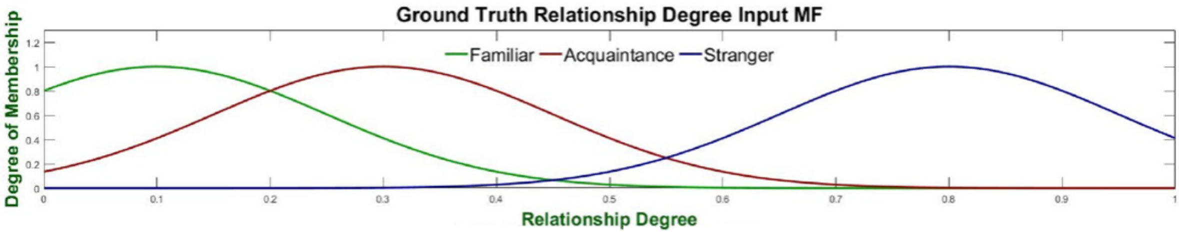
The gradient of the loss function is computed with respect to the weights to minimize the loss function. It is given by:

$$\nabla_{\theta_i^Q} L_i(\theta_i^Q) = \mathbb{E}_{s\sim\rho_\pi(\cdot),a\sim\pi(\cdot),s'\sim\varepsilon}\left[\left(r + \gamma \max_{a'} Q(s', a', \theta_{i-1}^Q) - Q(s, a, \theta_{i-1}^Q)\right) \nabla_{\theta_i^Q} Q(s, a, \theta_{i-1}^Q)\right] \tag{4.15}$$

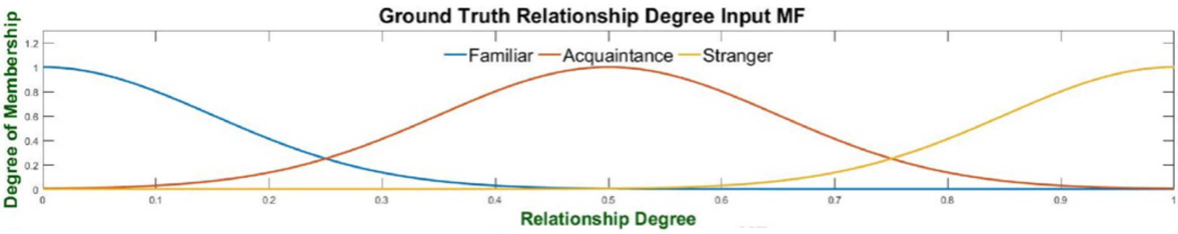
The loss function is minimized using stochastic gradient decent. The behaviour policy is an ϵ-greedy policy to ensure sufficient exploration.

Experience replay is the key to make DQN work. In general, DNN is easily over-fitting current episode. This due to the sample trajectories are correlated and it would lead the network would not be able to learn effectively.

In order to solve the correlation problem while training, experience replay stores the agent's experience at each time step (s_t, a_t, r_t, s_{t+1}) in the replay buffer. The information in this replay buffer are random extracted at every time instant and used to train the network. This make the network to learn more efficiency by increase learning speed from mini-batched and reduce correlation between experiences in updating DNN. The complete algorithm of DQN is given at Algorithm 7.



(a)



(b)

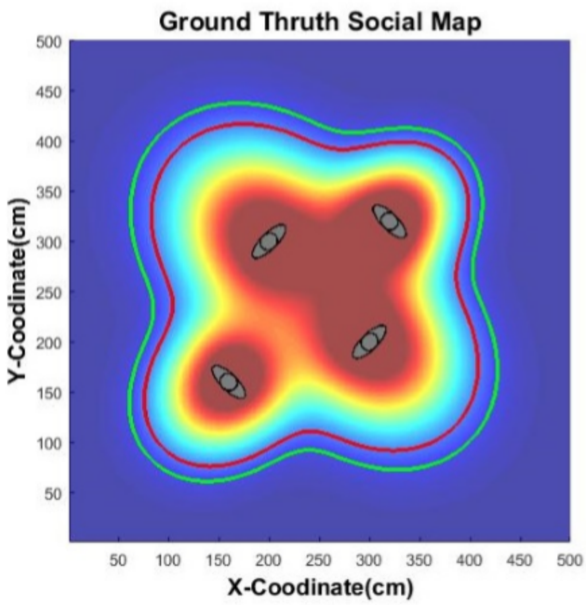
Figure 4.4 The different of ground truth relationship level MFs (4.4a) and initial setting of relationship MFs (4.4b)

4.3 Simulation and Results

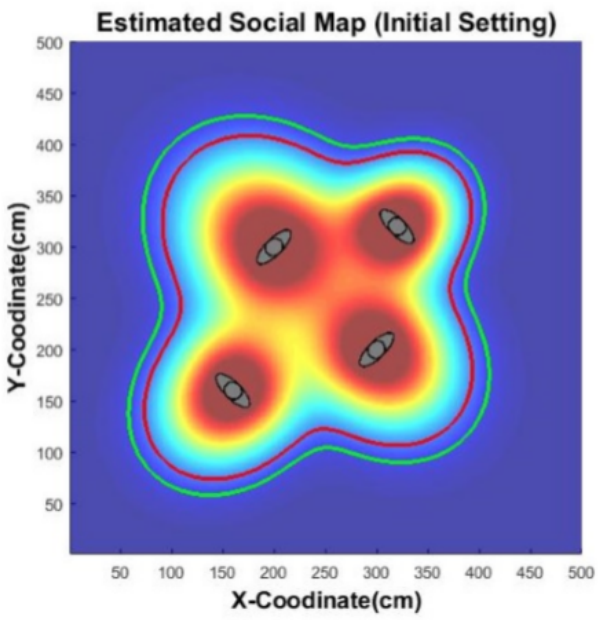
This section shows the results of our proposed strategy to correct the human personal space estimation method with a machine learning method. Here, learning fuzzy social model is introduced to estimate human personal space. By integrating the reinforcement learning with fuzzy social model, the robot has the ability to learn and modify the pre-design MFs to match to humans in a specific group of humans in the shared environment. Therefore, the proposed model assists the robot in planning paths to visit every person in the environment without trespassing on their private area, but to keep the distance from which people are able to have high-quality interactions.

4.3.1 Simulation Setup

At the beginning of this chapter, RL's algorithms are described and used as the learner for parameters adaptation problem. This section compares the efficacy of each algorithm to parameters adaptation problem in three properties. First is convergent of the algorithm that compares which algorithms can convert the social map error to a minimum. This property means that the estimated social model possible to modify and get similar to the realistic with RL algorithms.



(a)



(b)

Figure 4.5 The different of ground truth social map (4.5a) and initial estimate social map from our proposed fuzzy social model estimation (4.5b)

Second is the learning time which presents how fast that RL's algorithms use to modify the MFs' parameters until the social map error converts to the minimum. The third is the exploration rate of the state-action space which is the value to compare the algorithms how much state-action space has been explored. This value relates to the optimal action or best action that learning algorithm selected at each state.

The results of learning fuzzy social model that integrate RL to fuzzy social model estimation are also presented with different conditions, such as the face direction of humans or the number of humans in the environment, to show that our proposed method assist the robot to maximize interaction degree and minimize unacceptable degree in various conditions.

In the simulation, we assume that a geometric map is given or created by the robot. Our proposed model is to generate the social map by computing and updating social cost assigned to the geometric map. This social map is used to plan the robot navigation path in the environment. To validate the proposed model, we need to receive the reward from people. Therefore, the fuzzy social model is used with different parameters to model the ground truth social map which is used as the response or reward from humans for our proposed learning fuzzy social model.

The parameters to model the ground truth social map are set as follows: The parameters for relationship level MFs, according to equation (3.9), s_{Fam} = 0.15, μ_{Fam} = 0.1 to *Fam* set, s_{Acq} = 0.15, μ_{Acq} = 0.3 to *Acq* and s_{Str} = 0.15, μ_{Str} = 0.8 to *Str* set.

For our estimation method, the initial parameters of the relationship level MFs in equation (3.9) are designed as follows: s_{Fam} = 0.15, μ_{Fam} = 0 to *Fam* set, s_{Acq} = 0.15, μ_{Acq} = 0.5 to *Acq* set, and s_{Str} = 0.15, μ_{Str} = 1 to *Str*. These parameters can be adjusted by the learning process. Likewise the perception range MFs parameters, according to equation (3.7), are designed as follows: a_{Near} = -0.35, c_{Near} = 300 to *Near* set and a_{Far} = 0.35, c_{Far} = 300 to *Far* set. The relationship level MFs of ground truth and initial setting can be seen in Figure 4.4.

For the output function, the social interaction area is split into four Gaussian sets. The parameters of equation (3.10) are as follows: μ_{PA} = 0.035, s_{PA} = 0.005, μ_{SA} = 0.045, s_{SA} = 0.005, μ_{FPA} = 0.0035, s_{FPA} = 0.06, μ_{NPA} = 0.0035, s_{NPA} = 0.065. These parameters are decided based on the human interaction area concept [40] which determined the range of an individual's interpersonal space with different social factors when the robot approached the person. Reflecting their results, the parameters for the output membership functions can be determined. The different characteristic of ground truth social map and an initial estimated social map of males with different relationship levels to the robot can be seen in Figure 4.5. The figure shows that with initial setting, our fuzzy social model estimate incorrect personal space of each person which effect to incorrect social map (smaller than ground truth). If the robot uses this estimated social map to planning the path, it will cause human discomfort feeling. Therefore, RL is used to modify the MFs until the estimate social map is similar to the ground truth.

4.3.2 Efficacy of RL's Algorithms

For the reinforcement learning process, the discrete states s in equation 4.2 which consist of three mean values of each relationship MF are set as the value between 0 to 1, increasing by 0.1. The total number of states is 10^3. The action set of each MFs (equation 4.3) is defined as stay, increase, decrease, *i.e.*, 0, +0.1, -0.1. The total number of the action set is 3^3. Therefore, the total state-action space is equal to 10^3x3^3 = 27000 state-action spaces. The goal to use the RLs is to modify the MFs through iterative learning processes until gaining a maximum reward signal.

The RL's algorithms that are chosen to compare in this thesis are Q-learning(Q), R-

RL ALgorithms	Learning Time (Iteration)	Converge Value		Exploration Rate (%)
		Social Map Error (10^{-3})	Reward Error	
Q-Learning (Q)	**2390**	**2.2723**	**4.486**	**26.96**
R-Learning (R)	**2112**	**7.6772**	**8.160**	**28.09**
Actor-Critic(AC)	**2581**	**1.3968**	**4.497**	**43.05**
Deep Q-Network (DQN)	**1840**	**1.7872**	**4.755**	**29.56**

Table 4.1 Summary Efficacy of Reinforcement Learning Algorithms

learning(R), Actor-Critic(AC) and Deep Q-Learning(DQN). The results of social error and reward value of each algorithms through the simulation are shown in Figure 4.6.To evaluate the efficacy of RL's algorithms, learning time, converge value, and exploration rate are used as the criterions.

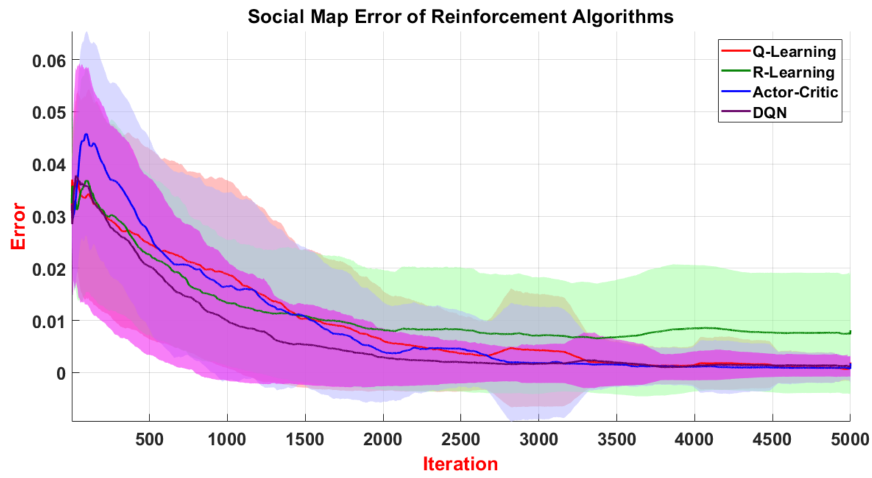
The first criterion, the learning time is used determine the time that the algorithm requires to reach and remain within design percentage of change. In this simulation result, learning time is determined from start to the time that error of personal space or reward oscillate with in five percent of minimum value. The results show that DQN is reqiure less learning time which compare to others algorithm. DQN used less learning time 17.6% compare to R-learning, 23% compare to Q-Learning, and 28.7% compare to Actor-Critc.

For the second criterion, the converge value is used to determine the final value of social map error that compare between estimated and ground truth social map, and final value of reward. The converge value of social map error and reward can be determine at the end of learning time. The results present that Actor-Critic which is policy-gradient method converge to smallest value of social map error and approach to gain highest reward. It converge less than DQN, Q-Learning and R-Learning at 21.8%, 38.5%, 80.8% respectively.

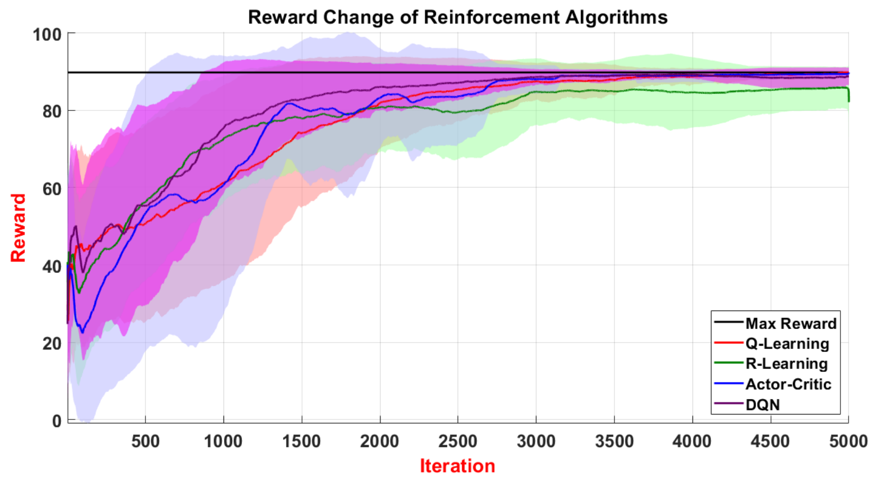
For the last criterion, the exploration rate can be determined by the percentage of the explored state-action pair (N_{exp}) to total number of state-action pair (N_{Total}), as follows:

$$ExplorationRate = \frac{N_{exp}}{N_{Total}} x100 \tag{4.16}$$

where N_{Total} is equal to 27000 state-action pairs. The results shows that Actor-Critic explore 48% of all state-action space which mean the algorithm is possible to determine the best solution



(a)



(b)

Figure 4.6 The error of social map that compared between the estimated and ground truth social map (4.6a) and the reward change of each algorithm (4.6b)

or action in each state. Actor-Critic has overcome Q-Learning, R-Learning, DQN by 59.7%, 53.2% and 45.6% respectively. The results of efficacy of reinforcement's algorithms can be summarized in Table 4.1.

Table 4.2 Results of learning social model with the number of people

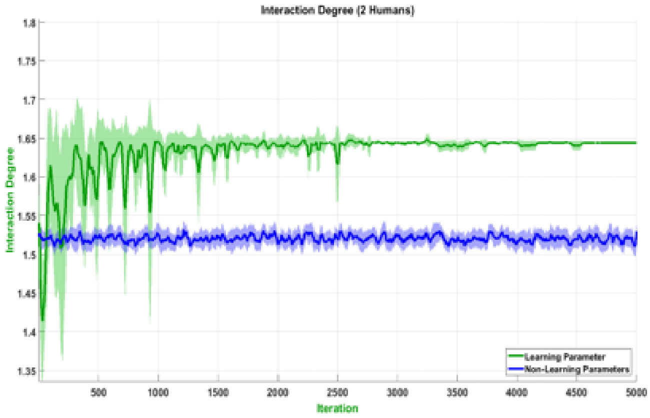
No. of Humans	**Average Interaction Degree**		**Average Unacceptable Degree**	
	Initial Parameters	**After Learning**	**Initial Parameters**	**After Learning**
2	5.6804	5.7923	0.9742	0.2528
3	5.5694	7.6095	1.8829	0.2321
4	5.261	6.6644	3.4072	0.9995

Table 4.3 Results of learning social model with people facing different directions

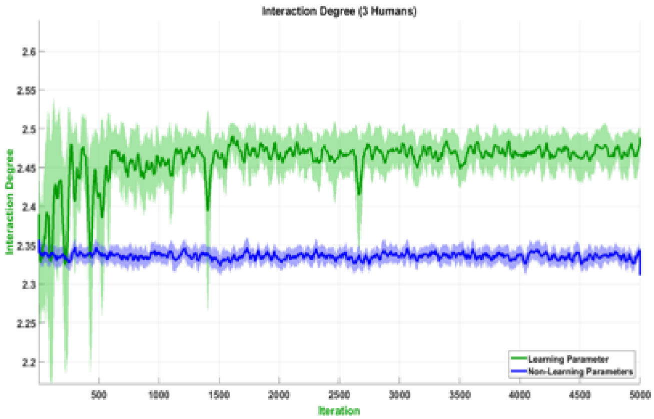
Facing Direction of 4 Humans	**Average Interaction Degree**		**Average Unacceptable Degree**	
	Initial Parameters	**After Learning**	**Initial Parameters**	**After Learning**
Into the center of group	5.5787	7.1985	1.7458	0.2129
Out of the center of group	5.5695	7.6095	1.8829	0.2321

4.3.3 Learning Fuzzy Social Model with Different Conditions

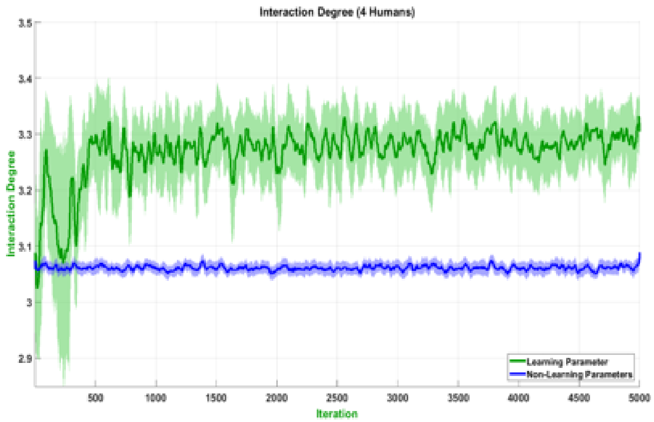
In this paper, we define the quality interaction area and the private area. Fig. 4.7 shows the interaction degree with three, four, and five subjects, respectively. The results show that our proposed method increases the interaction degree of subjects during their interaction with the robot until it suits everyone. Fig. 4.8 shows the results of the unacceptable degree. The results show that our method can reduce the unacceptable degree of subjects until they feel comfortable to interact with the robot. These results show that our proposed model outperforms the fixed-parameter for estimated the privacy area and more clearly with the number of humans in the environment. The results can be summarized in Table 4.2. We also perform the simulation with four subjects facing different directions. The results are consistent with the previous results obtained from the simulations with different numbers of subjects. The results show that our proposed method increases the quality interaction degree and reduces the unacceptable degree of the subjects, as shown in Table 4.3.



(a)

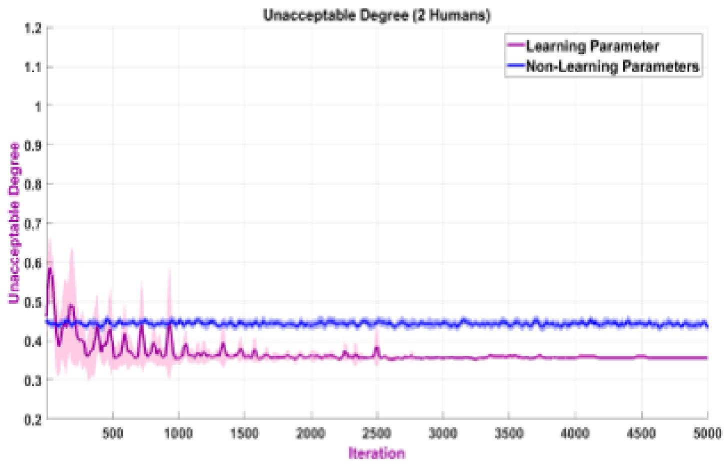


(b)

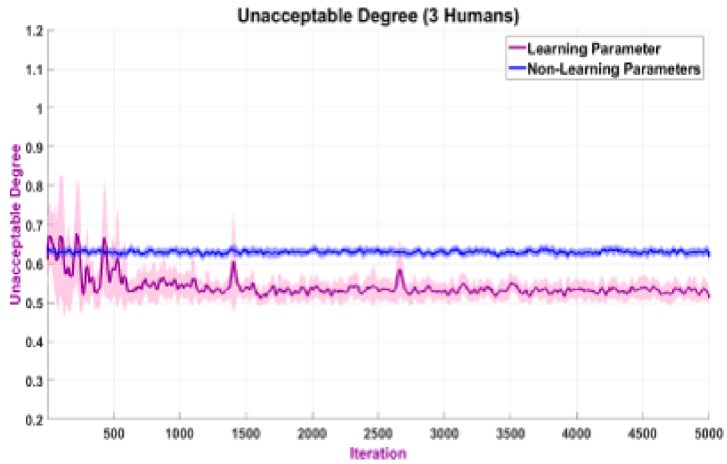


(c)

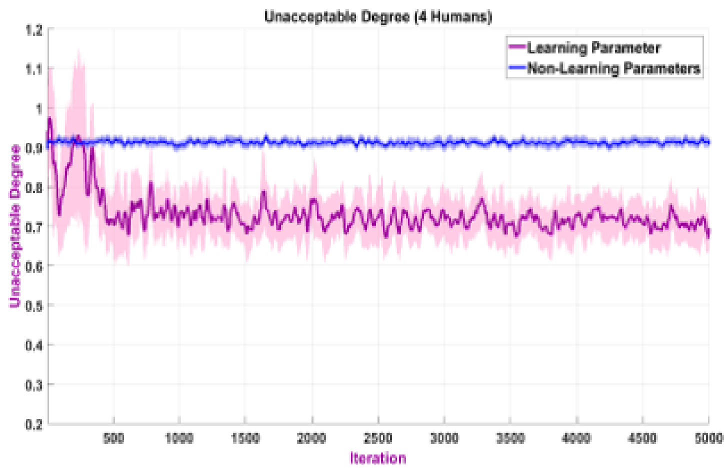
Figure 4.7 Interaction degree (ID) represent the acceptable or quality of interaction that the robot can receive from people along generated path. High interaction degree means that the robot approaches close enough to have interactions with humans.



(a)



(b)



(c)

Figure 4.8 Unacceptable Degree (UD) presents the total discomfort feeling that the robot receives from humans along the generated path. The robot should plan the path without entering the human private area.

4.4 Summary

This section began with the problem of fuzzy social model estimation method that use to estimate human personal space for a navigation task. The problem of the fuzzy social model method is that pre-designed MFs parameters cause the robot to estimate incorrect personal space for humans in the environment. This problem makes the robot to perform in unacceptable ways such as intruding into human privacy area or out of interaction range. These improper performances cause humans' feeling and make them refuse to operate with the robot.

This chapter applied reinforcement learning, which is one of machine learning, to solve the problem. By giving the ability to learn from humans' responses, the robot is able to modify the parameters that estimate human personal space suit to humans in the environment. However, there are lots of reinforcement learning algorithms; therefore, this chapter selected some algorithms to test and determine which algorithm is suited to our problem.

To find the efficacy of reinforcement learning algorithms which apply to the parameters adaptation. The convergent, learning time and exploration rate are used to evaluate algorithms. The convergent describes the possibility of algorithms to solve the parameters adaptation. The learning time determines the speed of the algorithm to solve the problems. The exploration rate is related to describe that the algorithm possible to find the best action to modify parameters in each situation.

The results show that most reinforcement algorithms can solve the problem and lead the robot to estimate perfect human personal space. Most of them convert the estimate social map to approach to the ground truth or realistic social map, except R-learning that approach with small error. Deep Q-Network has the best performance to converge the estimate social map to approach to the ground truth or realistic social map faster than other due to its experience replay memory that can reuse its past experience to find the appropriate action in each state or situation. However, its exploration rate is less than Actor-Critic which means that DQN can perform to approach the optimal value but the action that is selected might not be the best action in the state.

Chapter 5

Humanoid Robot Experiment

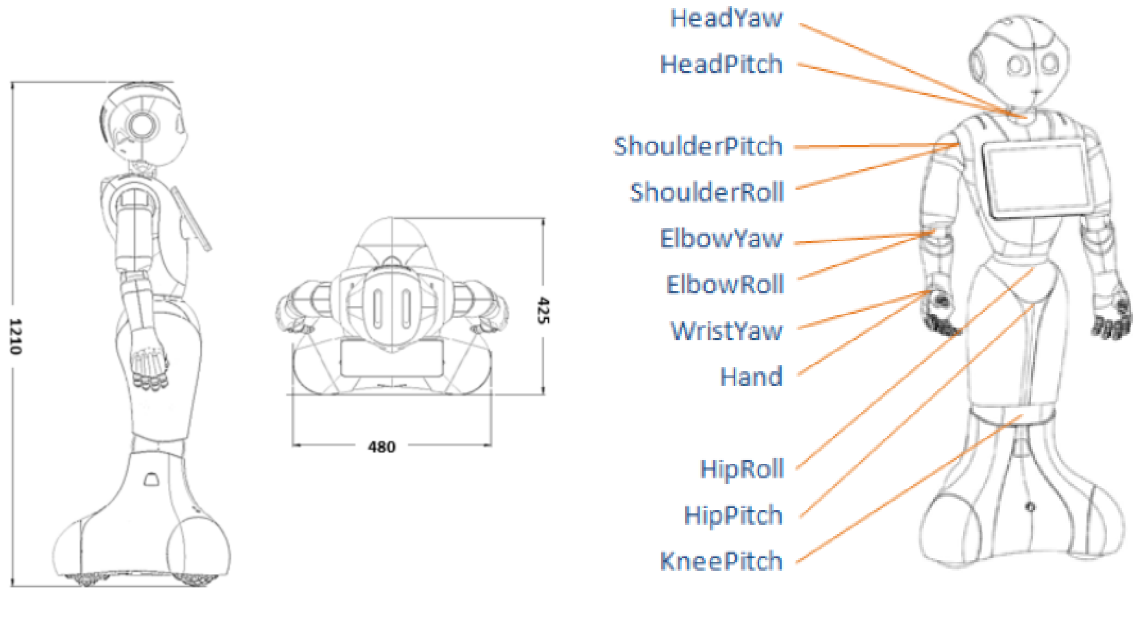
The proposed method has been tested and simulated in different conditions with simulation as described in the previous chapter. This chapter presents the results of implementing the experiment with a humanoid robot and conducted in the office-like environment to examine and validate the practicable of the proposed method. First of this chapter presents the information of humanoid robot and software which used to validate the learning fuzzy social model. Pepper robot which is the humanoid robot developed by SoftBank Robotics Corp is integrated with Robot Operating System (ROS) to implement the experiment. Then, the experimental procedure on the humanoid robot with participants is described to show the possible to applied our proposed method to the real world task.

5.1 Humanoid Robot and Software

This section introduces the necessary information of the humanoid robot and the interface software that is used in this research. To evaluate the possibility of our proposed method to the real-world task, we have to experiment with the real robot. Here, the Pepper robot is used to interact with humans and collect participants' response to our method. The robot control based on the Robot Operating System (ROS) which is the open-source environment.

5.1.1 Pepper Humanoid Robot

Pepper is a humanoid robot which developed by SoftBank Robotics Corp. The company aims to build a friendly robot that approachable for every type of customer. Its significant features are



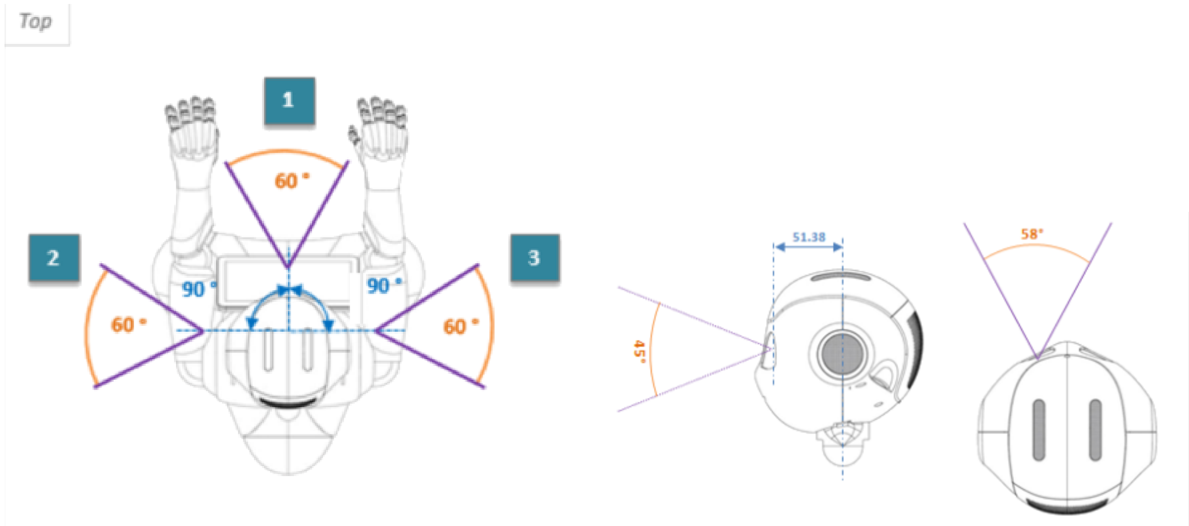
(a) (b)

Figure 5.1 Pepper dimensions and joint location [6].

ability to understand the human emotion and proximity sensors that enable the robot to track, localize and navigate in the environment. These abilities are useful and appropriate for research in HRI area.

Pepper's is holonomic type mobile robot which allows a wide range of movement that suite in social navigation. The total weight of Pepper distributed in the lower part of its body. The dimensions (5.1a) and body part (5.1b) are shown in Figure 5.1.

Most of the sensors that used in this research are lasers sensors which used to map generation and localization. At Pepper's knees have three sets of lasers as identified in Figure 5.1. The original document of Pepper, the control of these lasers are control separately as different sensors [6], but in the ROS packager [71] these three lasers are grouped together in single node. Another sensor is a depth camera which is ASUS Xtion depth camera. This camera is used to detect the person in the experiment. It located in the eye of the robot. The range and field of detection of both sensors are shown in Figure 5.2.



(a) (b)

Figure 5.2 Laser sensors and depth camera detection field [6].

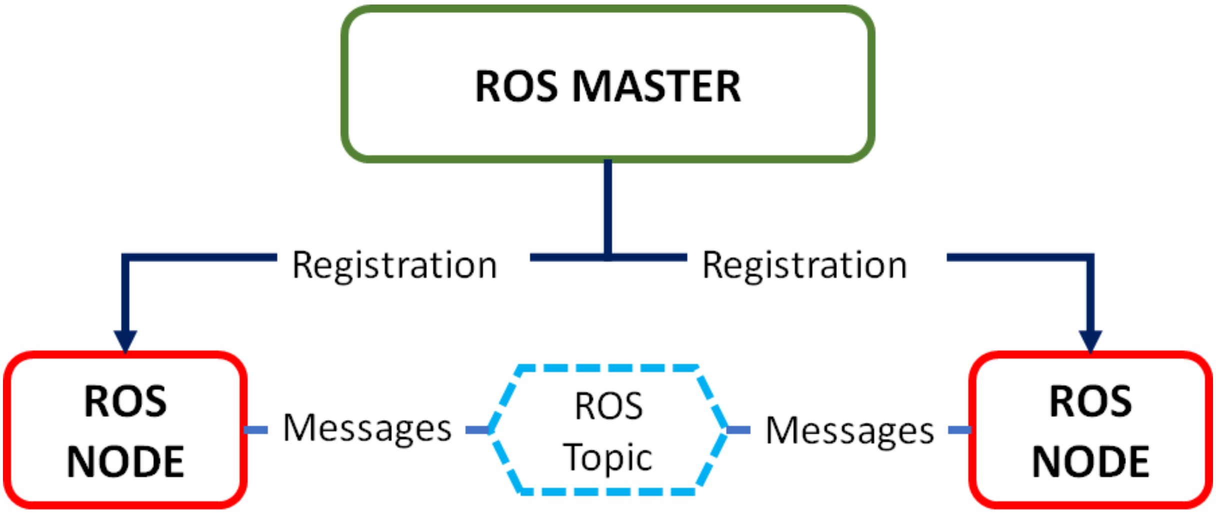


Figure 5.3 ROS Structure

5.1.2 Robot Operating System

The robot operating system(ROS) is an open-source, flexible framework. It is a collection of tools, libraries and conventions that can be applied to to simplify the task of designing the complex robot performance across a variety of robotic platform [72]. The flexible communication between different specific component is the main advantage of ROS. The basic structure of ROS is illustrated in Figure 5.3. The *Master* manages the communication between nodes by giving names and registration to nodes. Nodes transfer the data through *Topics* that receives data packages in term of *Messages*. There are two types of nodes in ROS, *Publishers* which advertise information to a topic, and *Subscribers* which receive the topics and retrieve the important information.

Pepper has its own software to control itself, called the NAOqi framework. However, with the limitation of its application programming interface(API), so we use ROS to develop and program our proposed model to the robot. ROS Package which is provided by the robot developing company is used to link specific parts of NAOqi's API as a component of a ROS system, called *NAOqi Drive*. This ROS Package is created to connect between specific module in the NAOqi framework as a ROS package.

Since the NAOqi driver provided a link between the NAOqi framework and ROS, additional programming are required to control specific components of the robot. For example, we can write the program the control Pepper to approach the person with our proposed method.

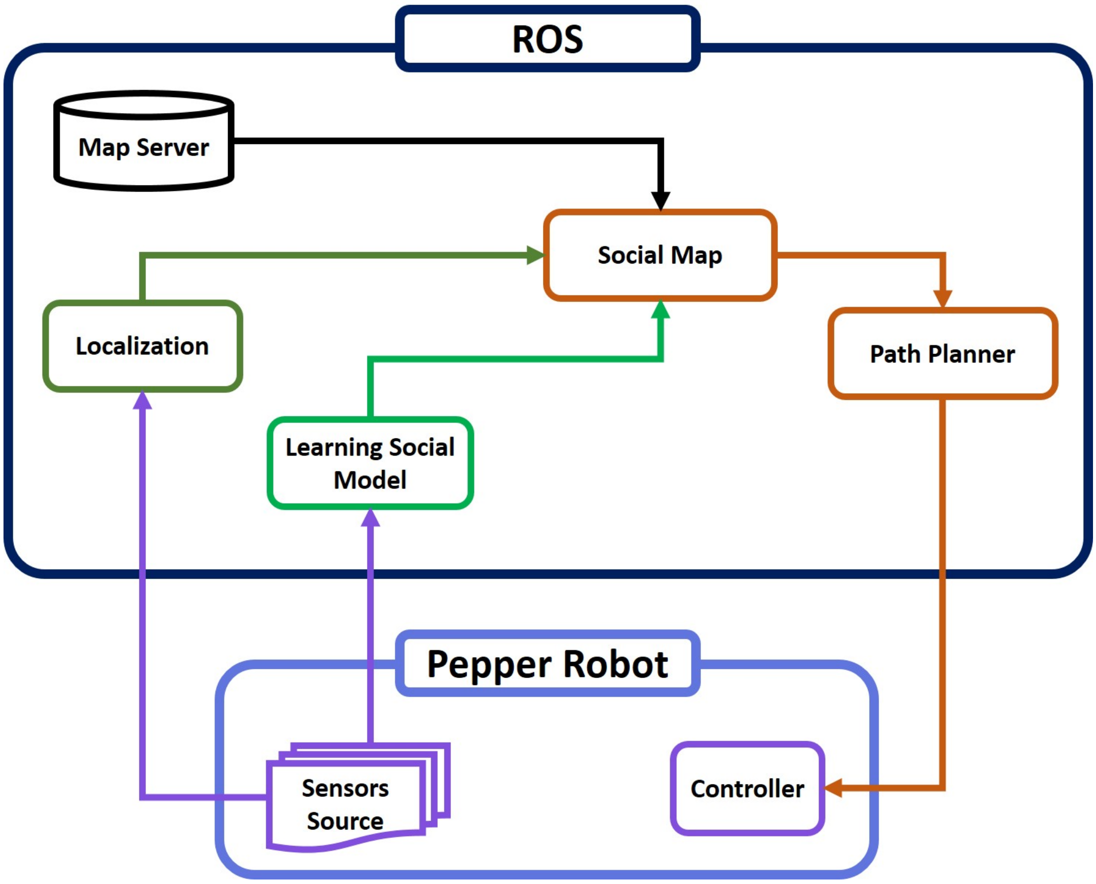


Figure 5.4 Humanoid Robot Experiment Overall Process

5.2 Humanoid Robot Experiment

This section briefly describes the related ROS packages that are used to implement the experiment. Then, the results of the experiment that test with participants who have different social information are described to validate our proposed method. The overall process of the experiment is illustrated in Figure 5.4. Specifically, Pepper needs to have prior knowledge about the environmental geometric map which can be stored in the map server. With several sensors, Pepper can localize itself required for the navigation task. Pepper also can detect and receive the human state and social factors to generate the social map to assign the social cost to the geometric map. This social map imposes constraints on the robot path, enabling the robot to avoid or interact with people. The robot also receives a reward from people to update the parameters of MFs to re-compute and update the social map.

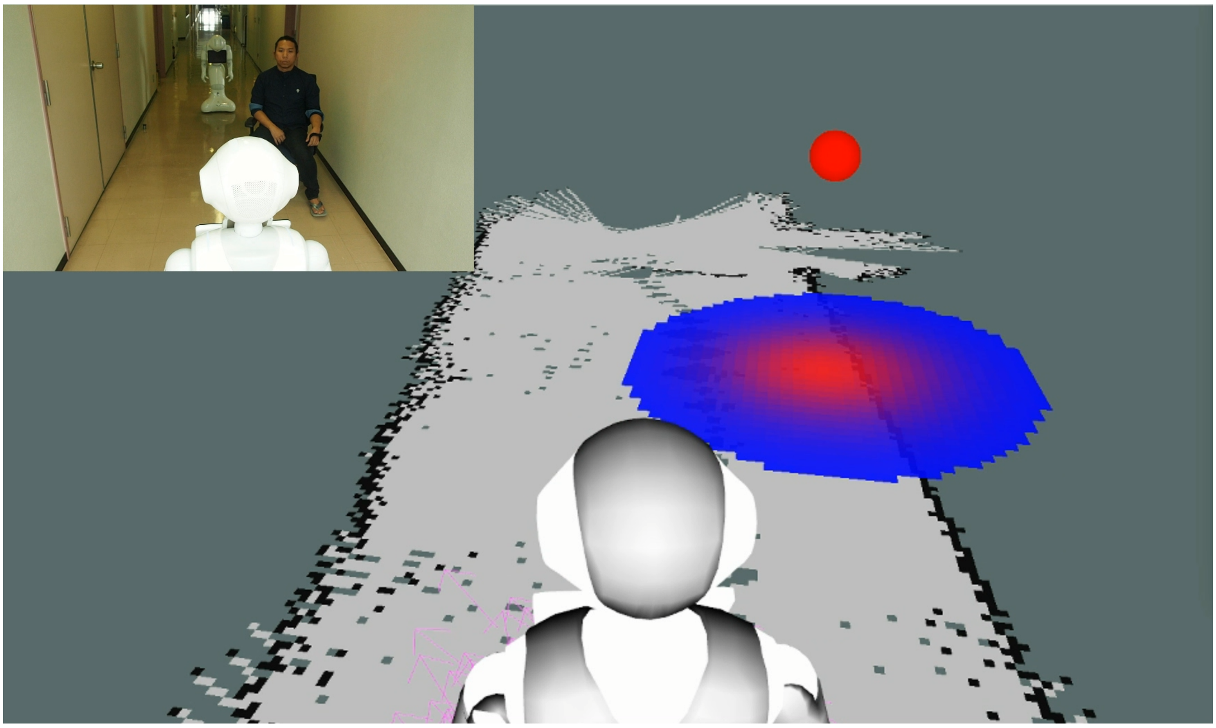


Figure 5.5 Human detection which used leg_detector package to detect participants.

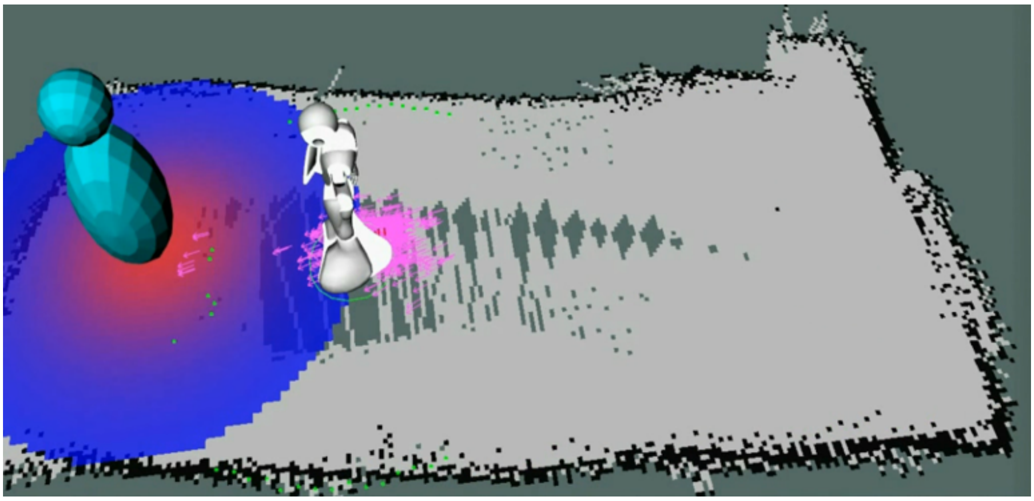
5.2.1 ROS Related Packages

This dissertation concerned the feasibility of implementing a social interaction that runs in Pepper. There are two tasks that the robot should perform. First is the estimation of the human model which create the cost map for a navigation task. Second is the navigation task that the robot performs to approach participants in the environment. Therefore, this section presents the related packages in the ROS environment to implement our proposed method with Pepper robot.

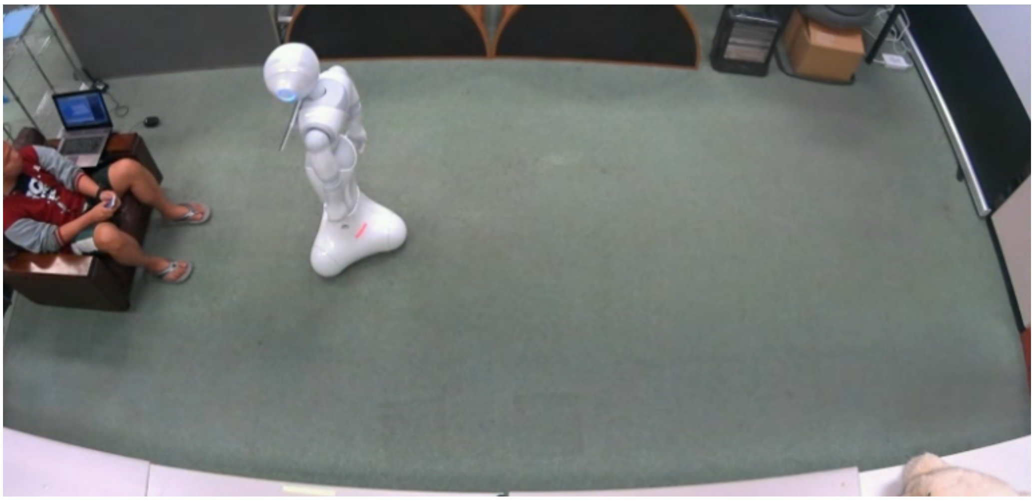
Human Social Cost

To generate human social cost that used for robot navigation. Two ROS packages are briefly described. First is the human detection that detects the human and give the location in the generated geometric map. Second is the cost map package which uses to give the value to every position on the geometric map.

- **Human Detection** The process of detecting human is achieved through the use of ROS Package called **leg_detector**. The NAOqi framework has API to detect human with face recognition by using the depth camera in Pepper robot. However, the depth camera has a short range to perceive the information which is not suite to our experiment that the



(a)



(b)

Figure 5.6 Private area of the participant which generated according to our proposed method by using **costmap_2d** package (5.6a) compare to real-world environment (5.6b)

robot should be able to detect the human from far distant. Therefore, This **leg_detector** package takes the message from laser scans, which has longer detect range, for human detection. This is more suitable for our experiment. The package uses the message from laser scan as input and uses a machine-learning-trained classifier to detect groups of laser readings as possible legs [73]. Then, the algorithm provides the position of the human which is used as the input to produce or estimate the private area of the human. Figure 5.5 illustrates the human detection experiment. The red ball represents the location of the human. This location is used as the input to generate the personal space cost which has high cost at the human position, which is shown by a blue circle on the map, and high cost

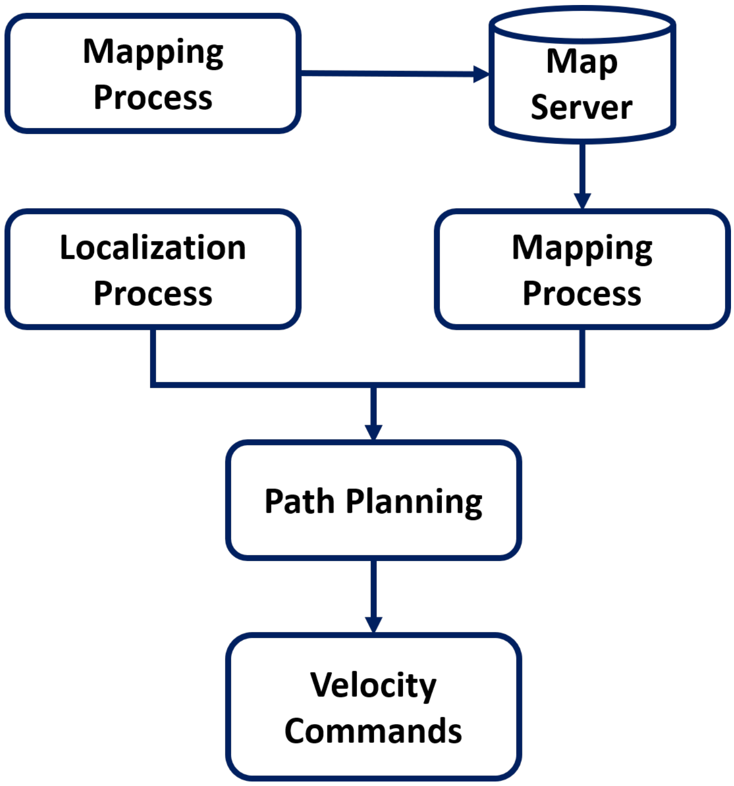


Figure 5.7 ROS Navigation Structure

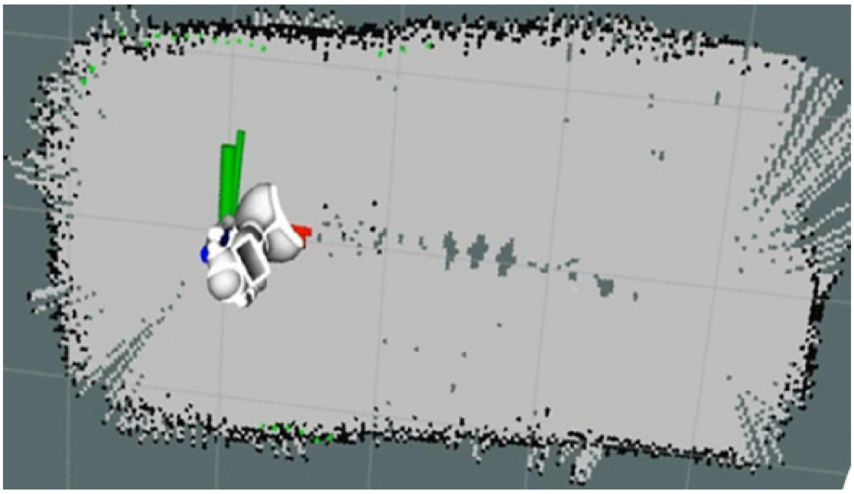
with a red circle.

- **Cost map generator package** Our method estimates and generates the cost surround the human. **costmap_2d** package is used to provides an implementation of a 2D cost map that takes sensor data from the environment to build 2D cost map [74]. The general **costmap_2d** generates the cost only for the obstacle, but we can use it API to generate the private area cost for the human, as shown in Figure 5.6.

Robot Navigation

The robot navigation is an important task to examine and validate our proposed method. ROS provides the packages for the robot platform to run the navigation task. The overview of the navigation structure is shown in Figure 5.7. Therefore, this section gives brief information about the related package for the navigation process.

- **Mapping Process:** The process of mapping the environment was achieved through the use of a *Simulataneous Localization and Mapping (SLAM)* algorithm. Fortunately, this algorithm is already implemented as a ROS Package called **GMapping** [75]. The basic



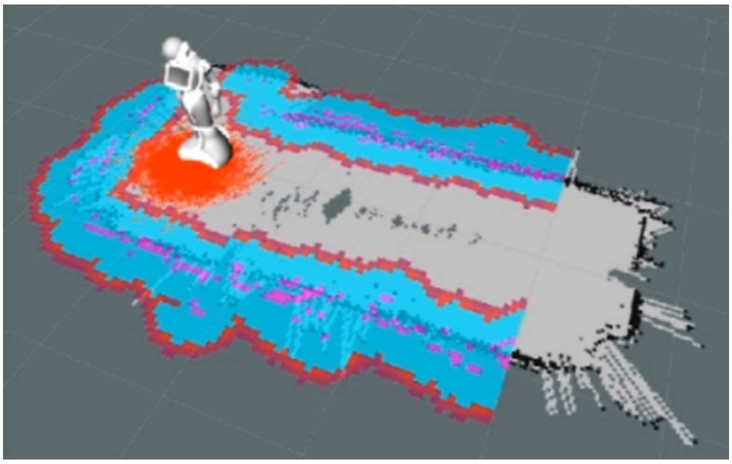
(a)

(b)

Figure 5.8 Map is constructed by Pepper (5.8a) compare to real-world environment (5.8b)

requirement to run this package is a robot that has horizontal laser scan which has already fulfilled by Pepper robot. The algorithm that built-in **GMapping** provides parameters that can be tuned to accommodate the hardware specification of the robot platform. To make Pepper robot to obtain a good result in the mapping process, the following parameters in **GMapping** package are tuned:

- **map_update_interval** This parameter is used to tune the frequency of updating occupancy grid for construct map.
- **angular/linear update** These parameters define how fast the map is created as Pepper move.



(a)

(b)

Figure 5.9 Pose estimation of the robot in RViz (5.9a) and real-world (5.9b)

- **minimumScore** This parameter is used to reduce the interval between matching particles in the map while Pepper is moving.
- **maxUrange and maxRange** The value of *maxUrange* needs to be lower than *maxRange* to clear areas of the map where a laser beam traverses but does not hit an obstacle within the range of the laser.
- **particles** This parameter is used for make algorithm more precise. However, the high value of these values also increases the computational power.

Once the settings of the algorithm were tuned to Pepper, the map of the environment stored in the map server that can be called to use in the navigation process. The result of the mapping process can be shown in Figure 5.8.

- **Localization** After the map has already construct, the localization process which used to estimate the location of the robot during the navigation is used to localize the robot position during the navigation. In this research, Adaptive Monte Carlo Localization (AMCL) algorithm is the common tool for localization purpose. AMCL is the ROS package that frequently used in robot navigation task. The algorithm is based on a particle filter that, give the map of environment, estimates the position and orientation of a robot as it moves and sense the environment [76]. Similar to gmapping package, AMCL package also has the parameter to tune for the best results as follows:

  - **min_particles** This value is used to proved the convergence of the pose estimation.
  - **max_particles** This parameter allows to process more particles in each iteration to overcome the problem of missing laser scan.
  - **odom_alpha** These values were increased to spread out the samples and still obtain a good sample set.

  Once the setting of the algorithm were tubed to pepper, the Pepper can localize itself in the map. Figure 5.9 presents the image that capture from RViz that show the location of the robot compare real environment.

- **Path Planning** Our proposed method provides estimate personal area of the human by giving the cost surround that person. Therefore, to implement the experiment, the famous ROS path planning package which based on Dynamic Window Approach (DWA) [77] is used as the path planner. By using a map, the path planner creates a kinematic trajectory for the robot to get from a start to a goal location. Along the way, the planner uses the cost of surrounding environment which represented as a grid map. This cost value encodes the costs of traversing through the grid cells. The path planner's job is to use this value function to determine linear and orientation velocities to send to the robot.

5.2.2 Experiment Results

In human-aware navigation, there are many application or robot motion schemes that can be tested, for example, the person following, accompanying persons and approaching and passing in a sociable way. The personal following is the task that the robot tries to move around or fellow humans. Accompanying person task is similar to the person following task but for the

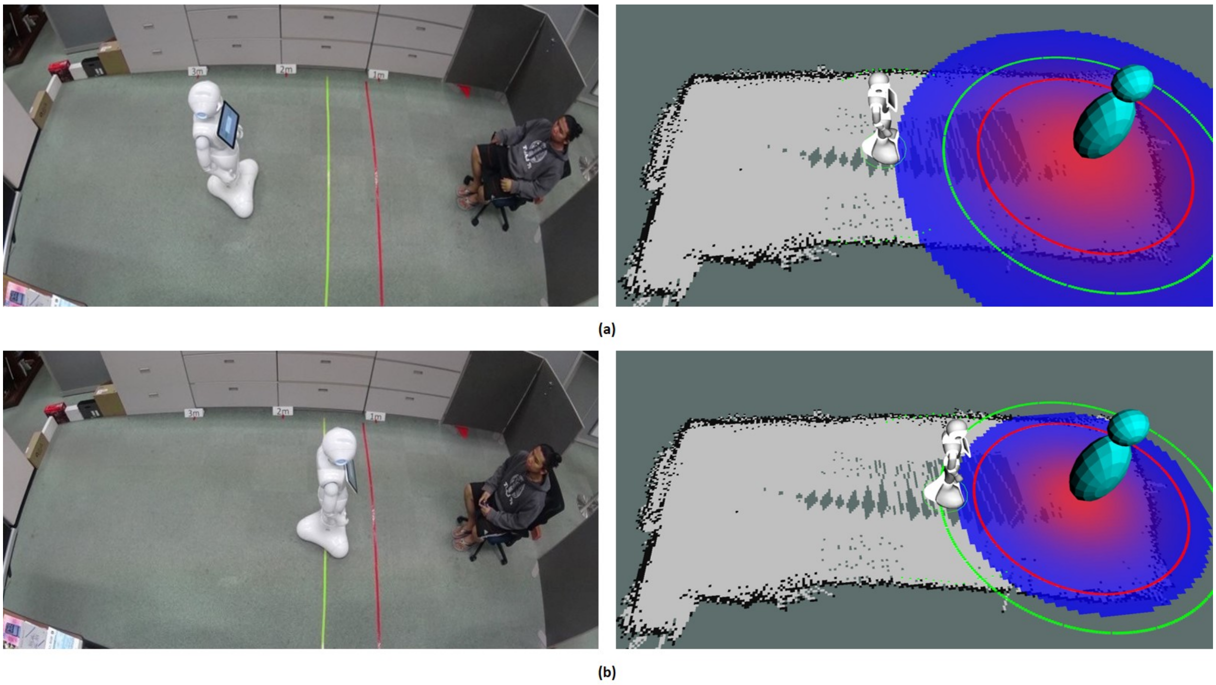


Figure 5.10 **Humanoid Robot Experiment:** (Left) The real experiments with Pepper. (Right) The blue area visualizes the estimate private area. The green line is the quality interaction area boundary Bi. The red line is the private area boundary Bp.

accompanying task, the robot tries to move with the human in the same direction and also try to guide the direction of human group by its behaviour. For approaching and passing humans, the research in this area is opposite to person following and accompanying tasks. The research in the area tries to design the robot behaviour to approach human or passing the human in a natural way or without disturbing human's feeling. Our experiment with the real robot like pepper is focused on approaching the person task to evaluate that our proposed method can make the robot maintain the distance between human and robot by learning from human's feedback.

In the experiment, the Pepper robot has a task to visits participants and keeps the distance to make them feel comfortable around it. However, as many uncertainties exist, it is likely that Pepper initially makes a rough estimate of the size of the personal area which may not be suitable for him/her to comfortably interact with it. Here, participants have been asked for their private range and the interaction range that they feel comfortable to interact with pepper. For instance, Figure 5.10 (a) shows that Pepper is outside the boundary of the quality interaction area (green line). During the interaction with Pepper, participants give reward by the verbal answer to the question from the robot. This reward allows Pepper to evaluate the distance between participants and itself. The positive reward is given when Pepper is within the area where the participants feel comfortable to interact with it. In the other hand, the negative reward is given for the distance from which they feel difficult to interact or discomfort which is outside the quality interaction area boundary (green line) or inside the private area boundary (red line). Learning fuzzy social model helps Pepper to re-estimate the human personal area until gaining a maximum positive reward. Finally, Pepper can locate itself within the area to interact with people that separates the private area as shown in Fig. 5.10(b).

In order to evaluate our proposed model, a total of five subjects participated in the experiment. Each person has a different range of quality interaction area, which is represented by the green line B_i and private areas, which is represented by the red line B_p. The results are shown in Fig. 5.11 to Fig. 5.15. It was confirmed that the social map might not clearly designate the private area at the initial phase of interaction, which is unsuitable for the subjects. In the case of Figs. 5.11, 5.12, 5.14, and 5.15, the robot is located away from the quality interaction area. Therefore the robot receives a hardly noticeable response from participants, which is considered to be the negative reward, to update its parameters associated with the MF of the interaction degree.

On the other hand, the robot receives a positive reward for updating its parameters for the MF of the personal area. In the case of Fig. 5.13, the robot is initially located inside the private area. Therefore, the robot receives a negative reward to decrease the unacceptable degree and the positive reward to update the parameters associated with the interaction degree. Finally, our proposed personal space learning model enabled the robot to interact with the participants at the proper distance between the boundaries of interaction and private areas as shown in Fig. 5.11 to Fig. 5.15.

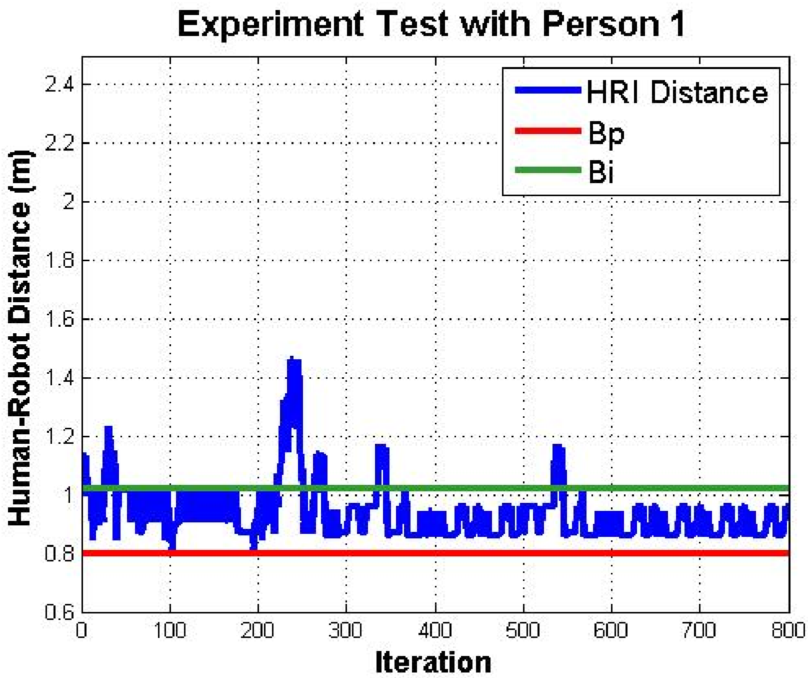


Figure 5.11 **Experiment Result with Pepper Robot:** the interaction distance (blue line) converges to the area between the quality interaction area boundary B_i and the private area boundary B_p of Person 1.

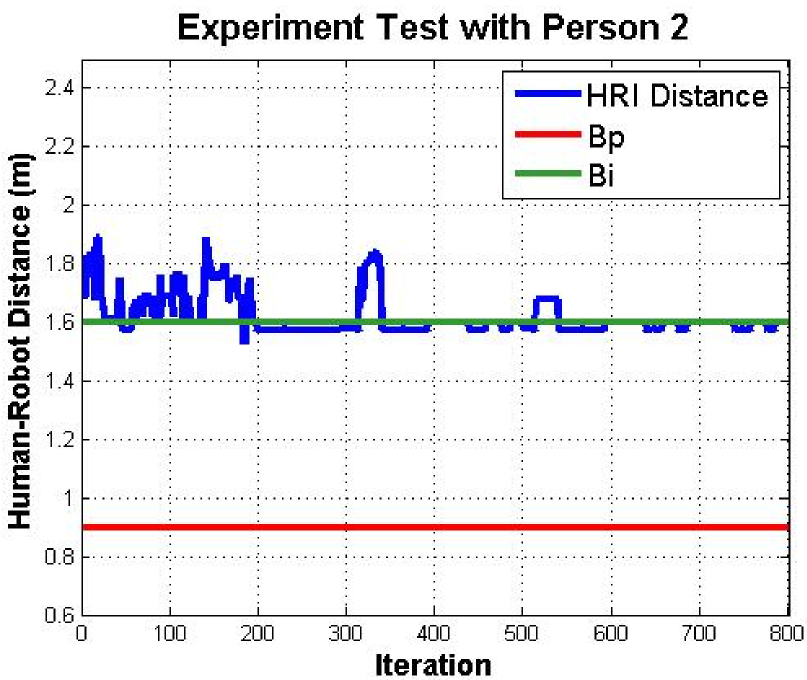


Figure 5.12 **Experiment Result with Pepper Robot:** the interaction distance (blue line) converges to the area between the quality interaction area boundary B_i and the private area boundary B_p of Person 2.

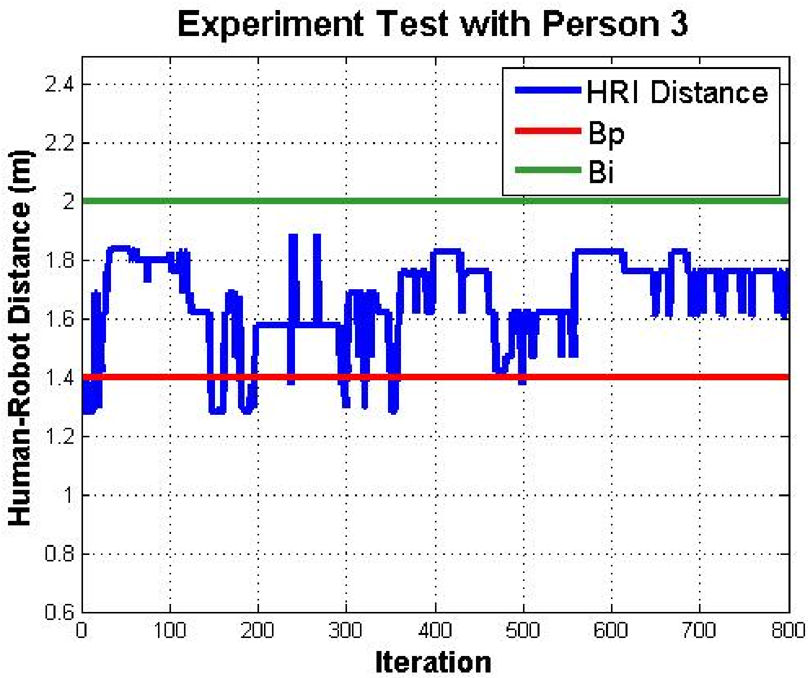


Figure 5.13 **Experiment Result with Pepper Robot:** the interaction distance (blue line) converges to the area between the quality interaction area boundary B_i and the private area boundary B_p of Person 3.

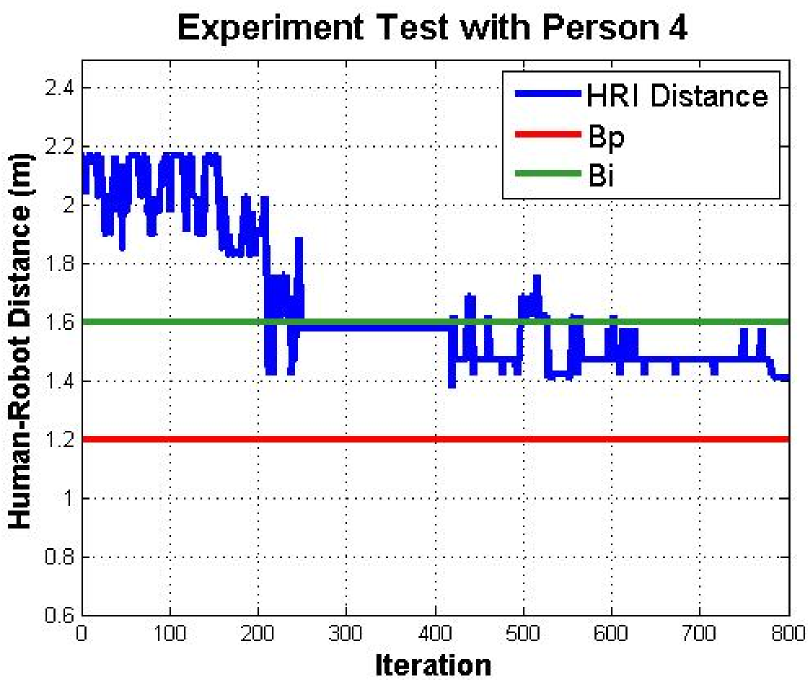


Figure 5.14 **Experiment Result with Pepper Robot:** the interaction distance (blue line) converges to the area between the quality interaction area boundary B_i and the private area boundary B_p of Person 4.

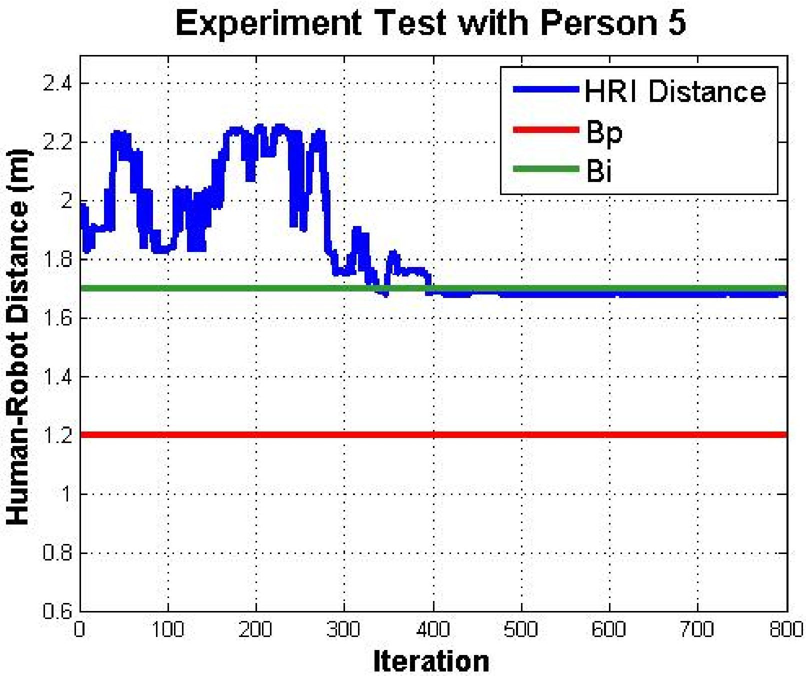


Figure 5.15 **Experiment Result with Pepper Robot:** the interaction distance (blue line) converges to the area between the quality interaction area boundary B_i and the private area boundary B_p of Person 5.

5.3 Summary

In this chapter, The results of the proposed method have been collected with the experiment with humans in the real world environment. Pepper robot, which is the humanoid robot developed by SoftBank Robotics Corp, is used in the experiment. Pepper has a variety of sensors, and its perception capabilities make pepper suitable to implement human-robot social interaction. The experiment of our proposed method is implemented on the open-source environment of Robot Operating System (ROS) which has many packages that can be used for our experiment.

The implement mostly used ROS packages which consist of human detection, cost map generation, mapping process, localization and path planner package. These packages have briefly summarized and showed the preparation results.

For the experiment, the robot has a task to visit the participants and keeps the distance to make them feel comfortable around it. However, as many uncertainties exist, it is likely that Pepper initially makes a rough estimate of the size of the personal area which may not be suitable for him/her to comfortably interact with it. In this case, the robot asks participant a question to provide the reward to it. This reward is used to evaluate the distance that the robot keeps from the participant. The reward is positive if the robot approaches the participant in the area between the private area boundary and interaction area boundary. After the reward is given, the robot re-estimate the participant's personal space and approach the participant again. The process run until the robot gain continues a positive reward.

Five participants participate in the experiment with Pepper. Each person has a different range of interaction area and private area. The results confirmed that the learning fuzzy social model enabled the robot to interact with the participants at the proper position which is between the boundaries of quality interaction and private areas.

Chapter 6

Conclusion

This chapter presents the conclusions of the work regarding the proposed method. To close this chapter, future work is proposed to give the idea of some improvements needed to solve some issues in the current approach.

6.1 Conclusion

The mobile robots are widely used in many application, for example, health care service, customer service, or event in the household. Most of the tasks requires robots to move and provide services in different locations. Therefore, robot navigation is important and widely developed to enable them to operate safely. The safe navigation is the critical requirement for robot navigation which enables robots to move around the environment without harm to the surrounding environment. However, even when the robots move safely, sometimes human feel that the motion of robots is not safe. The reason for unsafe feeling is the unfamiliar or lack of trusting in the technology. Therefore, social competence which is the rules of society that humans use to act with other humans is applied to the robot navigation task. This social competence makes robots behave more naturally and acceptable for humans to feel safe and comfort while robots operate around in the shared environment.

Human-aware navigation is the challenge for human-robot symbiosis that considers both safe and socially navigation. The research of human-aware navigation can be separated into three different approaches. First is the naturalness which is the development of low-level behaviour of the robot. The research in this approach strives to imitate human motion as the target behaviour and recreate the robot's behaviour. Second is sociability which applied high-level social conventions. The research in this approach is how to transfer the social convention into the robot. The last approach is the comfort which considers to human's feeling. The research in this area considers the motion of robots which is not only safe but also move in a way that makes human more relax.

The Proxemics theory which is a social science and psychological theory is mostly used to formalize to mathematic model. This theory describes how human use of space to different humans in the environment. Therefore, to use this theory into human-aware navigation. The researcher has to model it into mathematical formulations. Two popular methods to model the private area or personal area according to the Proxemic theory are a geometric method and cost-based method. In this work, the cost-based method is used to model the human's personal space.

An asymmetric Gaussian function is a function in the cost-based method. It provides the degree to different locations in the environment which can be used as the cost for the robot navigation. The variance parameters are essential parameters to model the shape and size of a human's personal space. They can be determined by social information. However, most of the work considered only single social information which is not enough to estimate correct personal space.

This dissertation contributes the method to estimate the personal space of each person from his/her social information and has the ability to learn from the human's response. The premise is that the model can estimate the individual's personal space reflecting individual's social information, and possible to update the personal area according to the response of the human.

Here, Learning Fuzzy Social Model is proposed. The method consists of two part. First is the personal space estimation which uses a fuzzy inference system to determine parameters for asymmetric Gaussian function. However, with pre-design membership function parameters of FIS, the estimated personal space might not be correct. Therefore, reinforcement learning is integrated into fuzzy logic to update the membership functions' parameters from the response of humans.

This dissertation also presents the result of different reinforcement learning algorithms for parameters adaptation problem in our model. Three criteria are used to evaluate the algorithms. First is the convergent of the error between the estimated and the real social map. This convergent can describe the ability to learn to approach realistic social space. The second is the learning period which describes how fast algorithms to learn and the accumulate reward converges to a maximum value. The third is the exploration rate of algorithms used to explain how many state-action pairs that have been explored. This exploration rate is used to describe a possibility that selected action in each state is the best action or optimal action.

The results show that most reinforcement learning algorithms can modify the fuzzy membership functions that cause the estimated personal space similar to ground truth. In detail, for learning time, Deep Q-Network is overcome other algorithms. This because Deep Q-Network has memory to store the experience which can be reused to learn again. However, the fast learning time may have the trade-off with the state-action space exploration rate. In case, Actor-Critic can explore the state-action space better than other algorithms but has the trade-off to the learning time.

This contribution is useful to integrate into the mobile service robot to service humans in a health-care center or household. The robot is able to estimate the users' personal space according to the users' gender, experience with his/her robot and the range of the robot location to themselves. During the operation, the robot also has the ability to adapt its estimation according to the users' feeling. This process will be operated automatically by the robot. Therefore, the user will feel more relaxed to have the robot to service in their environment.

6.2 Future work

In recognition of the never-ending nature of research, there are always some aspects of any work that can be improved and expanded upon by future research. This may be caused by the limitations of time and resources, or other unforeseen difficulties. The same is true for this thesis, where some problem remains unsolved, and new topics appear continuously. Therefore, it is the wish of the author to suggest some ideas regarding the future direction of this work.

First, our proposed method is simulated and implemented with the assumption that social information is already processed and ready to use. Therefore, the process to get the social information should be considered to use in the practical or implement in the real world.

Second, the simulation results presented in this thesis is based on some period of time which means this simulation is tested based on the static environment. However, in realistic, humans are moving which means the environment is dynamics. This dynamics may cause the efficacy of reinforcement learning algorithms. Therefore, the proposed method should be considered in the dynamic environment which has human's motion

Third, our reward function is assumed to collect from the distance between humans and the robot. However, this reward can be collected from other processes. For example, emotion, behaviour, feeling, even with the questioner. Therefore, instead of using only distance information. The developer should determine the process to gain the response of human which make the proposed method more accurate.

Fourth, the selected reinforcement learning algorithms in this dissertation used only one action selection strategy (ϵ- greedy) affects the efficacy and performance of reinforcement learning algorithms. Therefore, the action selection strategy should be considered and investigated to find which strategy is suited to each algorithm.

Fifth, the learning period of all algorithms is still substantial which caused by the action selection strategy, hyperparameter of algorithms or algorithm is not suited to the problem. Therefore, the development of reinforcement learning algorithms still needs to accelerate the learning period and suit our problem.

Lastly, the learning fuzzy social model has experimented with the real robot to obtain the proof of validity and effectiveness of the proposed model. However, with the limited area and robot capabilities, the implement with many people in the same is omitted.

Bibliography

[1] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 17–24, March 2007.

[2] J. Mumm and B. Mutlu, "Human-robot proxemics: Physical and psychological distancing in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 331–=338, March 2011.

[3] E. Pacchierotti, H. I. Christensen, and P. Jensfelt, "Evaluation of passing distance for social robots," in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 315–320, Sept 2006.

[4] Y. Nakauchi and R. Simmons, "A social robot that stands in line," in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)*, (Takamatsu, Japan), pp. 357–364 vol.1, 2000.

[5] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, pp. 874–883, Oct 2007.

[6] A. S. Group, "Aldebaran documentation," 2019. 2019-01-30.

[7] 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.

[8] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics Automation Magazine*, vol. 4, pp. 23–33, March 1997.

[9] C. Stachniss and W. Burgard, "An integrated approach to goal-directed obstacle avoidance under dynamic constraints for dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 508–513 vol.1, Sept 2002.

[10] H. T. Edward, *The Hidden Dimension : man's use of space in public and in private*. London, UK: The Bodley Head Ltd, 1969.

[11] F. Lindner, "A conceptual model of personal space for human-aware robot activity placement," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Hamburg), pp. 5770–5775, Sept 2015.

[12] A. K. Pandey and R. Alami, "A framework towards a socially aware mobile robot motion in human-centered dynamic environment," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Taipei), pp. 5855–5860, Oct 2010.

[13] C. P. Lam, C. T. Chou, K. H. Chiang, and L. C. Fu, "Human-centered robot navigation;towards a harmoniously human; robot coexisting environment," *IEEE Transactions on Robotics*, vol. 27, pp. 99–112, Feb 2011.

[14] S. T. Hansen, M. Svenstrup, H. J. Andersen, and T. Bak, "Adaptive human aware navigation based on motion pattern analysis," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, (Toyama,Japan), pp. 927–932, Sept 2009.

[15] R. Kirby, R. Simmons, and J. Forlizzi, "Companion: A constraint-optimizing method for person-acceptable navigation," in *in the Proceedings of the IEEE international Symposium on Robot and Human Interactive Communication*, 2009.

[16] P. Patompak, S. Jeong, I. Nilkhamhang, and N. Y. Chong, "Learning social relations for culture aware interaction," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp. 26–31, June 2017.

[17] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, pp. 4282–4286, May 1995.

[18] T. B. Sheridan, "HumanâĂŞrobot interaction: Status and challenges," *Human Factors*, vol. 58, no. 4, pp. 525–532, 2016.

[19] J. Herring, *Medical Law and Ethics*. Oxford University, 2014.

[20] A. Westin, *Privacy and Freedom*. Bodley Head, 1970.

[21] H.-G. S. Zeeger S. K., Readdick C.A., "Daycare children's establishment of territory to experience privacy," vol. 11, 1994.

[22] P. B. Newell, "A cross-cultural comparison of privacy definitions and functions: A systems approach," *Journal of Environmental Psychology*, vol. 18, no. 4, pp. 357 – 371, 1998.

[23] L. Takayama and C. Pantofaru, "Influences on proxemic behaviors in human-robot interaction," in *Intelligent robots and systems, 2009. IROS 2009. international conference on IEEE/RSJ*, pp. 5495–5502, 2009.

[24] E. A. Martinez-Garcia, O. Akihisa, and S. Yuta, "Crowding and guiding groups of humans by teams of mobile robots," in *IEEE Workshop on Advanced Robotics and its Social Impacts, 2005.*, pp. 91–96, June 2005.

[25] P. J. Elena Pacchierotti, Henrik I. Christensen, *Embodied Social Interaction for Service Robots in Hallway Environments*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[26] E. Martinson and D. Brock, "Improving human-robot interaction through adaptation to the auditory scene," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 113–120, March 2007.

[27] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz, "On the nonholonomic nature of human locomotion," *Autonomous Robots*, vol. 25, pp. 25–35, Aug 2008.

[28] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H. I. Christensen, "Navigation for human-robot interaction tasks," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 2, pp. 1894–1900 Vol.2, April 2004.

[29] P. Saulnier, E. Sharlin, and S. Greenberg, "Exploring minimal nonverbal interruption in hri," in *2011 RO-MAN*, pp. 79–86, July 2011.

[30] Y. Tamura, P. D. Le, K. Hitomi, N. P. Chandrasiri, T. Bando, A. Yamashita, and H. Asama, "Development of pedestrian behavior model taking account of intention," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 382–387, Oct 2012.

[31] J. Müller, C. Stachniss, K. O. Arras, and W. Burgard, "Socially inspired motion planning for mobile robots in populated environments," 2008.

[32] J. T. Butler and A. Agah, "Psychological effects of behavior patterns of a mobile personal robot," *Auton. Robots*, vol. 10, pp. 185–202, Mar. 2001.

[33] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon, "How may i serve you?: A robot companion approaching a seated person in a helping context," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, (New York, NY, USA), pp. 172–179, ACM, 2006.

[34] H.-M. G. Jens Kessler, Christof Schroeter, *Approaching a Person in a Socially Acceptable Manner Using a Fast Marching Planner*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[35] D. S. S. M. L. W. K. D. K. L. Koay, E. A. Sisbot and R. Alami, "Exploratory study of a robot approaching a person in the context of handing over an object," in *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, p. 18==24, 2007.

[36] E. Martinson, "Hiding the acoustic signature of a mobile robot," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 985–990, Oct 2007.

[37] E. Martinson, *Acoustical awareness for intelligent robotic action*. 2007.

[38] G. Diego and T. K. O. Arras, "Please do not disturb! minimum interference coverage for social robots," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1968–1973, Sept 2011.

[39] K. Huang, J. Li, and L. Fu, "Human-oriented navigation for service providing in home environment," in *Proceedings of SICE Annual Conference 2010*, pp. 1892–1897, Aug 2010.

[40] R. Tomari, Y. Kobayashi, and Y. Kuno, "Empirical framework for autonomous wheelchair systems in human-shared environments," in *2012 IEEE International Conference on Mechatronics and Automation*, (Chengdu), pp. 493–498, Aug 2012.

[41] L. Scandolo and T. Fraichard, "An anthropomorphic navigation scheme for dynamic scenarios," in *2011 IEEE International Conference on Robotics and Automation*, pp. 809–814, May 2011.

[42] M. Svenstrup, T. Bak, and H. J. Andersen, "Trajectory planning for robots in dynamic human environments," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Taipei), pp. 4293–4298, Oct 2010.

[43] M.Luber, L. Spinello, J. Silva, and K. O. Arras, "Socially-aware robot navigation: A learning approach," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 902–907, Oct 2012.

[44] P. Papadakis, P. Rives, and A. Spalanzani, "Adaptive spacing in human-robot interactions," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Chicago, USA), pp. 2627–2632, Sept 2014.

[45] P. Papadakis, A. Spalanzani, and C. Laugier, "Social mapping of human-populated environments by implicit function learning," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1701–1706, Nov 2013.

[46] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965.

[47] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice Hall Press, 3rd ed., 2009.

[48] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," (Amsterdam, The Netherlands, The Netherlands), pp. 3–24, IOS Press, 2007.

[49] D. Schuurmans and M. Zinkevich, "Deep learning games," 2016.

[50] H. Cuayáhuitl, N. Dethlefs, L. Frommberger, K.-F. Richter, and J. Bateman, "Generating adaptive route instructions using hierarchical reinforcement learning," in *Spatial Cognition VII* (C. Hölscher, T. F. Shipley, M. Olivetti Belardinelli, J. A. Bateman, and N. S. Newcombe, eds.), (Berlin, Heidelberg), pp. 319–334, Springer Berlin Heidelberg, 2010.

[51] I. Kastanis and M. Slater, "Reinforcement learning utilizes proxemics: An avatar learns to manipulate the position of people in immersive virtual reality," *TAP*, vol. 9, pp. 3:1–3:15, 2012.

[52] J. Hao and H.-f. Leung, "Achieving socially optimal outcomes in multiagent systems with reinforcement social learning," vol. 8, 09 2013.

[53] F. Martinez-Gil, M. Lozano, and F. Fernández, "Calibrating a motion model based on reinforcement learning for pedestrian simulation," in *Motion in Games* (M. Kallmann and K. Bekris, eds.), (Berlin, Heidelberg), pp. 302–313, Springer Berlin Heidelberg, 2012.

[54] R. Beheshti and G. R. Sukthankar, "A normative agent-based model for predicting smoking cessation trends," in *AAMAS*, 2014.

[55] D. Berenson, T. SimÃľon, and S. S. Srinivasa, "Addressing cost-space chasms in manipulation planning," in *2011 IEEE International Conference on Robotics and Automation*, pp. 4561–4568, May 2011.

[56] R. Iehl, J. CortÃľs, and T. SimÃľon, "Costmap planning in high dimensional configuration spaces," in *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 166–172, July 2012.

[57] L. Jaillet, J. CortÃľs, and T. SimÃľon, "Sampling-based path planning on configuration-space costmaps," *IEEE Transactions on Robotics*, vol. 26, pp. 635–646, Aug 2010.

[58] J. Mainprice, E. A. Sisbot, L. Jaillet, J. CortÃľs, R. Alami, and T. SimÃľon, "Planning human-aware motions using a sampling-based costmap planner," in *2011 IEEE International Conference on Robotics and Automation*, pp. 5012–5017, May 2011.

[59] L. Jaillet, F. J. Corcho, and J. J. PeÌĄrez, "Randomized tree construction algorithm to explore energy landscapes," *J. Comput. Chem*, p. 2011.

[60] L. A. Hayduk, "Personal space: An evaluative and orienting overview," *Psychological Bulletin*, 01 1987.

[61] M. Aliakbari, E. Faraji, and P. Pourshakibaee, "Investigation of the proxemic behavior of iranian professors and university students: Effects of gender and status," *Journal of Pragmatics*, vol. 43, no. 5, pp. 1392 – 1402, 2011. Multilingual structures and agencies.

[62] P. Kalbfleisch and M. Cody, *Gender, Power, and Communication in Human Relationships*. Routledge Communication Series, Taylor & Francis, 2012.

[63] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Interpersonal distance in immersive virtual environments," *Personality and Social Psychology Bulletin*, vol. 29, no. 7, pp. 819–833, 2003.

[64] C. J. Erkelens, "Perspective Space as a Model for Distance and Size Perception," *Iperception*, 10 2017.

[65] A. Schiano Lomoriello, F. Meconi, I. Rinaldi, and P. Sessa, "Out of sight out of mind: Perceived physical distance between the observer and someone in pain shapes observer's neural empathic reactions," *ArXiv e-prints*, 08 2018.

[66] R. Gifford, "Projected interpersonal distance and orientation choices: Personality, sex, and social situation," *American Sociological Association*, vol. 45, no. 3, pp. 145–152, 1982.

[67] M. L. Walters, "The design space for robot appearance and behaviour for social robot companions," 2008.

[68] P. Tarnowski, M. KoÅĆodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Procedia Computer Science*, vol. 108, pp. 1175 – 1184, 2017. International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland.

[69] A. Schwartz, "A reinforcement learning method for maximizing undiscounted rewards," in *Proceedings of the Tenth International Conference on International Conference on Machine Learning*, (San Francisco, CA, USA), pp. 298–305, 1993.

[70] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[71] A. S. Group, "Pepper," 2019. 2019-01-30.

[72] O. Foundation, "Ros about," 2019. 2019-01-30.

[73] "Leg detector," 2019. 2019-01-30.

[74] "costmap_2d," 2019. 2019-01-30.

[75] “Gmapping,” 2019. 2019-01-30.

[76] “Adaptive monte carlo localization (amcl),” 2019. 2019-01-30.

[77] “Dynamic window approach (dwa),” 2019. 2019-01-30.

Publications

Journal

[1] P. Patompak, S. Jeong, I. Nilkhamhang and N. Y. Chong, "Learning proxemicsfor personalized human−robot social interaction", International Journal of Social Robotics, May 2019.

International Conference

[2] P. Patompak, S. Jeong, N. Y. Chong and I. Nilkhamhang, "Mobile robot navigation for human-robot social interaction," 2016 16th International Conference on Control, Automation and Systems (ICCAS), Gyeongju, 2016, pp. 1298-1303.

[3] P. Patompak, S. Jeong, I. Nilkhamhang and N. Y. Chong, "Learning social relations for culture aware interaction," 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, 2017, pp. 26-31.