Observing People's Behaviors in Public Spaces for Initiating Proactive Human-Robot Interaction by Social Robots

Md. Golam Rashed

Graduate School of Science and Engineering

Saitama University, Japan

Supervisor: Professor Yoshinori Kuno

A thesis submitted in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

September 2016

To my son,
Zunaid Mahmud
for his endless love.

Acknowledgements

First of all, I thank ALLAH for giving me strength and ability to complete this study. Besides, there are many people who have supported and encouraged me throughout this doctoral research that I would like to acknowledge.

I would like to thank my supervisor, *Professor Yoshinori Kuno*, Saitama University, Japan, for his trust, continuous support, guidance, and patience over the past few years. He has provided me with the opportunity to work in the field of Human Behavior Tracking for Human-Robot Interaction. I feel incredibly lucky and honor to be one of his doctoral students. I am sincerely grateful to him for his constant encouragement, and his stimulating ideas.

I am incredibly grateful to *Professor Yoshinori Kobayashi*, *Professor Tetsuya Shimamura*, and *Professor Takashi Komuro* for their comments, insightful discussions, valuable advice, and feedback during their evaluations of my dissertation work.

The Computer Vision (CV) Laboratory at Saitama University as a whole has been a wonderful and supportive group. I am grateful to all the members of the CV Lab for their hard work, thoughtful, and valuable contributions to my research over the past few years. I must acknowledge *Dr. Antony Lam* for his invaluable help on the research that I have completed and published. I am forever thankful to *Dr. Hisato Fukuda*, and *Dr. Ryota Suzuki*. I will never forget the valuable time that you have given me.

I owe a huge debt to my teacher, colleague *Professor Dipankar Das*, University of Rajshahi, Bangladesh, for his unconditional support and cooperation on my doctoral study.

The generous financial support from JSPS KAKENHI (Grant number 26330186), Japan Science and Technology Agency (JST), Core Research for Evolution Science and Technology (CREST) has made it possible to pursue my research.

Finally, but very certainly not least, I am grateful to my son *Zunaid Mahmud*, and my wife, *Zinat Musfera Ferdous* who had to bear with my moods and impatience whenever the research did not work out the way I wanted. Without their love, care, support, and presence, I would not have been able to complete my doctoral study in Japan. I thank my parents and parents-in-law for their continuing *DUA* throughout all the phases of my life that have led to this point. There are too many to name everybody personally, so I won't, I a,m sure you know who you are. I am lucky to have you all, and this dissertation could not have happened without you.

Abstract

Traditionally, humans have viewed robots as a "mechanical machines", designed to perform a variety of industrial tasks. But within the last decades, the reality of robots is quite different from the traditional view and has enabled us to start developing social robots to support humans in their daily activities. The concept of the social robot is rapidly emerging and gradually being introduced as a part of human society where interaction among humans and social robots seems to be important to provide mental, communicational, and physical support to humans in society. As a consequence, many social robots have already been deployed in social spaces, where humans interact with reactive services in which social robots wait until the human proactively seeks services. Nevertheless, nowadays we are moving in a direction where we introduce social robots in social spaces with the ability to proactively offer services to humans in which social robots estimate human intentions, and can offer services only to those who would need it. To achieve such capabilities, social robots should have the capacity to observe human behaviors so that they can easily identify humans who are in need. But, observing human behaviors is a challenging task for social robots. This dissertation deals with making human-robot interaction systems capable of observing human behaviors so that social robots can understand their intentions, interests, and preferences concerning surrounding environments. Our findings will help social robots to proactively offer services to those humans who may want to be serviced. In this dissertation, a real life museum guide robot scenario is considered as a testbed for my proactive social robotics research.

The first part of the work is on developing a guide robot system which observes people's interests and intentions towards paintings in museum scenarios and proactively offers guidance to them using a guide robot, if needed. To do that, multiple USB video camera sensors are utilized to support the guide robot in detecting and tracking people's visual focus of attention (VFOA) toward paintings. Further, each person's head orientation and profile information and computed importance values are considered as *local behavior* to identify a target-person that may be interested in a particular painting. After identifying the target-person, the guide robot moves autonomously through an appropriate motion path from the so called *public-distance* to his/her *social-distance* to explain details about the painting to which s/he is interested. Furthermore, the viability of the proposed guide robot system is demonstrated by experimenting with the *Robovie-R3* as a museum guide robot. Finally, the system is tested to validate its effectiveness. Continuing to improve the recognition of people's interests, intentions, and preferences concerning paintings in the museum, a network enabled sensing system is designed and implemented by incorporating different sensing modalities in combination where sensors are distributed in the environments as opposed to conventional sensing systems that are usually on-board the robot. This network enabled sensing system may assist the guide robot to recognize human intentions before proactively approaching people that may want guidance or commentary about the paintings. To do that, first, observational experiments are conducted in a museum with participants. From these experiments, mainly three kinds of walking trajectory patterns are found, which characterize *global behavior*, and additionally, visual attentional information are also found that indicates the *local behavior* of the people. These behaviors ultimately indicate whether certain people are interested in the exhibits and could benefit from the guide robot system providing additional details about the paintings. Based on the findings, a network enabled Human Robot Interaction (HRI) system is designed and implemented for the museum. Finally, the viability of the proposed HRI system is demonstrated by experimenting with a set of *Desktop Robots* as guide robots. Experiments reveal that the proposed HRI system is effective for the network enabled *Desktop Robots* to proactively provide guidance.

To detect and track all the people inside any real public social spaces for reading an individual's interests, intentions as well as extracting knowledge on their actual expectations from their surroundings, a social robot should have robust human sensing systems. Most state-of-the-art human sensing systems fail to track any initially detected person, especially in crowded large scale social spaces where potential partial and full occlusion between persons and/or objects frequently happen. To combat this issue in observing people's behaviors for social robots, in the final part of this dissertation, a new method is introduced which uses LIDAR to identify humans and track their positions, body orientation, and movement trajectories in any public space to read their various types of behavioral responses to surroundings. We install a network of LIDAR poles at the shoulder level of typical adults to reduce potential occlusion between persons and/or objects even in large scale social environments. With this arrangement, a simple but effective human tracking method is proposed that works by combining multiple sensors' data so that large-scale areas can be covered. How valuable information related to people's behaviors can be autonomously collected and analyzed using this method is also described. Additionally, a solution to visualize people's movement patterns and preferences with respect to any social space is presented. Thereafter, the effectiveness of the proposed human detection and tracking method is evaluated in an art gallery of a real museum. Ultimately, results revealed good human tracking performance and provided valuable behavioral information related to the art gallery which are very important to deploy in any museum guide robot system in the future.

Keywords

HRI, network enabled sensing system, human behavior tracking, local behavior, global behavior, VFOA, gaze direction, head orientation, walking trajectory patterns, stereotypical movements, network enabled HRI system, museum, visiting styles, animal behaviors, museum guide robot, LIDARs, art gallery visitor, participants impressions, accuracy index, evaluation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

As technology advances, robots will soon be a part of our everyday lives. As a consequence, with the development of robotic technologies over the last decades, robots are already appearing in industry and becoming more commonplace in human society. As such, social robots will play an important role as partners of human beings. From the deployment of social robots in human society, the new field of Human Robot Interaction (HRI) has emerged. However, in early work, the main goal of HRI research has been to establish interaction channels between social robots and humans in social perspectives. With these goals, early social robots were deployed in many social spaces and used as tutors in schools [126], as museum guides [29], as receptionists for visitors [70], in the context of mental-care for elderly people [201], in autism therapy [119], child care [189], and so on. Gradually, social robots achieved the capabilities to take orders and serve customers in restaurants [36], built trust [158], express emotions [167], provide varieties of reactive advice or information to the people in many public spaces, and be recognized in some capacity as social peers. But in most social spaces, social robots were only bound to engage in conversation and social interactions with those people who proactively chose to interact with it.

Nowadays HRI research is investigating the elements necessary to create social robots so that they are able to act in advance to deal with expected difficulties. In such cases, a social robot recognizes human intentions and then offers assistance or gives a solution to them proactively. If social robots were enriched with such capabilities for real world environments then they would offer several benefits for human society. One of the benefits of the social robots is that in its proactive conditions the more human-like interaction makes the social robot less machine-like [107] because the social robots would be able to read human intentions before offering their proactive services. Dealing with developing such types of social robot's proactive behaviors is quite difficult. However, recently the HRI research community has been dealing such types of situations (for example, [69, 103, 104]) where perceiving human behaviors so that they can offer services to humans who seem to need or want their potential services. These types of research activities are emerging for social robots in which they can proactively serve people as opposed to the conventional reactive approach where robots wait until humans explicitly request them for their services.

In general, we can often tell what other people would like to do by observing their behaviors in social spaces. There are various such behaviors: *global behaviors* such as walking trajectories patterns, and *local behaviors* such as eye gaze patterns. Usually, such types of human behavioral information helps the people in public spaces (for example, museums, shopping malls, train stations, exhibition centers) to read other people's interests, intentions, and preferences to know whether anyone needs any services or not before offering proactive assistance. But, practically in a large scale environment, making manual large scale observations of human behaviors using only a limited number of human (guides/staff) is a very difficult and complicated task. Thus, to make up for the shortcomings in observing human behaviors in public spaces as well as offering proactive services, social robots can play important roles by utilizing modern human sensing technologies and their proactive behaviors. In recent years there have been a variety of initiatives involving the introduction of emerging human sensing technologies in supporting social robots to perceive human behaviors. But, in practice, it is difficult to devise a single sensor system to detect and recognize the *local* and *global behaviors* of all the possible people inside the large scale observation area. The demonstration of a working system with such abilities would present a new research direction in introducing the use of proactive social robots services for humans.

So, it is very necessary to design a human behavior perceiving system by which a social robot can observe people's behaviors inside any public space to read their interests, intentions, and preferences concerning surrounding environments. A robust human behavior perceiving system may enable a social robot to proactively approach those people who are in need. This dissertation addresses this issue through developing a real life museum guide robot system as a testbed by which a museum guide robot will approach humans proactively to offer guidance about the exhibits after perceiving their behaviors with, the help of environmentally distributed human sensing system.

1.2 Objectives

The above described future visions for social robots motivated further work during this doctoral study. Thus, the dissertation deals with enabling social robots with capabilities to guide humans in future museums as safe, natural, and socially acceptable actors. More specificically, the emphasis is on enabling a social robot to act as a museum guide robot so that it can perceive human behaviors to read their interests, intentions, and preferences inside a museum using modern remote human sensing technologies. Our intention, however, is to share an approach that can be useful in social robot deployment scenarios like those explored by other works; for example, pedestrian guidance [58], assisting people in schools [101], shopping malls [69, 103], and hospitals [146]. We propose a museum guide robot system, in which environmental distributed sensor networks are used to augment a social robot's recognition capabilities to observe human behaviors to identify those humans who may need guidance. Based on the observations, the museum guide robot provides guidance proactively to humans, as shown in Figure 1.1. Additionally, we propose a robust human detection and tracking system for social robots so that it can track all the people in large scale real public spaces even in crowded situations by combating against partial and full occlusion. Some of the challenges have led to interesting questions, which are investigated through the dissertation.

The research questions are:

•**Person detection and tracking:** How can social robots detect and track humans in any given social space by using wearable-free (environmentally distributed) sensors ?

Figure 1.1: Example photo of museum guide robot: Illustrating museum guide robot's proactive guidance to the humans in an art gallery of a real museum scenario.

•**Intention and interest recognition:** How can social robots find out whether some one is interested in something (human/object) from his/her surrounding environment, based on previous observations ?

•**Proactive approach:** How a social robot's behaviors can be designed to facilitate its sociable and comfortable proactive mannerisms towards humans ?

•**Evaluation:** How are the abilities of the museum guide robot evaluated and tested on a prototype social robot ? How are the abilities of wearable-free sensor systems evaluated in tracking humans in an art gallery of a real museum?

1.3 Research Contribution

This research resulted in designing museum guide robot systems with the ability to proactively offer guidance to humans based on perceiving their interests, intentions, and preferences towards the exhibits using robust environmentally distributed human behavior tracking systems.

•An experimental paradigm for studying how a museum guide robot system detects humans inside the museum and tracks their *local behaviors* to estimate their interests, intentions and preferences towards the exhibits before proactively offer guidance about any exhibit (Chapter 3).

•An experimental paradigm for studying how *Roboevie-R3* as a museum guide robot proactively approaches the target human at the beginning of guidance about any exhibit. Furthermore, the effectiveness of this museum guide robot system is validated through tests and evaluations (Chapter 3).

•An observational experiment paradigm for studying humans interests and preferences toward exhibits based on their *local* and *global behaviors* inside the museum (Chapter 4).

•An experimental paradigm for studying how a network enabled wearable-free human sensing system determines visitors' interest and intentions concerning exhibits in a museum by observing both their *local* and *global behaviors* (Chapter 4).

•An experimental paradigm for studying how a designed and implemented network enabled guide robot system proactively provides additional details to the people about the exhibits based on the their observed *local* and *global behaviors* inside the museum (Chapter 4).

•An experimental paradigm for designing a robust human behavior tracking system by which a robotic system can track people's position, body orientation, and movement pattern inside any social public spaces even in crowded situations by combating against partial and full occlusion between humans and objects (Chapter 5).

•An experimental paradigm for describing how the deployed robust human behavior tracking system in an art gallery of a real museum can autonomously collect and analyze valuable information related to human behaviors about the interests and preferences towards exhibits (Chapter 5).

•An experimental paradigm for describing how the autonomously collected data by the deployed robust human behavior tracking system in an art gallery is very important in making decisions on improving the services and guide system (Chapter 5).

1.4 Organization of Sections

Chapter 2 -Interdisciplinary Background: This chapter provides the interdisciplinary background and current state-of-the-art related to the different subjects treated in this dissertation.

Chapter 3 - A Vision Based Guide Robot System-Initiating Proactive Social Human Robot Interaction in Museum Scenarios: In Chapter 3, a guide robot system is presented which observes people's interests, intentions and preferences towards paintings by tracking their *local behaviors* in museum scenarios and proactively offers guidance to them using a guide robot, if needed. The proposed model of expected behavior of the guide robots to initiate interaction with the human is explained in this chapter. The results of our experiments are reported at the end of the chapter.

Chapter 4 - Network Guide Robot System Proactively Initiating Interaction with Humans Based on Their Local and Global Behaviors : A network enabled HRI system is proposed which can determine people's interests, intentions, and preferences concerning exhibits in a museum by tracking both their *local* and *global behaviors* using different sensing modalities in combination. The experiments are conducted with a set of *Desktop Robots* to demonstrate the viability of the proposed HRI system, which is described at the end of this chapter.

Chapter 5 - Robustly Tracking People with LIDARs in a Crowded Museum for Behavioral Analysis: In this chapter a LIDAR based people behavior tracking system is presented, which is robust in tracking people even in a real world public space. The effectiveness and usefulness of the proposed system in an art gallery of a real museum is evaluated at the end of this chapter.

Chapter 6 - Conclusions : We conclude the dissertation with a summary of the concepts and designed human behavior tracking and HRI systems introduced in this dissertation followed by the potential future works and application.

Chapter 2

Interdisciplinary Background

This chapter provides the literature review covering the theoretical background and the state-of-the-art research accomplishments in the areas of Social Robotics and Human Robot Interaction (HRI) in order to frame the context of the presented dissertation work. The research topics of this dissertation, which are outlined in Section 1.2, can be categorized into three different research areas in HRI. First, detection and tracking people to observe their behaviors to recognize their interests, intentions and preferences concerning surrounding environments. Secondly, the different modes of approach for social robots to provide services to people based on their observed behaviors. Finally, the behaviors of the social robots during interaction with humans. These three research topics are highly linked to HRI research.

Therefore, first a brief and general overview of Social Robotics is given in Section 2.1. Next, we begin by highlighting the purpose of research in HRI and its supporting research fields are presented in Section 2.3. Then, the more detailed state-of-the-art descriptions of the related technologies to detect and track people are presented in Section 2.3.1. In Section 2.3.2, a description of current HRI research is provided where consideration is given towards observing people's behaviors for supporting robots to estimate their interests and intentions. Finally, modeling the behaviors of social robots (Section 2.3.3) from the perspective of interaction with people (Section 2.3.4) in public spaces are described. Each section gives an overview over the state-of-the-art research and concludes with an evaluation of how the work in this dissertation fits into the field.

2.1 Definitions of Social Robots

At the beginning of the robotic era, robots have been loosely defined as "reprogrammable multi-functional manipulators" that were designed to move materials, parts, tools, or specialized devices through variable programmed motions for the performance of a variety of industrial tasks. But over the last decades, robots have moved form purely industrial use into human society to play an important role in society as a partner of human beings in daily life. Thus, the above mentioned definition is outdated for those robots that are called "social robots". So, we can define an autonomous robot as a social robot if it can interact and communicate with humans or other autonomous physical agents by following social behaviors and rules attached to its role. Social Robotics is a fairly recent branch of robotics.

2.1.0.1 Socially Interactive Robots

Fong et al. [53] compiled several observations concerning socially interactive robots after an extensive survey of social robots. From their point of view, "Social robots are embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other" [53]. Therefore, the socially interactive robot requires some specific capabilities: it has to be able to express and perceive emotions, communicate with high-level dialogue, and learn and recognize models of other agents. Furthermore, it has to be capable of establishing and maintaining social relationships, using natural cues (gaze, gestures, etc.), and exhibit distinctive personality and character. Finally, the robot may also develop social competencies.

2.1.0.2 Sociable Robots

According to the study by Breazeal in [24], a sociable robot is able to communicate with us, understands and even relates to us, in a personal way. It should be able to understand humans and itself in social terms. In turn, human beings should be able to understand the robot in the same social terms to be able to relate to the robot and to empathize with it. Such a robot must be able to adapt and learn throughout its lifetime, incorporating shared experiences with other individuals into its understanding of itself, of others, and of the relationships they share. In short, a sociable robot is socially intelligent in a human-like way.

2.1.0.3 Design-Centered Social Robots

A design-centered social robot is an autonomous or semi-autonomous robot that interacts with humans by following the behavioral norms expected by the people with whom the robot is intended to interact. Their definition presupposes three conditions: the robot has to be autonomous, depending on the case it has to interact cooperatively or non-cooperatively, and it has to recognise human values, roles, etc.

2.1.1 Towards a Definition of Social Robots

It can be easily realized from the above discussion that social robots contain aspects about form and function. However, within the definitions of industrial robots the form is not mentioned. Thus, there is a difference between (industrial) robots and social robots relating to form. All the researchers in [16, 24, 53, 164] argue that the embodiment and form of social robots are important aspects. In human society, humans have different expectations due to the aesthetic form of robots. The aesthetic form communicates social cues and signals and the behavior of a robot is mediated somehow through its physical form.

In comparison with an industrial robot, a social robot combines technical aspects as well as social aspects but the social aspects are the core purpose of social robots. The industrial robot is not a social robot, because it needs specific communicative capabilities to become a social robot. First, it implies the robot behaves (functions) socially within a context and second, it implies the robot to have an appearance (form) that explicitly social with respect to any person. From this point of view, a social robot contains a robot and a social interface (see Figure 2.1). A social interface encloses all the designed features by which a user judges the robot as having social qualities.

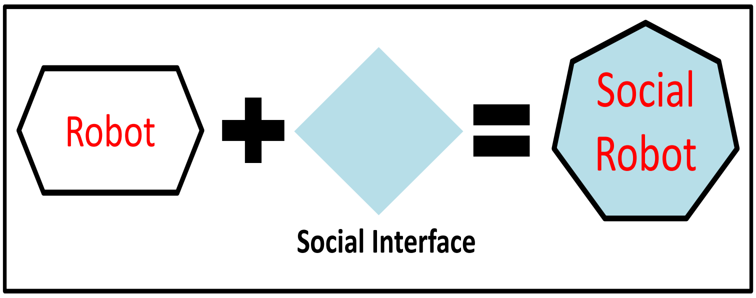

Figure 2.1: A Social Interface creates a social robot (Source: [84]).

2.2 Potential Applications of Social Robots

The main focus of this dissertation is on "social robots" in the role of peer-type human partners in social environments. Thus, applications for social robots would include services that are typically provided by people. A few examples of social robot application concepts are illustrated in Figure 2.2. This section will discuss examples of services that could be performed by social robots in real world environments.

2.2.1 Guidance Services

Social robots can provide guides in a way that are typically provided by human guides. Some examples of guiding services include providing public service announcements in public spaces, directing people in emergency situations such as the evacuation of a building, and providing guided tours or personalized proactive commentary about the exhibits in museums. Another example can be found in how the size of shopping malls and train stations continue to become larger and larger. Sometimes people get lost in such large spaces and ask for directions. Even though all such types of public spaces have maps, many people still prefer to ask for help. In such situations, social robots can play important roles, providing route guidance to direct them according to their preferences.

2.2.2 Informational Services

From the shopping mall management's point of view, information services are one of the most important facilities they need. For instance, posters and signs

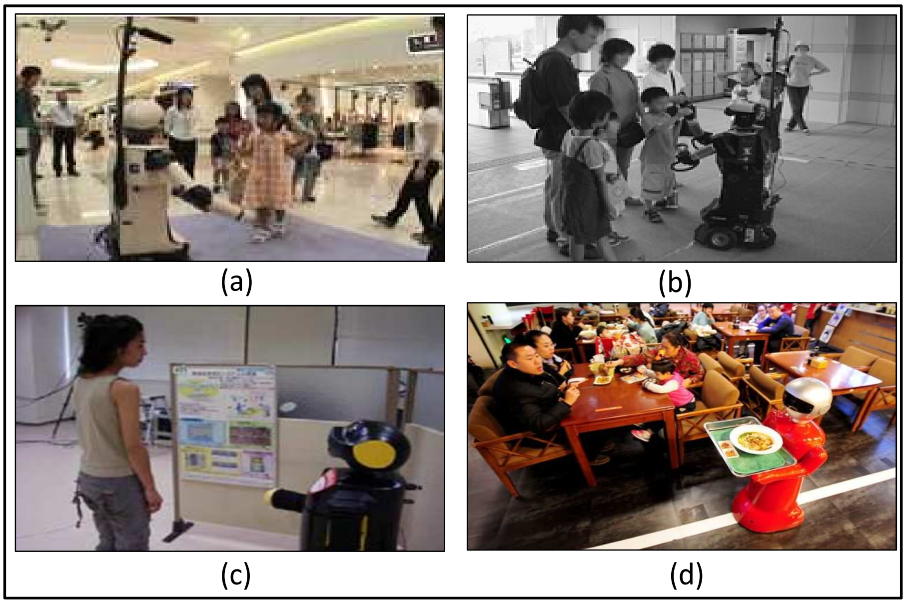

Figure 2.2: Potential application of social robot in various public space scenarios; (a) guiding customer in shopping mall (Source: [103]), (b) Providing route guidance at a train station (Source: [177]), (c) guiding museum visitor (Source: [121]), (d) serving customers in a restaurant (Source: [159]).

are placed everywhere in malls. In recent times IT services have also been put to use. We believe that a social robot can also be a powerful tool for this purpose. The physical presence of social robots in such social spaces gives them a great advantages over those posters, signs, and IT based informational services. The authors in [100] claimed that since a social robot is novel, it can attract people's attention and redirect their interest to the information it provides. Some other examples of information services could include answering people's questions at a convention or other event, and soliciting responses to survey questions [67].

2.2.3 Assistance

Social robots could provide physical, mental, or social assistance to persons who could benefit from it such as the elderly or disabled. Social robots may help them to carry bags or groceries while shopping, or help to carry heavy luggage at an airport. They can also be used to carry people in and out of bed in hospitals. Some studies have investigated the use of "intelligent shopping carts" to provide both information and baggage carrying services in a shopping scenario [75]. For example, DF Glas et al. [69] deployed non-humanoid cart robots in shopping mall to provide on-demand baggage-carrying services, to carry customer's bags to various destinations in a mall. For a closer-to-home example, if we can request that the robot take out the trash and it could find the garbage can, remove the bag, and place it in the right location on the curb, that would be seen as very helpful to have around our home.

2.2.4 Entertainment Services and Companionship

Social robots can provide service in the form of entertainment. In such situations, the value of the service primarily lies in the content of the information provided by the social robot. These sorts of services have been demonstrated in the form of a robot playing with children at an elementary school [101, 127] and day care centers [189]. In addition, there is already a great use of robots on the silver screen (for example, the *Jurassic Park* and the *Terminator* movies). On the other hand, robots are also considered as companions of humans in the public spaces where the main goal is to know the feelings of the person when interacting with the robot. In [91] the authors did some work like that with a social robot accompanying an elderly person through a super market, chatting about topics like the weather while they were shopping. Furthermore, authors on the mentioned works addressed the quality of interaction between humans and robots.

2.2.5 Autism Therapy

It has been seen from a significant amount of robotics research over the last decade that many children with autism spectrum disorder (ASD) have a strong interest in robots, and further, robots are considered as a potential tools for the therapy of ASD. Robotics research has demonstrated that many individuals with ASD express elevated enthusiasm (for example, increase in attention [49], imitation ability [55], verbal utterances [108], social activities [202]) while interacting with robots. A comprehensive survey on this application of robots for the therapy of ASD is available in [31, 168]. Recently, Begum et al. conducted a study in [17] on measuring the efficacy of robots in ASD therapy.

2.2.6 Peer, Tool, Tutorship in Education

It has been shown through years in the HRI research that social robots are more crucial for children and teenagers, where robots can be used for their development and intellectual growth. As a consequence, greater attention has already been paid to use social robots in education to provide language, science or technology education and that a robot can take on the role of a tutor, tools, or peer in the learning activity. It has been shown in that young children performed better on post-learning examinations and generated more interest when language learning took place with the help of robots as compared to audiotapes and books [94]. Nowadays, education robots are a subset of educational technology, where they are used to facilitate learning and improve the academic performance of students. An analytical overview of the prevailing fields of social robots in education can be found in [141].

2.3 Human Robot Interaction

Human Robot Interaction (HRI) is the interdisciplinary study of interaction dynamics between humans and robots. Mostly, humans express their intentions via speech, gestures, expressions, and sounds. In response, to such types of human behaviors, social robots must be aware and also be able to understand them [9]. Researchers and practitioners specializing in HRI come from a variety of fields, including engineering (electrical, mechanical, industrial, and design), computer science (human-computer interaction, artificial intelligence, robotics, natural language understanding, and computer vision), social sciences (psychology, cognitive science, communications, anthropology, and human factors), and humanities (ethics and philosophy). Human-Robot Interaction (HRI) differs fundamentally from typical Human-Computer Interaction (HCI) in several dimensions. Yanco et al. state in [209] that HRI can be seen as a subset of HCI. Figure 2.3 shows the HRI which is placed within the multidisciplinary field of research.

The HRI research area started growing roughly a one and half decade ago, and has been fast growing ever since. To highlight the research progress in HRI, the HRI community has been arranging annual conferences on Human-Robot Interaction since 2006. There are quite a few distinct subject areas related to

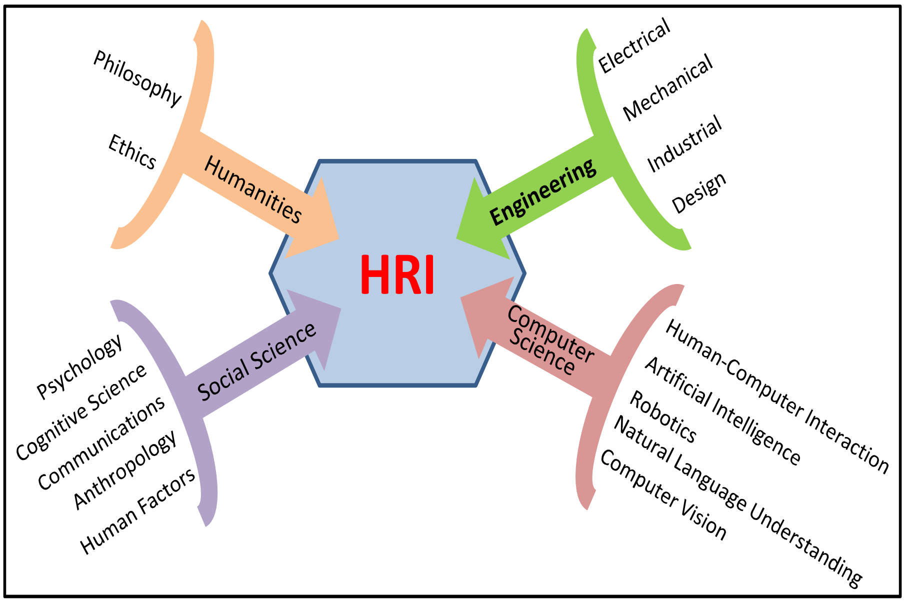

Figure 2.3: HRI- a multidisciplinary field of research.

HRI. Most of the research related to HRI can be categorized within one or more of the subjects illustrated in Figure 2.4. The categorization of these research areas is based on an analysis of HRI related publications at conferences/symposiums such as HRI, IROS, RO-MAN, ICRA, and ICSR. The research areas which are directly treated in this dissertation are marked in orange , whereas the ones marked with light green are not explicitly considered in this dissertation but have some background relation to this dissertation. The light aqua ones are not directly related to the presented work in this dissertation. The research areas which are not directly treated in this dissertation are briefly presented here:

Facial Expressions: Facial expressions are responsible for a huge proportion of nonverbal communication. While nonverbal communication and behavior can vary dramatically between cultures [88]. Social robots can express human emotions, to be able to be sociable [187].

Gestures: Deliberate movements and signals are an important way to communicate without words. Common gestures include waving, pointing, and using fingers to indicate numeric amounts [93]. Other gestures are arbitrary and related to culture.

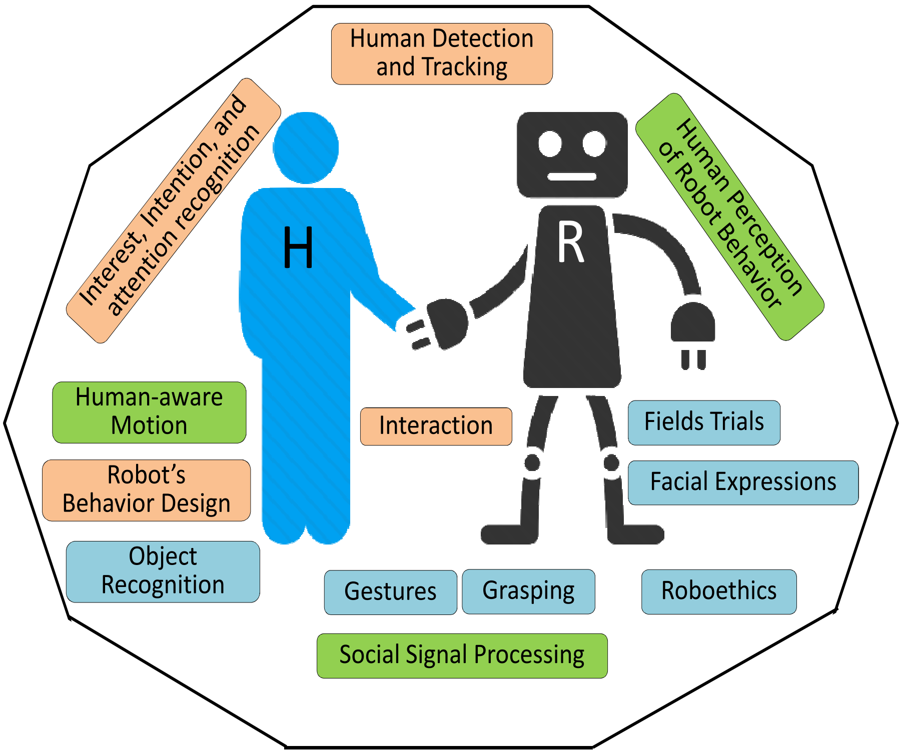

Figure 2.4: Research within HRI can be grouped into a range of areas. The orange areas are the ones that are most closely related to the work presented in this dissertation. The light green areas are relation to the dissertation, and the light aqua ones are not directly related to this dissertation.

Grasping: Autonomously grasping a previously unknown object still remains a challenging problem for a social robot. Modern social robots should be able to grasp things [162] carefully through hand-programmed or "scripted" to move in the physical world.

Roboethics: During the last decade, social robots have gradually entered many of our public spaces such as assistance robots for the elderly in day-care service centers [77], general service robots for hospitals [51] and office buildings [11], [12], visitor guide robots for museums [19], [30], [181], shop clerk robots for shopping centers [74], and general guide robots for passengers in train stations [177]. Thus, a fruitful discussion on how the social robot should be allowed to behave with humans is very relevant [187]. Movies like; Blade Runner; Terminator; A.I.; I, Robot, etc., contribute to the popular belief that robots might be dangerous when becoming intelligent parts of our everyday environments [10].

Object Recognition: The human visual system is equipped with extremely high selectivity that allows us to distinguish among even very similar objects, like the faces of identical twins [120]. Before deploying a social robot in public spaces, the visual system of a social robot should have the capability to identify objects being observed in real world spaces where objects are considered as information about the environment and for navigation. Computer vision is usually used for these tasks in robotics [13].

Field Trials: This specific area of HRI concerns experiments, where robotic systems have been put out into real world environments to validate their effectiveness to provide services to humans. So far many such field trials have been conducted by the researchers in many public spaces. For example, in shopping malls [99, 104], train stations [177], and museums [175, 194] to see how the robot behaves in the real world, or controlled real world settings.

Social Signal Processing (SSP): With this term, SSP, researchers addresses the ability to estimate human social signals and behaviors [199]. SSP is not a focus in this dissertation, but it has some resemblance to what is done, since this dissertation deals with estimating human intentions, interests, and preferences by observing their various bodily actions as behaviors for social robots.

Human Perception of Robot's Behaviors: How do people actually feel when a social robot enters or coexists in real world social environments. Many researchers addressed this topics in different real world scenarios, for example, in shopping malls [69, 99, 103], museums [173], and train stations [177]. These topics are also relevant in relation to this dissertation, where it is desired that the social robot will be a part of the social environments for the people. No experiments are done in real world environments but are done in designed real world controlled environments, where human perception is measured. However, it is relevant in relation to how and when a robot should approach humans.

Table 2.1: Relationship between space, distance, and degree of familiarity among the interacting humans and the number of interactors.

Space	Range	Situation
Intimate space	$0 \sim 50$cm	Unmistakable involvement with another body (lover or close friend)
Personal space	50cm $\sim$120cm	Comfortable separation, interaction with friends
Social space	120cm$\sim$350cm	Reduced involvement, interaction with non-friends
Public space	>350cm	Outside circle of meaningful involvement, public speaking

Human-aware Motion: In real world environments, social robots should move naturally, comfortably and safely, such that they are socially acceptable to people. Although the aim of the work in this dissertation is not centered around enabling a robot to move in the social space, there is some resemblance to what is done, since in some of our conducted experiments we deployed social robots which moved through some predefined locations to interact with humans in controlled conditions to offer proactive services. Thus, to find out how a social robot should move in a real environment, a good place to start, is to study how humans position themselves relatives to each other. Edward T. Hall proposed a theory-called "proxemics theory" on human-human spatial placement behavior in his large study presented in [79], where he categorizes distances between people into four classes depending on the degree of familiarity between interacting humans and the number of interactors. The distances are categorized as: intimate, personal, social, and public (Table 2.1). These spatial zones are static zones, which apply to how people position themselves relative to each other. Figure 2.5 shows four spaces during human-human interaction. Scientists in the robotics community have strived to move social robots in human environments for the past decades. Yoda and Shiota generated robot motion in [211] based on their own studies on how humans pass each other [210]. Pacchierotti et al. conducted an experiment where people in a corridor evaluated the behavior of a robot that acted according to the proxemic theory [154].

In [203], Walters et al. asked the participants (both kids and adults) to approach a robot, and stop at a distance, where they felt comfortable. Authors

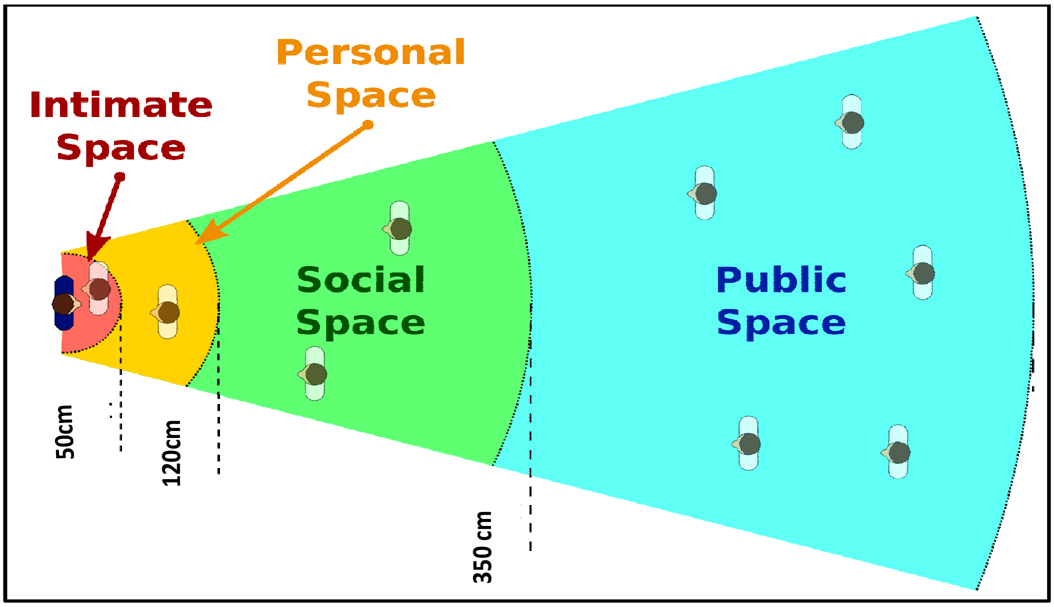

Figure 2.5: Different modes of interaction among humans (Source: [190]).

claimed that the children mostly place themselves in the social zone of the robot, but on the border to the personal zone. The result for adults was that they approach a bit closer to the robot. In the second part of the study, the roles were reversed, such that the robot approached the persons. In this experiment, the robot was allowed to approach a bit closer, than the participants approached in the first part of the experiment. Furthermore the authors concludes that there is a need for long term trials, since social distances may change over time. Follow-up studies on the stopping distance and direction of social robots prior to establishing interaction with humans were performed in [43, 44, 110, 111]. Here, various experiments are done with different approach directions. For example, in the museum scenario, Das et al. [43] claimed that the robots should select the motion path to reach a person from its public distance to his/her social distance to proactively approach him/her based on his/her orientation of attention so that the robot and the person could be face-to-face.

Based on the scope of this dissertation, the four central themes that are relevant in relation to HRI are; Human detection and tracking in spaces, human intention recognition, interaction among humans and robots, and the design of the robot's behaviors. These are described more in detail in the following sections.

2.3.1 Human Detection and Tracking in Spaces

Human detection and tracking is increasingly demanded in various applications ranging from human-computer interfaces to robotics, crowd analysis, surveillance, automation, and medical purposes. Specifically, in robotics, to implement a successful HRI system, a robot should have advanced abilities to carry out complex and complicated tasks. One such ability is robust people detection and tracking even in crowded social spaces. It has been identified as important tool in HRI not only for safe operation [170] but also for collision avoidance [178] or to implement following behaviors [60, 71], or most importantly to provide proactive services by the robot to the people through observing their behaviors about the surrounding environments. But, most of the state-of-the-art people detection and tracking systems for robots in HRI itself have on-board sensing capability can only track people at short range and are usually operated in controlled environments. Such types of on-board human sensing systems are insufficient for observing people's behaviors in real world large-scale environments for successful HRI. Thus, an environmentally distributed people detection and tracking system for social robot is desirable to meet HRI's demands in real world environments. To realize accurate detection and tracking of people in different environments, many studies have been proposed in the fields of robotics using vision sensors, laser sensors, floor sensors, and wearable sensors, or combination of them (For examples, see in Table 2.2) which are briefly described in the following sections.

2.3.1.1 Vision Based System:

Detection and tracking of persons using camera images has been reviewed in [136],[215],[192],[165]. The authors in [6, 63, 85, 95, 105, 151] have used cameras to detect the position of humans from the robot in the fields of HRI. Recent works presented in [8, 86, 179] on standard RGB camera based human detection and tracking present some impressive tracking results, even in very crowded scenes. In real-world environments, their on-board sensing system was not robust enough to track human behaviors in realtime [99]. Again, the authors in [153, 161] addressed the issue of tracking both position and body direction estimation in the fields of computer vision. In addition, the problem of using vision sensors is that they do not automatically provide the identification of each person and

Table 2.2: Different state-of-the-art human sensing and tracking modalities in different domains.

Sensing modalities	RGB-Camera	3D Range	Laser	Floor Sensor	Ubiquitous Sensor
RGB-Camera	[147], [204], [61], [95], [6], [63], [74], [42], [85], [85], [151], [105], [122], [19], [106]				
3D Range	[33], [73]	[104], [27], [26], [76]	[68], [72]		
Laser	[7], [194], [133], [113], [163], [23], [18], [59]	[3]	[97], [64], [152], [35], [87], [4], [37], [39], [66], [40], [156], [216], [18], [170], [116], [176], [174], [27], [29], [99]		[69]
Floor Sensor				[177]	[103]
Ubiquitous Sensor	[62], [117]				[175], [57], [81], [197], [78], [32], [173]

there are ambiguities when people cross paths [89]. For those reasons, the number of applications of cameras for tracking in a large public space is still small, and they are often also limited to single camera views, relatively simple and static backgrounds, or do not discuss the long-term use of the method with changing illumination. In addition, although unrelated to tracking performance, privacy issues can often pose an obstacle to the introduction of such systems in public areas [26]. Vision based trackers also generally lack robustness and scalability, especially in open environments where lighting conditions vary over time [133]. Thus, obtaining a robust and autonomous vision based solution for continuous human position detection and tracking in public spaces appears to be a hard problem.

2.3.1.2 Laser Based System

In applications where it is only desired to find the persons positions and not posture or facial expressions, laser scanner based people detection and tracking is prevalent. Therefore, Laser Range Finders (LRF) have successfully been applied to human position tracking in public spaces [11, 46, 52, 64, 66, 137, 142, 216] because it presents important advantages over vision based tracking system like high accuracy, effective sensing distance, wide view angles, high scanning rates, little sensitivity to illumination change, ease of use [5], and require far less processing than video tracking systems [102]. In addition, especially in HRI, laser range finder based human tracking has also replaced the floor sensor based human tracking systems [144] as they are easier to use and less obtrusive. The success of most of the early human-tracking systems using LRF has been based on leg tracking [25, 39, 87, 138, 170] because it is relative easy to extract information about a persons legs from a simple line scan compared to more sophisticated image processing. Kim et al. [109] presented a method utilizing laser range scanners for mobile robots to detect human legs and to follow for interaction with him/her. Recently in [64, 152], the authors proposed systems for simultaneously tracking the position and body orientation of people, using a network of LRFs mounted at torso height which are distributed in the environment but their systems are sensitive to occluded views of the person, so these can work reliably only when the number of persons is small and there are line-of-sight views between the LRF sensors and the target persons.

2.3.1.3 3-D Range Based System

On the other hand, 3D range sensing has gained more attention recently due to the increased availability of such sensors. Some recent works on human detection and tracking have applied 3D range sensors [26, 148, 171, 183]. In [160], Piérard, Sébastien et al. presented an example of their using 3D range sensors for estimating the person's body orientation. All of these works use a horizontal view of the sensors. The application to detect and track people using 3D range sensors where sensors are overhead mounted are presented in [21, 82]. But they were limited to a single sensor and did not try to detect the body angle from range data. However this would have resulted in a significant and prohibitive increase in the cost to buy, install, and maintain the tracking sensors in large area tracking, especially in public spaces.

2.3.1.4 Ubiquitous Sensor Based System

With the grace of cutting edge technologies, ubiquitous computing devices are often used, such as GPS, or the signal strength of radio (GSM, WiFi, Bluetooth, RFID, and power line) to detect and track people [100]. For example, Eagle and Pentland developed a Bluetooth-based device attached to a mobile phone that enables the analysis of activities such as being at home, at the office, or elsewhere [47]. Currently, Bluetooth technology seems to be very promising to track people in public spaces [50, 131]. Authors in [128] also used location obtained via GPS with a relational Marcov model to discriminate location-based activities such as being at home, at the office, and out dining. Besides, indoor positioning systems using Wireless LAN access points like the EAGER system [96], Museum Experience Recorder (MER) system [140] and Radio frequency identification (RFID) systems [143, 175] are also very promising to track people in public spaces. But, these technologies can be obtrusive because they require wearable or mobile devices which need to be carried by the target visitors. Thus, these approaches have a number of limitations for applications in tracking people's actual behaviors in any social public spaces. For example, in the context of an art gallery, visitors may enter the gallery spontaneously, usually pass time based on their own interests, and may not be interested in actively engaging with the technology by wearing mobile devices. Thus, such types of wearable solutions to track humans in public spaces is not adoptable to observe their actual behaviors.

2.3.1.5 Different Sensing Modalities in Combination

It can be seen that from the above discussions, the different types of sensors have different capabilities and effectiveness. For example, laser based approaches tend to work at longer ranges than vision. On the other hand, using vision enables more accurate tracking in terms of finding features like the posture of the person. Also, face tracking is important for close interaction purposes, where it is necessary to find out if a person is looking at the robot, and what kind of facial expression the person has. To combine the possible capabilities into a single system, there are also some proposed works presented in [23, 33, 34, 41, 113, 130, 133, 163, 182] to detect and track humans in public spaces and/or laboratory environments where authors utilized different sensing modalities in combination, like various kinds of vision and laser based sensors. Authors in [33, 130] used RGB-D cameras which combine 3D range and camera measurements to track humans. Again, Kobayashi and Kuno proposed to use integrated sensors that are composed of an LRF and an omni-directional camera for precise tracking of human behaviors to establish HRI [113]. Blanco et. al. presented an approach to combine LRFs and vision data to robustly and efficiently detect persons even in cluttered environments [23]. Multiple persons tracking with data fusion of multiple cameras and floor sensors are illustrated in [139]. Guide robots in [103] utilized integrated sensor systems to track people in public spaces where for position estimation, floor sensors [144] are used to accurately and simultaneously identify the positions of multiple people. Furthermore, for person identification, a passive-type RFID was tagged to every person that would always provide accurate identification. Such RFID tags require intentional user contact with the RFID reader. But, from the practical implementation point of view combining multiple sensor types make the system more problematic and complex to install and maintain, especially for large area tracking in public spaces (e.g. shopping mall arcades, train stations, and museums) [26].

2.3.1.6 Occlusion Problems and Handling in Human Detection and Tracking

Numerous human detection and tracking solutions have been presented in different research fields (stated in Section 2.3.1) to tackle various challenges, such as illumination changes, fast motion, and so on in visual tracking. Laser, 3D-based tracking systems were utilized for seamless tracking with proper identifications of people in larger spaces. But, most importantly, many trackers have ignored occlusion or handled partial occlusion of people, while at the same time, occlusion is known to be one of the most challenging aspects in human detection and seamless tracking.

In human tracking, occlusion happens when a portion (or the whole) of the observed part of the target person disappears from the observed space due to obstruction of the sensors' line-of-sight to the target person. Such types of phenomenon occur due to numerous reasons and frequently happens in real world crowded social environments where there are complex interactions between people in which temporal and/or spatial occlusions may occur. Occlusions result in errors which may eventually cause the tracker to drift away from the target or be dropped in the middle of a tracking scenario. Hence, occlusion handling is necessary to robustly track people even in crowded real social spaces. Occlusion handling is the task of minimizing the impact of occlusion on the tracking, which is achieved by granting robustness to the tracker against occlusion or preventing/compensating for the disturbing effects of the occlusion.

In contrast to the wealth of literature in human detection and tracking using various sensor modalities, the occlusion problem has received little attention. Most of the proposed state-of-the-art methods simply either ignored the occlusion problems or claimed to handle just partial occlusion owing to their robust design. But dealing with occlusions of people in the sensors views is a critical issue in human behavior tracking research as we need to continuously track their positions and body orientation to ultimately know their interests and intentions in public spaces. The state-of-the-art in handling occlusions for visual tracking is presented in [118]. Carballo et al. dealt with partial occlusion in people detection and position estimation using multiple layers of laser range finders on a mobile robot [5]. Their proposed approach is well suited to tracking people that are relatively close to the robot. However, their tracker is not suited to larger spaces where a significant number of people may gather. In other recent work, multiple laser range finders were installed in a space for human and robot tracking in [66, 152] where each of the sensors deployed around the perimeter of the space would be covered by other sensors, to minimize occlusions. But under each of their implementations, the problem of persons occluding each other in the sensor view is a serious issue in crowded social spaces for real world use.

On the other hand, 3D range sensing is also used to minimize the error due to occlusion while detecting and tracking people in public social spaces. The authors in [171] presented methods for explicitly taking into account the possibility that persons are temporarily occluded. Works presented in [26] used 3D range sensing systems where sensors are overhead mounted in the public space to allow for a good view of all the persons and minimizes occlusions due to surrounding objects or other persons, which is especially important for situations when the density of persons in the space is high. However, 3D range sensing devices are commercially expensive and used only for research purposes which would result in a significant and prohibitive increase in the cost to buy, install, and maintain in public spaces for real world use.

Thus, in this dissertation, we are interested in making use of people tracking for a wearable-free solution where people do not need to attach markers to themselves or carry special devices so that users may observe humans in an unrestricted manner. We also would like to minimize the effect of occlusions while tracking people even in crowded public spaces where nowadays HRI systems are introduced in which robots sense human behaviors using the various modern human sensing technologies.

2.3.2 Human Intention Recognition in HRI

In human societies, knowing the attention and intent of others is fundamental to collaborative interaction in which interactors work collaboratively [104]. But, in an HRI platform, identifying people's attentions and interests is a challenging task for service robots in real social environments. Usually, from people's behaviors, we can often say things like, *What they would like to do?, What are their needs?, What are their interests, intentions?, What are their actual expectations from their surroundings?.* In daily life, for example, if we find a person who is looking around with a map in his/her hand at a train station, we may offer assistance to that person. We, the human beings can determine others' behaviors by observing their various bodily actions. Such bodily actions can be observed in large areas as well as in small areas. The observation of bodily actions in a large area and small area can be defined as *global behavior* and *local behavior*, respectively. People's overall walking trajectory patterns (see Figure 2.6), such as, "entering through the entrance of an art gallery, walking across all the paintings, stopping in front of a few of the paintings, and finally, leaving the art gallery" can be good examples of people's *global behaviors*, on the other hand, people's basic motion primitives, such as fast walking, idle walking, wandering, stoping [99], various facial expressions (e.g. disgust, anger, fear, sadness, happiness, surprise), visual focus of attention, head movements, eye gaze movements can be considered as their very well known *local behaviors* (see Figure 2.7). Both people's *local* and *global behaviors* are highly dependent on the specific environments. For robots, dealing with observing such types of people's behaviors is a really challenging task. If robots could deal with such types of situations in observing people's behaviors, then it would enable the robot to anticipate the future behaviors of individuals thereby estimating people's attentions, intentions, and interests about the surroundings. But, in HRI, very few research studies have been conducted that considered the cases where robots are expected to observe both people's *local* and *global behaviors* to recognize their intentions. Table 2.3 provides a summary of the state-of-the-art HRI research where researchers considered observing either people's *local* or *global behaviors* or both in supporting robots to estimate interests, intentions, and preferences before initiating interaction with people in different social public location scenarios. For example, Kato et al. [104] developed an intention estimator for pedestrians to make a robot better initiate the interaction. The pedestrian intentions were learnt from their trajectories in [106] where the changes in velocity and distance were used as features, and intents such stopping, approaching, and following were modeled. But, the developed model was not used to make a robot better initiate interaction with human.

It is seen from the Table 2.3 that none of the state-of-the-art people's intention recognition system in HRI incorporated the sensing modalities from where the robot system estimates people's intentions by combining the bits of information from their *local* and *global behaviors*. But, considering people's *local* and *global*

Table 2.3: Related Studies in HRI concerning human behavior tracking, Intention recognition, followed by service robots' services in different scenarios

	Recognition				Services	
	Local Behavior	Global Behavior	Intention Recognition	Wearable Free?	Proactive ?	Location Scenario
W. Burgard et al. [29]	√	×	×	√	×	Museum
S. Thrun et al. [194]	√	×	×	√	×	Museum
Y. Koide et al. [117]	×	√	√	×	√	Exhibition*
M. Shiomi et al. [173]	√	×	√	×	×	Museum
R. Kelley et al. [106]	√	×	√	√	×	Personal
T. Kanda et al. [103]	×	×	×	×	×	Mall
T. Kanda et al. [99]	√	×	√	√	√	Mall*
M. Shiomi et al. [174]	√	×	√	√	√	Mall*
M. Shiomi et al. [176]	√	×	×	√	√	Mall*
M. Shiomi et al. [177]	√	×	×	√	×	Train Station*
DF. Glas et al. [69]	√	×	√	×	√	Mall*
A. Garrell et al. [59]	√	×	×	√	×	Personal
Y. Kato et al. [104]	√	×	√	√	√	Mall*
D. Brscić et al. [27]	√	×	√	√	×	Mall*
This Study	√	√	√	√	√	**Museum***

* Indicates human sensor systems are distributed in environments.

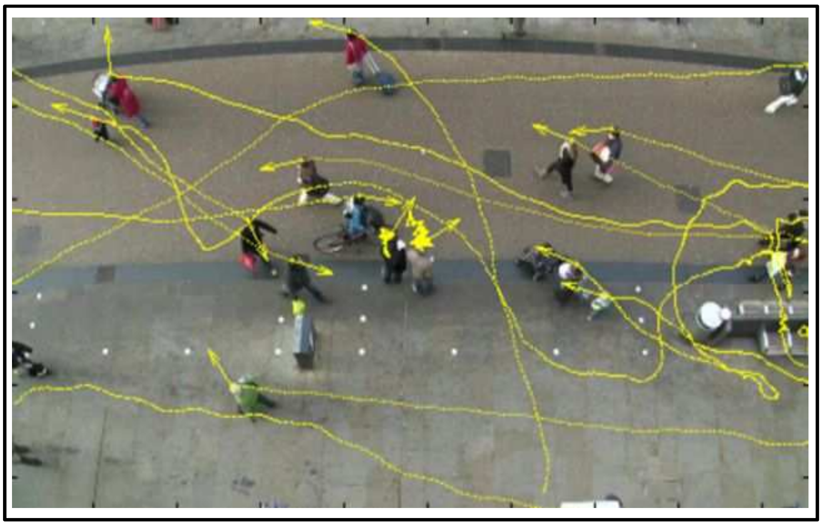

Figure 2.6: People's Walking Trajectory Patterns–an example of people's *global behavior* (Source: [212]).

behaviors is highly effective for estimating their intentions. It is also stated in Section 2.3.1 that observation of people's *local* and *global behaviors* should be unobtrusive to recognize their actual intentions because the active engagement of people with the sensory system may detract them from their actual intentions.

On the other hand, it is difficult to devise a single sensor system to detect and recognize both people's *local* and *global behaviors*. Thus it is necessary to design a sensor system for HRI which can recognize people's behaviors together. For this reason, a network sensor system is proposed in Chapter 4 for social robots, which consists of a global sensor subsystem observing a large area and a local sensor subsystem with a number of distributed sensors each observing a small area. The robot system estimates peoples intentions by combining the bits of information from the global and local sensor subsystems and proactively offers help to people in need.

2.3.3 Designing the Social Robot's Behaviors

Most of the state-of-the-art social robots (for example, ASIMO, Robovie) do not have any built-in artificial intelligence to adapt in public spaces to support human activities in real environments. Depending on the environments, the social robots' activities, objectives, and roles will be typically different. Thus social

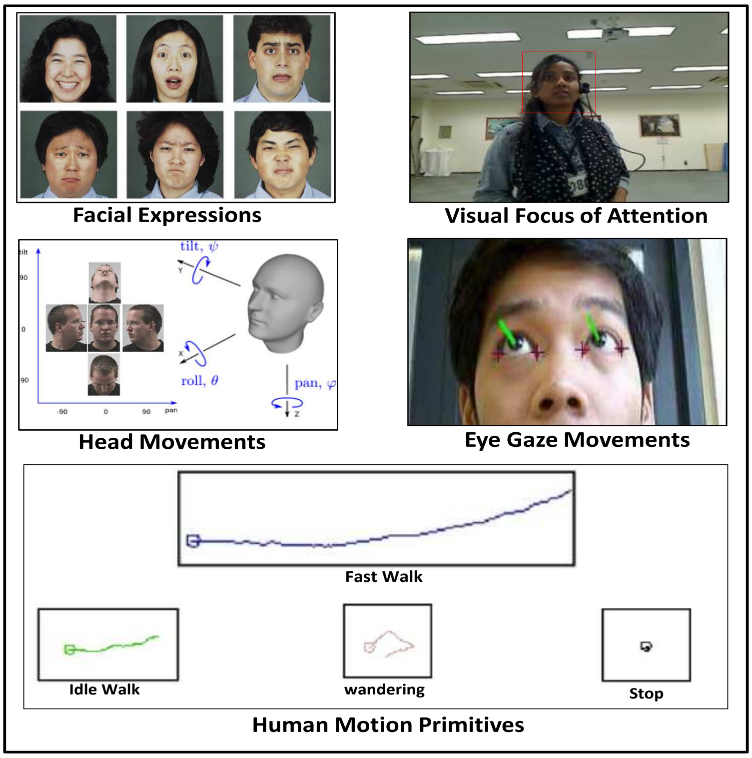

Figure 2.7: Illustration of people's various types of behaviors as examples of people's *local behavior*.

robots are programmed to perform various kinds of functions accordingly in real environments. For example, in train stations, sometimes people get lost and ask for directions. Even though every train station has maps, many people still prefer to ask for help. For example, some people may prefer to ask questions like, "Where is the platform headed to Tokyo?" as this information might be difficult to infer from a station map. Here, the robot's support would come in the form of route guidance for people. From the shopping mall's point of view, advertising is an important need. For instance, posters and signs are placed everywhere in malls. We believe that a social robot can also be a powerful tool for such purposes. Since a robot's presence is novel, it can attract people's attention and redirect their interest to the information it provides [100]. Thus, depending on the applications of social robots in real environments, a human designer should define the contents of the services as well as the context in which the robot should provide the services.

Before designing the social robot's behaviors to be able to interact with people, usually, researchers observe the behaviors of the persons who are in conversations in various real environments. Such types of observations are executed in very few HRI studies (for example, [69, 99, 103, 104, 112, 114, 206, 208]) before modeling the social robot's behaviors. For example, Kato et al. [104] observed visitors' behaviors together with the behaviors of the service staff in a shopping mall. They observed that the behaviors of the service staff is highly dependent on the behaviors of the visitors. Based on their observations, they developed a model with the focus on estimated intentions of the staff members with respect to visitors intentions. Later, the robot behaved according to the model as staff members to proactively serve future visitors in the same shopping mall. Kanda et al. in [99] present a series of abstraction techniques for people's trajectories to distinguish potential customers from the other people in a shopping mall and present a service framework for using these techniques in a social robot, which enables a designer to make the robot proactively approach customers by only providing information about the target local behavior. Also, in the context of a museum scenario, researchers undertook extensive ethnographic fieldwork in [208] on the interaction of human guides and visitors during gallery talks to discover a range of new features of conduct in the human guide's behavior. Based on their findings, the behaviors of a museum guide robot was modelled so that it could play the role of the human patron's guide in the museum. In most of their designed behaviors for social robots, the robots proactively approach humans by exhibiting verbal, gaze movements, gestural (head shaking, hand waving and body movement in between the target person and the objects) actions to draw attention and offer services. In the case of the reactive approach, social robots show their availability and recipiency by their behaviors so that people can easily start asking for help or assistance.

In this study, a museum guide robot system is developed, which proactively offers guidance to some category of visitors in the museum scenario, namely the ones who actually need guidance. Here, we developed artificial intelligence for the *Robovie-R3* and *Naoko Desktop Robots* so that they could play the role of human museum guides, in which they proactively approach visitors by exhibiting verbal, gestural actions to draw attention and offer service to the target visitors.

2.3.4 Interaction Between Humans and Social Robots

When a social robot needs to interact with humans, it should be cooperative and socially intelligent. Thus, socially intelligent robots are autonomous robots with a physical embodiments that can communicate with humans following the social rules attached to their roles [16]. Furthermore, a robot can be perceived as social when it understands and responds to people's behaviors in socially acceptable ways.

The interactions of a social robot with humans are designed to be as natural as possible, in such a way that they look like normal human-human interactions. People tend to react to systems as if they are social actors [166]. This supports the idea that theories and knowledge that apply to interpersonal relations will also apply to interaction between humans and robots. Social robots are embodied and share the same physical spaces as humans, so they should be able to interact by either touching (physically), by verbal or non-verbal means, or a combination of both of them. Physical contact with a social robot is necessary when a robot is working in a public space, for example, when in a hospital and it has to carry people in and out of bed [107]. But physical contact with a human is not essential for all social robots, especially not for a robot which has a task where it helps or guides or entertains someone in museums, train stations, or shopping malls like popular public spaces.

In the early stages of HRI, social robots interacted with human reactively. In such typical social robotic systems, people explicitly call the robot for help. There has been a great deal of research to extend the modalities so that we can use voice and gestures in these cases. In addition, Yamazaki et al. proposed that social robots should show their availability and recipiency by it's behaviors so that people could easily start asking for help or assistance [207]. In most cases, social robots wait until people willingly initiate interaction where the robot should only behave reactively, to respond to questions, for example, see Figure 2.8(Left).

On the other hand, in actual social interaction, people tend to do things automatically for each other, without being asked for help. In our daily life, for example, if we find a person who is looking around with a map in his/her hand at a shopping mall, we can anticipate on upcoming situations and be proactive. Consequently, we may ask him/her if we can help him/her, but only when the

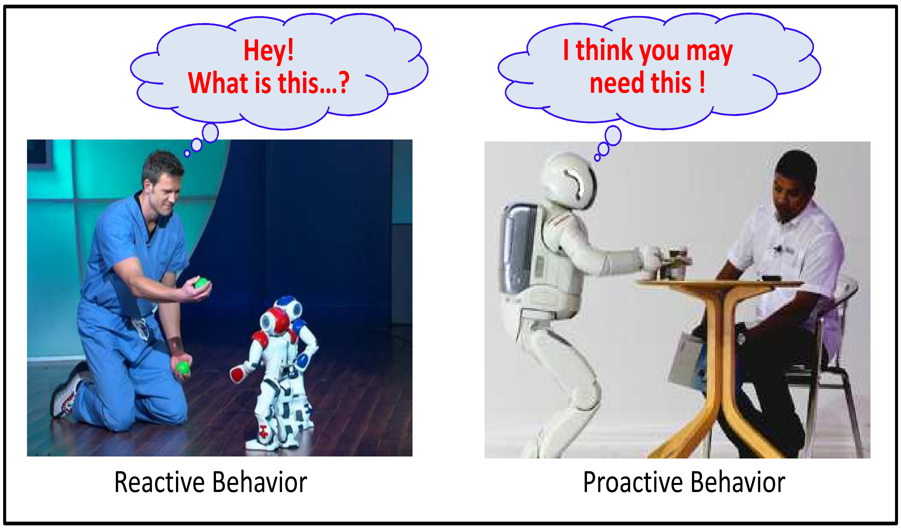

Figure 2.8: Different modes of interaction among human and robot.

upcoming situations do happen, otherwise it was a waste of time. During the middle of this era, scientist have tried to make the social robots as natural as possible to serve people proactively where social robots should estimate human intentions, and offer help only to those who would need it, for example, see Figure 2.8(Right). If it is possible to introduce the proactive behaviors of social robots then it can have several advantages, but if it does not work extremely well, it can be intrusive and disturbing. The advantages of proactive behavior are a more intuitive human-like interaction with robots [169] because the social robot reads human's intentions, s/he does not have to formulate a question (which can be difficult), thus a social robot's proactiveness reduces the human's effort. It is expected that in a more proactive state the more human-like interaction makes the social robot less machine-like [107].

So far very few works (see Table 2.3) have been proposed in HRI where the authors introduced a robot's proactive social behaviors in different scenarios. In most of these works, social robots estimate human intentions in advance by observing their *local behaviors* using different sensing modalities. The authors in [117] utilized different sensing modalities in combination in an exhibition area to enrich the guide robot with the information about each individual's history of activities as their *global behavior*, which allows the guide robot to provide them personalized guidance. But such types of sensing modalities are partially wearable to humans which highly influenced their actual intentions in the exhibition area.

In this dissertation, we intend to design a human-robot interaction framework where wearable-free behavior sensing modalities will be environmentally distributed by which social robots estimate human's intention by observing their varieties of bodily actions. After estimating intentions, the social robot proactively approach humans in socially acceptable ways to guide them in museum scenarios.

2.4 Tracking Human Behaviors in the Museum

Though the origin of paper-and-pencil based tracking of human behaviors in museums dates back to the early part of the 20th century [205], the acceptance of human behaviors observation through technology as a valid and reliable way began in the 1990s. A history of tracking visitors in museums along with a detailed explanation of methods used to record, analyze and report visitor tracking data can be found in S. S. Yalowitz and K. Bronnenkant [205]. With the grace of cutting-edge technologies, currently, technologies like Bluetooth [50, 131], indoor positioning systems using Wireless LAN access points like the EAGER system [96], Museum Experience Recorder (MER) system [140] and Radio frequency identification (RFID) systems [143, 175] seem to be very promising to observe various types of behaviors of the museum visitors. But, these technologies can be obtrusive because they require wearable or mobile devices which need to be carried by the target humans. Thus, these approaches have a number of limitations for applications in tracking human behaviors in museum like social public spaces. For example, in the context of an art gallery, visitors may enter the gallery spontaneously, usually pass time based on their own interests, and may not be interested in actively engaging with the technology by wearing mobile devices. Besides these techniques, system based on observation by video cameras [28] have also been used to gather insights into movements of humans. However, the number of applications of cameras for tracking humans are restricted to measurements in narrow fields of view. In addition, although unrelated to the issue of human behaviors observations, privacy issues can often pose an obstacle to the introduction of such system in museum like public spaces [26]. In this dissertation, a robust human behavior tracking system is developed specifically for supporting of museum guide robots for real museum application. The developed human behavior tracking system is not only helpful for deploying museum guide robots but also helpful for museum professionals and curators in adapting new services for visitors in the near future.

2.5 Museum Guide Robot

Before deploying a guide robot in social spaces, the way a guide robot approaches people should be in a polite, intuitive and appealing way [29]. There are a significant number of museum guide robot projects that have been carried out based on the robot's autonomous movement [65]. Robotic museum guides as *MINERVA* [194] and *ROBOVIE* [137] have proven to be effective in addressing people and maintaining their attention [45]. Burgard et al. in [29] presented an autonomous mobile robot, called *RHINO* to provide interactive tours to museum visitors. In [149], an autonomous robot, called *SAGE* is presented to provide educational content to people in museums. Sidner et al. in [180] designed a guide robot that was designed to explain some innovative items in a museum. These aforementioned studies have not considered gestures or other body movements during the approaching phase to begin interaction in HRI scenarios. In addition, Shiomi et al. in [173] conducted a study on HRI by deploying *Robovie-II* and *Robovie-M* as museum guide robots at Osaka Science Museum, Japan to encourage the visitors to study and develop an interest in science. Yousuf et al. [213] introduced a mobile museum guide robot that can configure spatial formations with visitors. Yamazaki et al. [208] provided a museum guide robot with resources to engage visitors in an interaction about exhibits. A novel museum guide robot is also presented by Y. Kobayashi et al. to interact with visitors through nonverbal behaviors [112]. In all these works [112, 173, 208, 213], the robot interacted with visitors via either gestures or utterances, or verbal or nonverbal behaviors or a combination of any of them in explaining exhibits. But, none of them already deployed guide robots in museum scenarios considered the robot's proactive behaviors to offer guidance to the visitors. Thus, introducing social robots for proactive guidance services for the visitors in museum scenarios is still a challenging task to the HRI research community. In this study, we integrate head gestures and other body movements of the robot during the approaching phase to proactively guide/assist target-people in more realistic ways. In this case, the guide robot proactively approaches the target-people to explain any particular painting. It is noted that the proactive behaviors of the museum guide robots to explain any particular paintings will depend on the people's observed behaviors.

2.6 Overall Summary

The goal of this dissertation work is to design a museum guide robot system which can estimate the interests and intentions of the people toward the exhibits inside the museum space with the help of modern human sensing technologies so that it can proactively provide guidance to them. To do this, theories from HRI are combined with state-of-the-art people detection and tracking research. There exist several methods for detecting and tracking persons in social public spaces using different sensing modalities ranging from vision based system to laser based system, but very few of them utilized those sensing modalities to estimate people's interest and intentions, which is very important for HRI tasked social robots to proactively provide service to the people. Researchers have also investigated how social robots should behave in social environments when people ask for help, but it has not been investigated widely how a social robot can seek out a person in need, which is also important for social robots. Motion around humans have also been investigated thoroughly with respect to how a robot should move in human environments. However it has not been investigated thoroughly how and when to move appropriately to offer proactive service, when the person seems to be interested in accepting help from someone. The next two chapters will describe a series of studies that are contextualized in and motivated by our objective to design a museum guide robot system to provide proactive guidance to the museum visitors. Each study involves a human-robot interaction situation where each system can estimate the behaviors of the peoples toward the exhibits inside the museum. After these two chapters, a chapter is explained which focused on designing a robust human behavior tracking system for real social spaces with the goal to track people even in crowded situations, and combat against partial and full occlusion cases.

Chapter 3

A Vision Based Guide Robot System: Initiating Proactive Social Human Robot Interaction in Museum Scenarios

3.1 Introduction

During the last decades, computer vision has offered the ability for service robots to understand human behavior. One of them is the visual focus of attention (VFOA) of a person as *local behavior*, which is the behavioral and cognitive process that indicates where and at what a person is looking, and that can be determined by eye gaze and/or head pose dynamics [43]. Thus, VFOA is an important key to understand human behaviors within a limited spaces. The VFOA of people can be used by service robots to estimate information such as their interests, intentions, and preferences in relation to the environment [15]. If any defined VFOA related to their interests and intentions are found from their gaze direction, then service robots can serve them accordingly. For example, with such information in an art museum, a human museum guide could guess which paintings patrons consider more attractive. Such types of information can be very helpful for developing and adapting various services for people in public spaces where service robots can exclusively provide services to them. Although dealing with such types of situations for a service robot is quite difficult, the purpose of this study in this chapter is to develop such a robotic system for serving people who are in need in a proactive manner as opposed to the conventional reactive approach where robots wait until people exclusively request them for their service.

The main research goal here is to find people that may want the service robot's help. To find such people, usually their VFOA are estimated from a computer vision point of view. It is believe that people can find such information by detecting and tracking either their eye gaze, head orientation, face profile information, or some combination of them. In this study, a solution through a museum guide robot application is demonstrated. There are several guide robots that have been proposed (for example, [29, 152]) that can provide a guided tour to people in a museum scenario. However, here developing a guide robot system is considered so that it can find a person who seems to be interested in a particular painting. In that case, the guide robot can proactively provide more information about the painting to the person to make his/her visit to a museum a more enjoyable.

When trying to find out where people are looking around in a museum, a first solution would be measure their VFOA which can be estimated by their gaze direction [186] because gaze direction typically follows the focus of his/her interest in the environment [188]. D. Todorovic et al. in [196] defined the gaze direction with respect to some references. There are several such references, and thus several different alternatives of the notion of gaze direction. One among them is *looker-related* gaze direction: the gaze direction that is determined in relation to the onlooker's head. *Observer-related* gaze direction is another way to define gaze direction where the direction of the onlooker's gaze is measured with respect to the observer (who from the on looker's point of view is just a part of the environment). The most important distinction in this case is whether the on looker looks at the observer or not. The final way to define the gaze direction is in the form of *environment-related* gaze direction: this is the direction of gaze specified with respect to some environmental reference [196]. In this study, people's *environment-related* gaze direction is considered to estimate their interest and intention to the painting inside the museum because in a museum scenario a human guide will most likely not be concerned with where the people look with respect to themselves, but rather what s/he looks at in the environment. In this experiments, environmental low cost USB video camera sensors are used as such references in the museum.

Due to the limitations (see next paragraph) of using eye gaze information to determine the VFOA of people to know their interests and intentions in museum like environments, in this study, a guide robot system is proposed in which the guide robot determines the gaze direction of people towards any paintings, but not the exact gaze point, using the head orientation information because tracking eye gaze is prohibitively difficult on low or mid resolution images [15] that are typical of low cost USB video cameras. We believe that the developed guide robot system can perform robust real-time detection of peoples interest and intention toward paintings using camera sensors in a museum scenario for providing proactive guidance to the museums patrons. For any real museum environment, a large viewing volume is desired for any guide robot system so that people in it can be tracked as they move from painting to painting. Multiple camera sensors afford this ability by providing a large combined viewing volume by which peoples attention can be tracked. The advantage of the proposed robotic system in this chapter is that multiple camera sensors cover a much larger field of view to observe people than that covered by a single camera sensor in our museum scenario.

It is noted that, many studies [38, 92, 123, 124, 132, 172] have found that vision enabled eye gaze tracking based information is not always feasible in obtaining people's VFOA due to several environmental constraints in open environments. For example, in open spaces such as offices, meeting rooms, and museums, where the motion and the head orientation of people are unconstrained, high resolution images of people's eyes are not available to track the eye center location [14]. In addition, eye-appearance vision-based gaze tracking systems restrict the mobility of people since their need of high resolution close-up eye images requires cameras with very narrow field-of-views [14]. Furthermore, detecting eye orientation from a distance is difficult in real life environments [83]. Again, it is not possible to track eyes when the eyes of a person are not visible (like low-resolution imagery, or in the presence of eye-occluding objects like sunglasses) [145]. Additionally, current eyeball orientation estimation systems either cumber users with head-mounted equipment, including cameras and special light sources, or set heavy restriction on users behavior [186]. These constraints also do not allow me to measure the VFOA using anything more than eye gaze. On the other hand, there

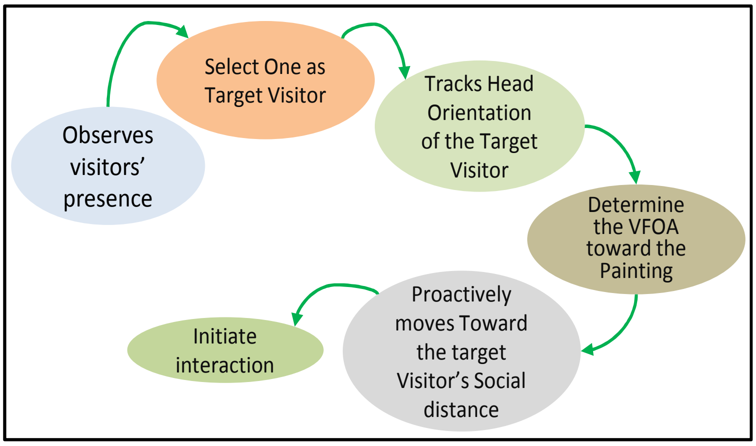

Figure 3.1: States of the proposed Human-Robot Interaction System.

are several reasons (described in [15]) to choose head orientation information to measure gaze direction (VFOA) instead of using eye gaze tracking information for the purpose of estimating the interest of the people about the paintings in museum environments.

3.2 Proposed Guide Robot System

The states of the proposed guide robot system is illustrated in Figure 3.1. In this proposed approach, a number of paintings (P_1, P_2, P_3, etc.) is considered as exhibits and a humanoid robot as the guide robot in a museum where paintings are hung on a wall at equally separated distances and the guide robot is placed at a *public-distance* [80] from the paintings and people's painting viewing regions[1]so that its presence may not interfere with each person's movements and attention. The objective of the guide robot is to provide more information proactively to people when their interest towards any painting is detected by the system.

To provide more information about any particular painting to any person to which s/he seems to be interested, first the guide robot system tries to determine the presence of people within the museum's vicinity using USB video camera sensors placed just above the paintings. If no person is detected by the sensors, then the system will continue its task to detect people. If only one person is

[1]The regions from where a person typically views exhibits in a museum.

detected by any one of the sensors then the system will treat that person as the target-person. If multiple people are detected, then the system will compute the importance value (described in Section 3.2.1.1) of each of the detected persons to select one as the target person. Thereafter, the system will obtain the head orientation information of the target person from the video frames and thereby extract gaze directional information to ultimately estimate his/her VFOA. The system will then trigger the guide robot to get the face profile information from the head orientation information to select the appropriate motion path to move from the *public-distance* to the *social-distance* [80] of the target-person to proactively approach him/her and explain the painting.

3.2.1 People Detection and Tracking Framework

The employed people visual detection and tracking system is based on the 3D head tracking method presented in [115]. This method used particle filter framework incorporating Ada-Boost based cascaded classifiers to detect and track people's head based on USB video camera captured image frames. Originally, the face detector is developed by Viola and Jones [200] using cascaded structure to reduce detection time, and to reliably detect faces without requiring a skin color model. This method works quickly and yields high detection rates [19]. Thus, with the 3D head tracking method, we can easily detect and track people in terms of tracking head orientation accurately in wider angular views in real time from video frames delivered by the USB video camera sensors where image sequences are captured at VGA resolution at a video frame rate about of 15 fps and processed by one PC (Intel(R) Core(TM) i5-2400 CPU @3.10GHz, Memory 4 GB).

3.2.1.1 Target-Person Selection Procedure

Under the proposed guide robot system, if only one person is detected by the video camera at any time then s/he is automatically selected as the target-person to track his/her VFOA. If more than one persons are detected by the cameras then our system computes their importance values individually. The importance value allows the system to choose the one target-person among all detected persons. In our present study, it currently depends on two parameters (**a**) the distance of the persons from the paintings, and (**b**) the face profile information. Here,

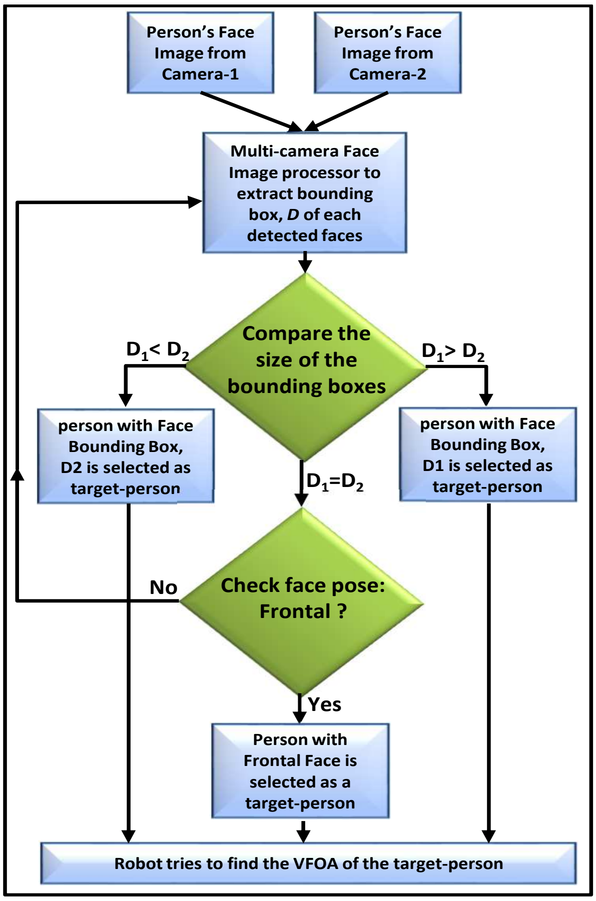

Figure 3.2: Flow diagram to obtain target-person using importance value. Two face trackers are shown to select target-person, though more could be possible.

the distance between the persons from the paintings is estimated by considering the size of the bounding box, D of their detected faces on the image planes. A person with his/her face closer to the paintings, will have a larger bounding box for his/her face and vice-versa. The procedure to compute the importance value of only two persons is illustrated in Figure 3.2, though more could be possible. Anyone with a greater sized bounding box gets a higher importance value over others. Here, System focuses its attention on the person who has the highest importance value. These scenarios are depicted in Figure 3.3(a) and (b). When the bounding box of both of the persons are the same in size the person with the most frontal face profile will be considered as the target-person. This scenario is depicted in Figure 3.3(c).

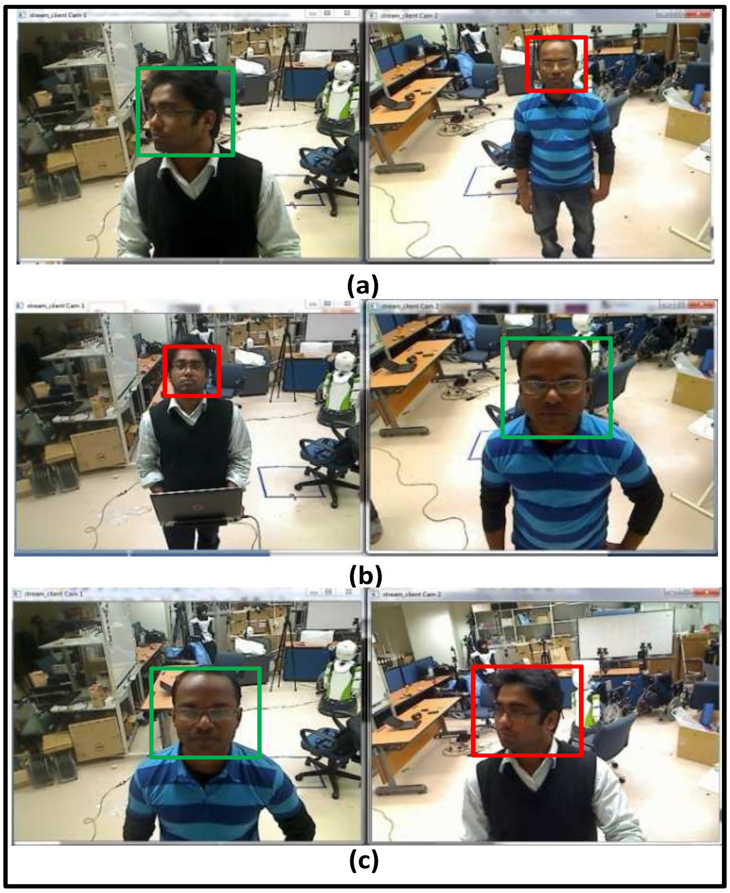

Figure 3.3: (a)-(b) People with greater green colored bounding box are treated as target-person, (c) Person with same bounding box size as other person but person with the frontal face is treated as the target-person.

3.2.1.2 Recognition of Target Person's VFOA

Usually yaw angle of human's head with respect to a video camera can be divided into four fundamental angular regions: Central Field of View (CFV), near peripheral field of view (NPFV), far peripheral field of view (FPFV), and out of field of view (OFV). Additionally, Each peripheral views (NPFV, FPFV) can be subdivided into two angular regions: Left Near Peripheral Field of View (LNPFV)and Right Near Peripheral Field of View (RNPFV), Left Far Peripheral Field of View (LFPFV),and Right Far Peripheral Field of View (RFPFV). Figure 3.4 illustrates these head orientation classifications with example face images when the camera is situated in the CFV angular region of any person. In this study, proposed guide robot system considered only the following three fundamental angular regions: The frontal view, which covers 30° (75° to 105°) is defined as the CFV. The right side angular region with respect to CFV, which covers about 45° (30° to 75°) is defined as the RNPFV. And the LNPFV is defined on the left side with respect to CFV, covers about 45°(105° to 150°). Other angular regions are not considered in this study. Once the target-person is selected, to measure the level of interest to any particular paintings, our system measures the head orientation stability by calculating the average head yaw angle for next the 30 consecutive video frames captured by the USB video camera sensor. If the target person's head orientation is stable within any particular angular region, then the guide robot will consider the orientation of his/her gaze direction toward that region and thereby measure the stable VFOA of the target-person toward any particular painting. For example, if the stable VFOA of the target-person is found in the CFV angular region, the guide robot system will detect only the frontal face of the target-person. However, in the other defined angular regions, the guide robot system may detect two face patterns which are either the 45° degree of right profile face or the 45° degree left profile face in the LNPFV and RNPFV angular region, respectively.

3.2.2 Guide Robot's Motion Path Planning

In this study, *Robovie-R3* (see Figure 3.5) is deployed as a humanoid museum guide robot to proactively offer guidance to the people in museum scenario. *Robovie-R3* itself does not have any built-in artificial intelligence, but in the implementation of my considered guide robot's behaviors, we programmed it to perform various kinds of actions. The properties of *Robovie-R3* is illustrated in Table 3.1. Because of *Robovie-R3*'s enriched human like embodiment, it is considered as a popular research platform for human robot interaction. In this works, the possible motion paths for *Robovie-R3* is restricted to only that of two people's possible predefined positions. For each position, two motion paths and positions for the *Robovie-R3* is programmed to reach the target person's *social-distance* from a *public-distance* and to initiate conversation. The schematic representation of the guide robot's motion paths for the people's two locations is depicted in Figure 3.6.

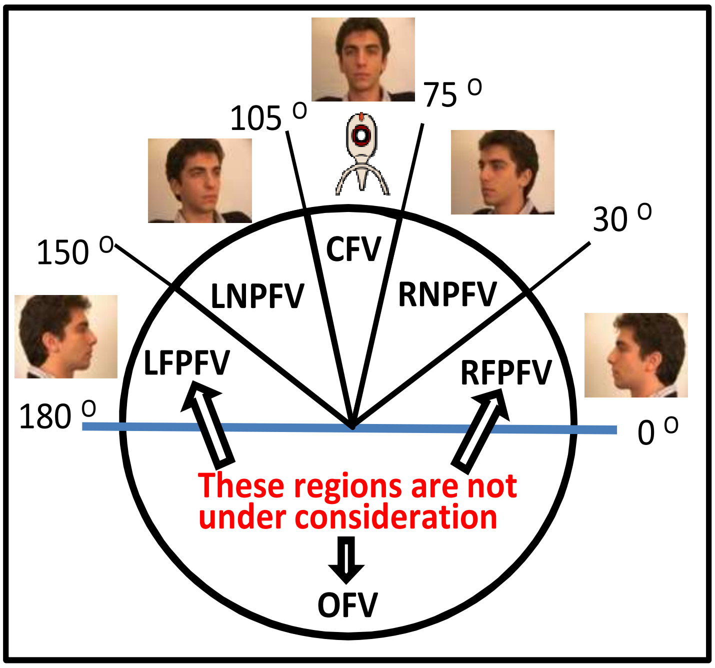

Figure 3.4: Head orientation classification into three angular regions in the proposed HRI system.

For position-1, the left-side motion path and position for LNPFV region of camera-1, and the straight motion path and position for RNPFV and CFV region of camera-1 are defined. For position-2, the right-side motion path and position for RNPFV regions for camera-2, and the straight motion path and position for the LNPFV and CFV region of camera-2 is designed. Thus, a total of four possible motion paths and positions are designed. As soon as the head stability of the target-person is found to reside in any of the angular regions described in Section 3.2.1.2, the guide robot starts to move from the *public-distance* through any of the preprogrammed motion paths to reach his/her *social-distance* to approach and explain that painting to which his/her VFOA is detected.

While the guide robot starts to approach him/her to explain any painting, the guide robot exhibits its verbal and gestural (head shaking, hand shaking and body base movement in between the target person and painting) actions to draw attention and offer commentary about the paintings. Simultaneously verbal and gestural action enhance the overall experience of the person in the museum by interacting with the guide robot.

Figure 3.5: *Robovie-R3*.

3.3 System Evaluation

Experiments is performed to verify that the proposed system is useful in museum guide robot scenario. In experiment, 10 people (8 males, average age 29.2 years) are participated from the university campus. Each person participated in the experiments four times in a total of two phases. In the first phase, two experiments were conducted to verify the flexibility of using multiple video camera sensors over a single video camera sensor based robotic system to detect and track people's VFOA towards the paintings, and in the second phase, two experiments were conducted to verify the effectiveness of my proposed system using *Robovie-R3* as a guide robot to proactively initiate interaction with target people to offer a explanation about any particular painting. In every case, people were asked to observe the exhibits from possible predefined positions randomly.

Table 3.1: Properties of *Robovie-R3*.

Parameter	Specifications
Size	108cm × 50cm × 52cm
Weight	35 kg
Degree of Freedom	17 (Eyes:2, Neck:3, Arms: 4×2, Base: 2 wheels)
Servo Motors	VS-SV1150×7, VS-SV3310J×4, MICRO STD/F×4
Motors	Max on Brushless Motor×2
Sub CPU board	VSRC003HV (ARM7 60 MHz)
I/O	Touch sensor×11, USB Camera × 2, Mono-Microphone ×2, Speaker ×1
Battery	12V 28 Ah

3.3.1 Experiment Design

In experiments, five paintings (P_1, P_2, P_3, P_4, P_5) were hung where each of the paintings are separated equidistant from their neighboring paintings and all at the same height. Two USB video cameras were mounted as environmental sensors on the top of paintings P_2 and P_4 to detect and track people's VFOA. Paintings P_1, P_2 were placed in the LNPFV, and CFV areas of camera-1, respectively whereas paintings P_4, P_5 were placed in the CFV and RNPFV areas of camera-2, respectively. Finally, painting P_3 was placed in between the RNPFV area of camera-1 and the LNPFV area of camera-2. The *Robovie-R3* was initially placed at the *public-distance* from the people's possible positions (usually far from exhibits and behind person's positions). The experimental setup for the museum scenario is illustrated in Figure 3.7. Finally, a total of four possible motion paths and positions (as described in Section 3.2.2) are designed so that the guide robot could move to suitable positions for approaching a person. Figure 3.8 shows some snapshots of the experiments. One video camera was also used in a appropriate position to document all experimental activities.

3.3.2 Experimental Cases

To validate the effective of our robotic system, two modes of experiments are conducted and compared them.

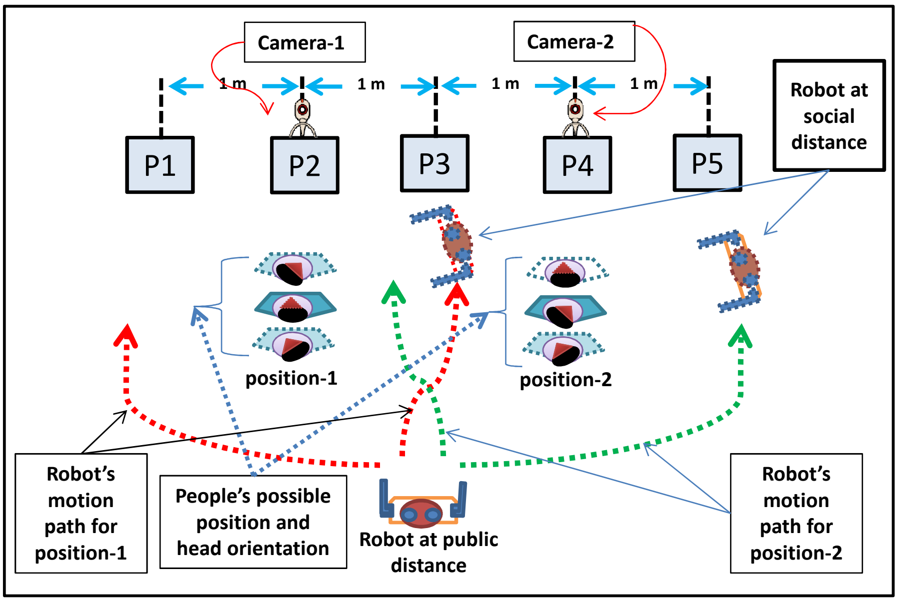

Figure 3.6: Robot's motion path and position planning scenario.

•*Multi-Sensor Mode (MSM) (Proposed mode).* In the current implementation of the proposed HRI system, the performance of the proposed HRI system is observed using two USB video cameras to select the target-person and track his/her VFOA toward the paintings while other people could also be viewing the paintings.

•*Single-Sensor Mode (SSM).* Under this system, only one USB video camera is utilized to detect the target-person and track his/her VFOA toward all the paintings while other people could also be viewing the paintings.

Additionally, experiments are also conducted with the following two modes within the proposed HRI approach to measure the effectiveness of the Robovie-R3 in dealing with guiding the target-person.

•*Single-Person Mode (SPM).* Only one person observes the paintings randomly where vision data from the two USB video cameras will be processed to detect him/her as the target-person and track his/her VFOA toward the paintings by guide robot.

•*Multi-Person Mode (MPM).* In this mode, multiple person will observe all the paintings randomly where the data of the two USB video cameras will be processed simultaneously to determine the target-person by the guide robot.

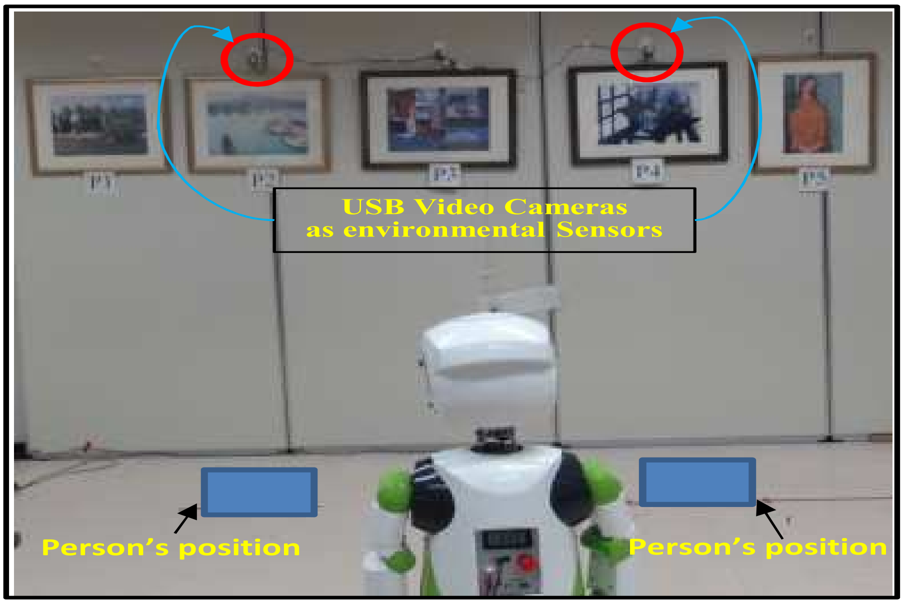

Figure 3.7: Experimental environment.

3.3.3 Measurements

In all of the conducted experiments, the following two cases are evaluated:

3.3.3.1 People's Impression

Participants were asked to fill out a questionnaire for each mode after completion of experiments. The measurement was a simple rating on a Likert scale of 1 to 7. There were three items in that questionnaire: (**Q1**) Did the robotic system estimate your gaze directions (VFOA) while you were viewing all the paintings (P_1 to P_5)?, (**Q2**) Did the robotic system effectively detect and track your VFOA while you were viewing any of the paintings?, (**Q3**) Was the *Robovie-R3* effective in approaching you in a timely manner to start interaction with you?

3.3.3.2 Success Rate

From the documented videos and experimental site, how many times the *Robovie-R3* detected the target-person's VFOA and established a successful interaction under the proposed approach is observed. The *Success Rate* (*SR*) was measured

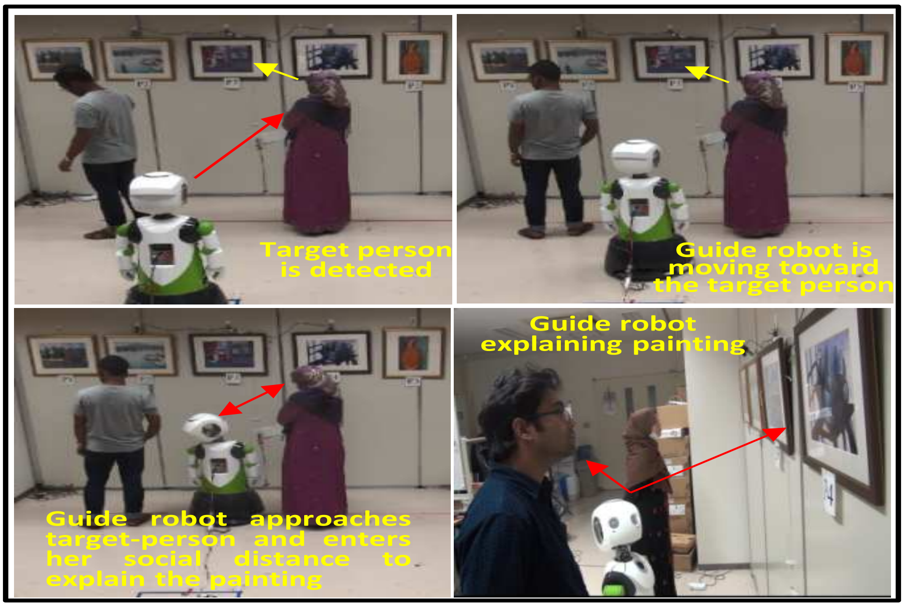

Figure 3.8: Some snapshots of conducted experimental scenes.

by the following formula:

$$SR = \frac{\text{The total number of successful interaction}}{\text{Total number of attempts that the } \textit{Robovie-R3} \text{ made}} \times 100 \qquad (3.1)$$

3.3.4 Results

The experiment was conducted in a within-participant design, and the order of all experimental trails was counterbalanced. We performed the repeated measures ANOVA for all measures.

3.3.4.1 People's Impression:

Subjective measures for all the modes are shown in Table 3.2. The result shows the questionnaires measure which represents the means (M), and standard deviations (SD), F, $p-value$, and η^2 in each condition for the participants.

For **Q1** (Table 3.2, 2$^{\text{nd}}$ row), the ANOVA analysis shows statistically significant differences between the modes (MSM and SSM) ($F(1,9) = 240, p < 0.01$, $\eta^2 = 0.88$) for the people. Thus, these results reveal that the performance of the multi-camera sensor based robotic system (proposed HRI system) outperforms

Table 3.2: People's Impression for various Questionnaires where 7 is strongly agree.

Questionnaire	Modes	Mean	SD	$F(1,9)$	$P-value$	η^2
Q1	*SSM*	2	0.67	240	0.00000008	0.88
	MSM	6	0.45			
Q2	*SPM*	4.4	1.16	19.23	0.001759	0.49
	MPM	6.3	0.9			
Q3	*SPM*	4.2	3.067	12.35	0.00656	0.25
	MPM	6.0	2.22			
Combined Q2 And **Q3**	*SPM*	4.3	1.07	25.62	0.00068	0.44
	MPM	6.15	1.4			

the single-camera sensor based robotic system in detecting people and tracking their VFOA during their viewing of all the paintings.

It is because multiple video camera sensors afford the ability to *Robovie-R3* to cover a much larger field of the viewing volume to detect people and track their VFOA towards all paintings.

For **Q2** (Table 3.2, 3$^{\text{rd}}$ row), the differences between the two modes were statistically significant ($F(1,9) = 19.23, p < .01, \eta^2 = 0.49$). Thus, these results mean that the proposed HRI system is more effective in detecting and tracking the target-person's VFOA when more than one person is detected by the guide robot.

In the case of the *SPM*, sometimes missed face detections occurred in cameras where actual human faces were not visible. In addition, static environmental objects (the robot, furniture) would sometimes be falsely detected as faces. But in case of the *MPM*, multiple persons are detected by the multiple video cameras and the target-person is chosen from among them by computing importance value (described in 3.2.1.1). In such a case, missed detections less likely appended while selecting the target-person.

For **Q3** (Table 3.2, 4$^{\text{th}}$ row), significant differences between the two modes ($F(1,9) = 12.35, p < 0.01$, $\eta^2 = 0.25$) were found under the proposed robotic system. Most of the participants felt that *Robovie-R3* was more effective in the case of *MPM* over *SPM* to approach the target-person accurately to explain the painting to which s/he was interested. The reason is that in case of *SPM*, when missed detections and tracking occurred, the system would cause the guide robot moved move to locations different from where the target person was really viewing the painting.

To obtain the overall impression of the people in between the two modes (*SPM* and *MPM*) for **Q2** and **Q3**, the average Likert scale for each participant is calculated and performed an ANOVA analysis on the Likert scale. The 5$^{\text{th}}$ row of Table 3.2 shows the result of the analysis. It is found that there were also significant differences between people's impressions of the two modes in both **Q2** and **Q3** ($F(1,9) = 25.62, p < .01$, $\eta^2 = 0.44$).

3.3.4.2 Success Rate

10 participants experienced a total of 48 trials as the target-person under the *MPM* settings. Among 48 trials, proposed HRI system was able to detect and track their VFOA and finally make successful interactions at a rate of 87.5%. Thus it can be said that proposed museum guide robot system is effective for initiating interaction with the target-person.

Thus, the experimental results revealed that the proposed robotic system is more effective for target-person detection from multiple people and tracking of his/her VFOA using multiple USB video cameras to obtain his/her interest and intention toward the painting. It is also revealed that *Robovie-R3* is much more operative at proactively initiating interaction with the target-person with its mobility features in the museum scenario.

3.4 Chapter Summary

This chapter presents a guide robot system which observes people's interest and intention towards paintings in museum scenarios and proactively offers guidance to them using a guide robot, if needed. To do that, multiple USB video camera sensors are utilized to support the guide robot in detecting and tracking people's visual focus of attention (VFOA) as *local behavior* toward paintings. In this study, each person's head orientation, profile information and compute importance values are considered to identify a target-person that may be interested in a particular painting. After identifying the target-person, the guide robot moves autonomously through an appropriate motion path from the so called *public-distance* to his/her *social-distance* to explain details about the painting to which s/he is interested. Furthermore, in this chapter, the presented guide robot system is demonstrated to proof it's viability by experimenting with the Humanoid Robot-*Robovie-R3* as a museum guide robot. Finally, the system is tested to validate its effectiveness.

3.4.1 Limitations

In this chapter, the positions of the persons are considered to be fixed to some particular locations that allow them to view all considered paintings. Only the VFOA is detected and tracked as *local behavior* to observe people's interests towards the paintings. It is revealed in our study that there are many other situations where people may view paintings from different locations and ways based on their own choosing. Thus, observing only *local behavior* is not enough to know someone's preferences towards the paintings. It is effetive to observe both *local* and *global behavior* towards paintings to precisely identify people's interests, intentions, and preferences towards the paintings. To do so, in the next chapter, a multiple LIDAR poles based *global behavior* tracking system together with a USB video camera sensor based *local behavior* tracking system is utilized for the robotic system to know the interests, intentions, and preferences of the people in museum-like public environments, which will provide more information to establish effective social interaction between people and the guide robot.

Chapter 4

Network Guide Robot System Proactively Initiating Interaction with Humans Based on Their Local and Global Behaviors

4.1 Introduction

With the development of HRI systems over the last decades, social robots have started to move from laboratories to real-world environments [189], where a robot interacts with ordinary people. But, in most of these typical HRI systems, the interaction partners (human and robot) are restricted to controlled conditions. There has been a great deal of research to extend to more interaction modalities such as voice and gesture [157]. In addition, Yamazaki et al. proposed that robots should show their availability and recipiency by nonverbal behaviors so that people can more easily ask robots for help [207]. But in real-world environments, the identification of people in need is also an important task for service robots. In daily life, for example, if someone find a person who is looking around with a map in his/her hand at a train station, s/he may offer assistance to that person. Dealing with such types of situations for a robot is quite difficult. However, very few research studies (for examples, [69, 103, 104]) have been conducted in the fields of HRI that consider cases where robots are expected to offer their services to people who seem to need or want their potential services.

It is noted that, there are several systems that address related issues in HRI. For example, robots have been deployed in public spaces, including day-care service centers [77], hospitals [51], train stations [177], office buildings [11, 12], museums [30, 150, 193], and shopping centers [104] that address navigational and perceptual problems. Service robots deployed in child care centers [189], autism therapy centers [17, 119] and in schools [101, 127] has addressed the quality of interaction between humans and robots. In addition, estimation of the intentions of the surrounding humans towards the robots are also highlighted in these works. Moreover, many robots have also been deployed in public spaces with the capability to encourage people to initiate interaction with them [20, 44]. What differentiates the works presented in this chapter is that here an HRI system has been developed for public spaces in which service robots can proactively serve people as opposed to the conventional reactive approach where robots wait until people explicitly request them for their service. The main research topic here is to find people who may want the robot to serve them.

In general, people can often tell what other people would like to do from their behaviors. There are various such behaviors: *global behaviors* such as walking trajectories, and *local behaviors* such as gaze patterns. It is difficult to devise a single sensor to detect and recognize all these behaviors. Thus, in this chapter, a network sensor system is presented and additionally, a network robot system is presented that uses all sensor data. The sensor system consists of a global sensor subsystem observing a large area and a local sensor subsystem with a number of distributed sensors each observing a small area. The robot system estimates people's intentions by combining the bits of information from the global and local sensor subsystems and proactively offers help to people in need. In this study, this issue is addressed by taking museum guide robots as an example and develop a network-enabled HRI system which can find people who seem to be interested in a particular exhibit and offer them accordingly, more detailed explanations about that exhibit by any one of the guide robots in the network robot system.

In addition, there has been much research on technologies that are employed to track people's *local* and *global behavior* in the fields of robotics and computer vision [26, 136]. In ubiquitous computing, positioning devices are often used. These include the use of GPS, or the signal strength of radios (GSM, WiFi, Bluetooth, RFID) [103]. These technologies all used wearable or mobile personal devices, but these approaches have a number of weaknesses for applications in large public spaces. Thus, in this work, a wearable-free solution is introduced where people do not need to attach markers to themselves or carry special devices, as the main goal is to observe the unrestricted interests and intentions of all the people in public spaces. Hence, a global sensor subsystem is used which is based on LIDAR (Laser Range Finder) poles [152] to track people freely and to obtain their *global behavior* information (i.g. walking trajectories). Additionally, a set of low cost USB video cameras is used as a local sensor subsystem to obtain people's *local behavior* information (e.g. visual attention). That is, the presented HRI system in this chapter tracks the interests and intentions of people in public spaces using technologies that incorporate global and local sensor subsystems that are independent of traditional wearable sensor systems.

In the preliminary stage of this work, an observational experiments has been conducted in an art museum and observed the behaviors of people with various levels of interest in the exhibits by using the technologies that have been discussed above. By analyzing the observed results, it is found that any observer (e.g. human museum guide) can detect people who may desire the robot's service from their walking trajectory patterns and visual attention information in the museum. Then a network enabled multi robot based HRI system is developed that can find such people to offer them guidance using the guide robot for their exhibit of interest. Finally, the presented HRI system is implemented by incorporating a set of four *Naoko desktop Robots* [1] as museum guide robots and tested the system with an art museum designed in a laboratory environment to confirm its effectiveness in proactively approaching selected people to provide guidance.

4.2 Observational Experiments

To conduct observational experiments in a museum scenario, an art museum room sized 8m×10m was set up in laboratory (Computer Vision Laboratory, Saitama University, Japan) where six paintings were hung. The overall setup of the designed art museum is illustrated in Figure 4.1.

The people at a museum may be there for various purposes with different intentions. Their interests in the exhibits may also vary. The following three cases are considered as the most typical cases as illustrated in Figure 4.2.

Figure 4.1: Setup of the designed art museum including the paintings, LIDAR poles and USB cameras.

•*Case1.* They are interested in the collection in the museum and would like to look at all the paintings carefully.

•*Case2.* They know that a famous painting by some painter is in this museum and would like to specifically appreciate it.

•*Case3.* They are not particularly interested in any of the exhibits. For example, they could have just been in town and decided to visit the museum to pass the time.

Four laser range sensor poles are used in the four corners of the designed art museum room to track people and to obtain *global behavior* information. The details of the tracking method can be found in [152]. This method tracks the locations and orientations of people by using a particle filter framework [90], which is an important requirement for maintaining the continuity of walking trajectories in public spaces. Tracking people's walking trajectories patterns via laser range sensor poles is illustrated in Figure 4.3. Low cost USB cameras were also placed just beneath each of the six paintings. The purpose of using the cameras is to obtain the *local behavior* information of people. In order to observe the above three types of people, 48 participants (42 males, 6 females, average age

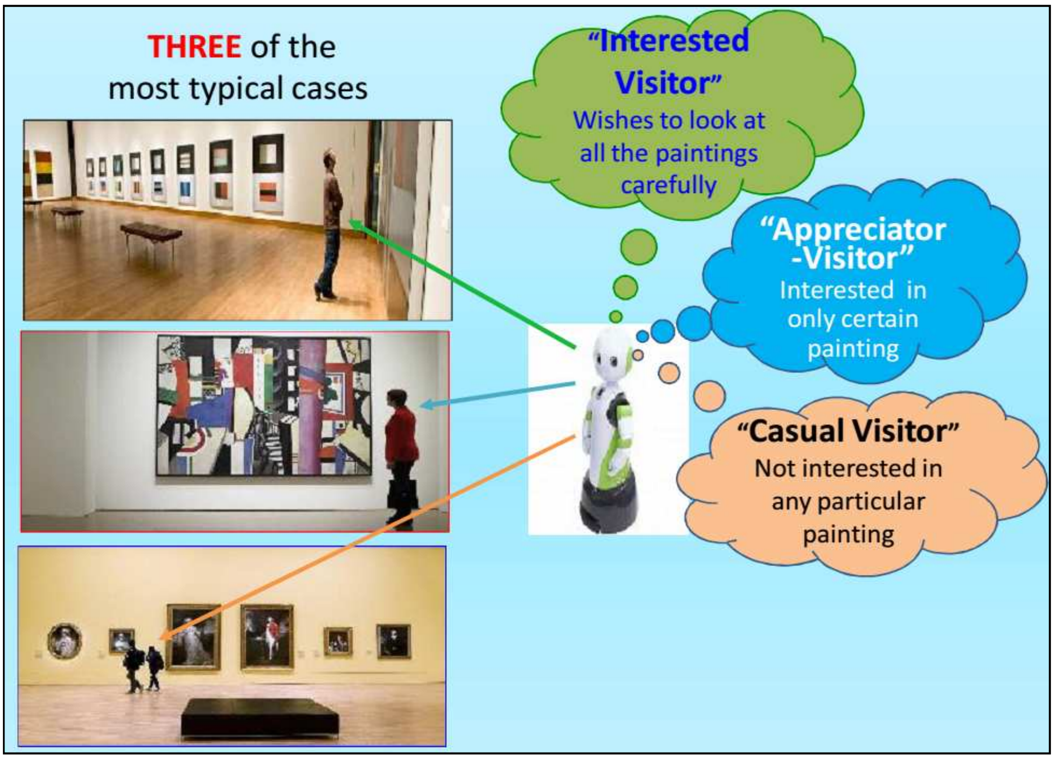

Figure 4.2: Three most typical cases of visiting a museum

25.2 years) from Saitama University were divided into three groups and instructed each group as follows.

•*Group1* (14 participants). Look at all paintings carefully and choose one that you most like. You will be asked about the painting later.

•*Group2* (18 participants). One of the paintings was showed to them and tell them to look at that painting. They were also asked about that painting later.

•*Group3* (16 participants). No specific instruction is given.

4.2.1 Findings of Conducted Observation Experiments

In this study, people's walking trajectory patterns were treated as *global behavior* and visual attention (for example, indicated via face detection) was treated as *local behavior*. In the conducted observational experiments, the following three major trajectories patterns were found.

•*T1*: *"trajectory where all exhibits were viewed sequentially followed by a return back to any one of the previously viewed exhibits"*,

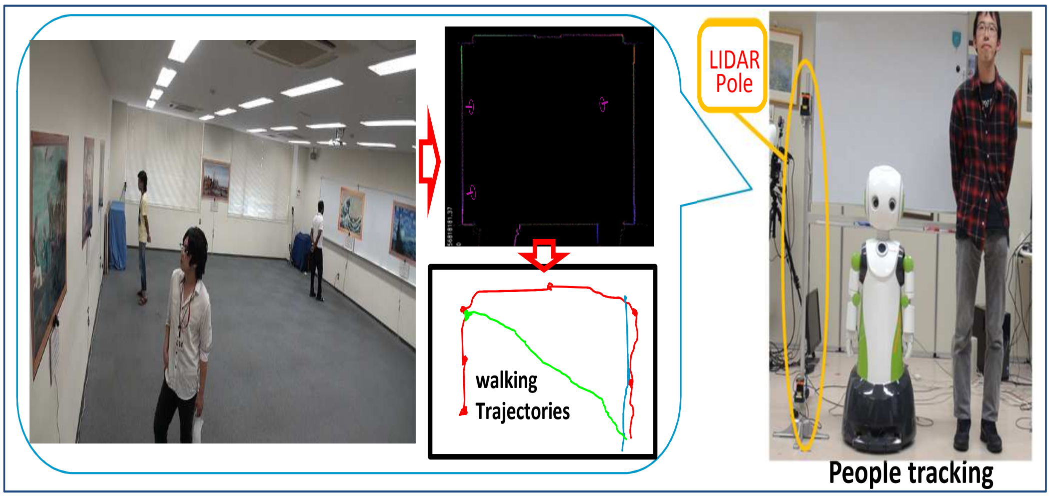

Figure 4.3: People's walking trajectories tracked using LIDAR poles.

Table 4.1: Relationship between different groups of participants and their walking trajectory patterns.

	T1	*T2*	*T3*	*Others*
Group1	09	00	04	01
Group2	00	11	06	01
Group3	01	00	15	00

• *T2*: *"trajectory where the person went straight to view only one exhibit before leaving the museum"*, and

• *T3*: *"trajectory where all exhibits were viewed sequentially"*.

Examples of found walking trajectory patterns of type $T1$, $T2$, and $T3$ are illustrated in Figures 4.4(a), (b), and (c), respectively. Table 4.1 illustrates the relationships between the three groups and the walking trajectory patterns. It is observed in videos that the faces of people viewing a given painting would always be frontal views in the painting specific USB video camera. Figure 4.4(d) shows the examples of visual attention observation of different participants in three different painting specific USB video cameras.

People with walking trajectory pattern $T2$ were found to be interested in a specific painting. Such people are the first candidates for the robot to offer its

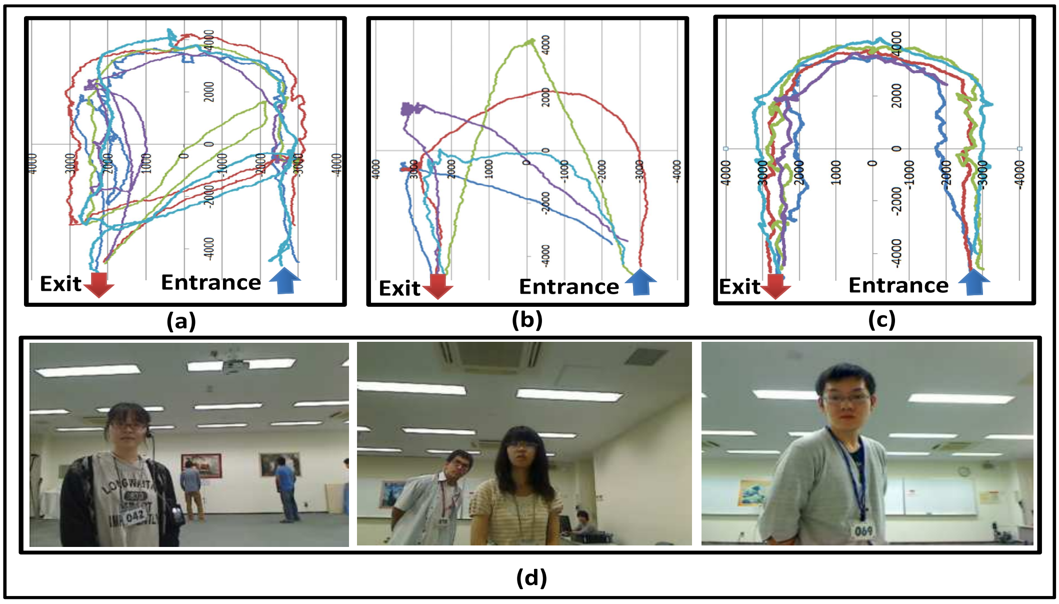

Figure 4.4: Participants' typical walking trajectories for (a) $T1$, (b) $T2$, (c) $T3$, (d) Examples of visual attention observation of participants in three different cameras.

guidance. It is also natural for people to see paintings that they like multiple times. Thus people with walking trajectory pattern $T1$ are assumed to be interested in the painting that they returned to and are also candidates for receiving extra commentary from the robot about the painting.

In addition, after detecting the trajectory patterns of the people, in order to observe their *global behavior*, it can be more sure about their interest towards any specific paintings if their visual attention as *local behavior* near the painting's specific USB video camera were detected. It can be seen from the stored USB video camera footage of the observational experiment that all the video cameras successfully captured the frontal faces of the participants, if and only if they really viewed the paintings for some extended period of time.

4.3 Proposed HRI System

Based on the findings from the observational experiments, a network enabled HRI system is proposed which is illustrated in Figure 4.5. Basically, this HRI system consists of two main types of sub-systems: the *Server Sub-System (SSS)*

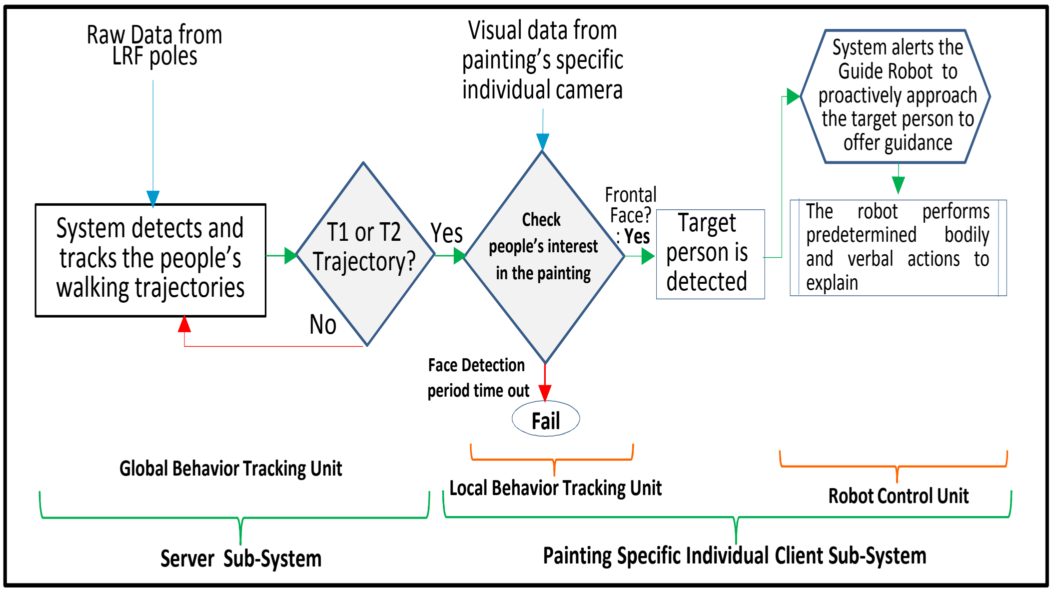

Figure 4.5: States of the proposed network enabled HRI system.

and painting specific individual client system-called *Client Sub-Systems (CSSs)*. A total of three software units was incorporated in these two types of sub-systems :

(a) The human positions and walking trajectory pattern tracking unit-called the *Global Behavior Tracking Unit (GBTU)* in the *SSS*,

(b) Visual attention detection and tracking units-called *Local Behavior Tracking Units (LBTU)*, and

(c) the *Robot Control Unit (RCU)*.

The latter two units are in each individual *CSS*. one *CSS* is employed for each of the paintings. For example, if there are n-exhibits in a museum then it will have one *SSS* and n-*CSSs* to implement the proposed HRI system. Our HRI system detects people's interests and intentions toward the exhibits and guides them accordingly, if and only if the three units in two sub-systems work in order in real time. The communication between the *SSS* and *CSS* units are done by a TCP network connection. The functionality of different units of the proposed HRI system are described in the following subsections.

It is assumed that people are free to move inside the museum to view exhibits based on their own interests and intentions. If any person with *local* and *global behaviors* indicating strong interest in an exhibit is found, the HRI system immediately sends a command to the assigned guide robot to proactively approach him/her to offer its commentary on the exhibit. The potential steps of our proposed HRI system are stated below.

•*Step-1*: The *SSS* tracks the *global behavior* of people using the *GBTU* (described in Section 4.3.1.1).

•*Step-2*: If the *GBTU* detects interest from a person's *global behavior* (e.g. either walking trajectory pattern *T1* or *T2*) in front of any painting viewing region then **goto** *Step-3*, otherwise **goto** *Step-1*.

•*Step-3*: The *SSS* sends commands to the *LBTU* (described in Section 4.3.2.1) of a specific *CSS* to track the *local behavior* of that person.

•*Step-4*: If the *LBTU* detects interest from the person's *local behavior*, then **goto** *Step 5*, otherwise **STOP** and wait to receive future commands from the *SSS*.

•*Step-5*: The *CSS* sends commands to its *RCU* (described in Section 4.3.2.2) to trigger the assigned guide robot to proactively offer extra commentary about the painting to the target person.

•*Step-6*: After finishing its commentary about the paintings, the *CSS* makes the *LBTU* and *RCU* go idle and wait for future commands from the *SSS* if it identifies interest from the *global behavior* of other people (i.e. **goto** *Step 2*).

4.3.1 Server Sub-System (SSS)

4.3.1.1 Global Behavior Tracking Unit (GBTU).

The basic aim in introducing this unit is to detect walking trajectories patterns of type $T1$ and $T2$ as the *global behavior* of the people in the museum. Painting viewing regions[1] with region-ID (rID) numbers for the n-paintings are assigned inside the museum. The schematic representation of the museum's painting viewing regions and the rID assignments for the 6-paintings are illustrated in Figure 4.6. The rIDs are used to construct a trajectory vector (TV) to define each person's walking trajectory pattern where each rID inside the TV represents the person's visited regions inside the museum. If a person views any paintings then the TV will be updated by concatenating the $rIDs$ of all the visited regions of those

[1]The regions from where a person typically views exhibits in a museum

paintings. An example of the TV for the $T1$ walking trajectory pattern can be expressed as:

$$TV_{T1} = [rID_1, rID_2, ..., rID_n, rID_2] \tag{4.1}$$

where it is seen that the person visited region rID_2 a second time to check the painting again.

Thus a person with this TV would be a candidate to offer commentary immediately from the robot about the painting residing inside region rID_2. To check whether a person's position is stable inside any viewing region, the variances of their position coordinate values ($Var(x)$ and $Var(y)$) for every past 10 consecutive frames are calculated and combined using the following equation:

$$D = \sqrt{(Var(x))^2 + (Var(y))^2} \tag{4.2}$$

For any person, if his/her position is inside any viewing region with body orientation toward the painting, and $D < D_{thres}$ for the next 100 consecutive frames, then the robot will assume that the person is stable at that region. Here D_{thres} defines some threshold integer values. Again, if any person goes directly to any region rID to view the painting without viewing other paintings at other regions in the museum, we treat the person as having a $T2$ walking trajectory. However, the conditions of stability in this case require that his/her position is be inside any painting viewing region with body orientation toward the painting, and $D < D_{thres}$ for next 200 consecutive frames. This is because it was observed from experiments that people with $T2$ trajectories spend more time viewing only one particular painting than the average time that people take to view a painting.

Performance evaluation. The performance of the above described method to recognize the $T1$ and $T2$ walking trajectory patterns (*global behavior*) is examined by applying it to the stored data recorded in the observation experiments. The obtained results on recognition of the $T1$ and $T2$ walking trajectory patterns is summarized in Table 4.2. Thus, the proposed method is effective to recognize the walking trajectory patterns, $T1$ and $T2$, thereby recognizing participants that were initially interested in any particular paintings.

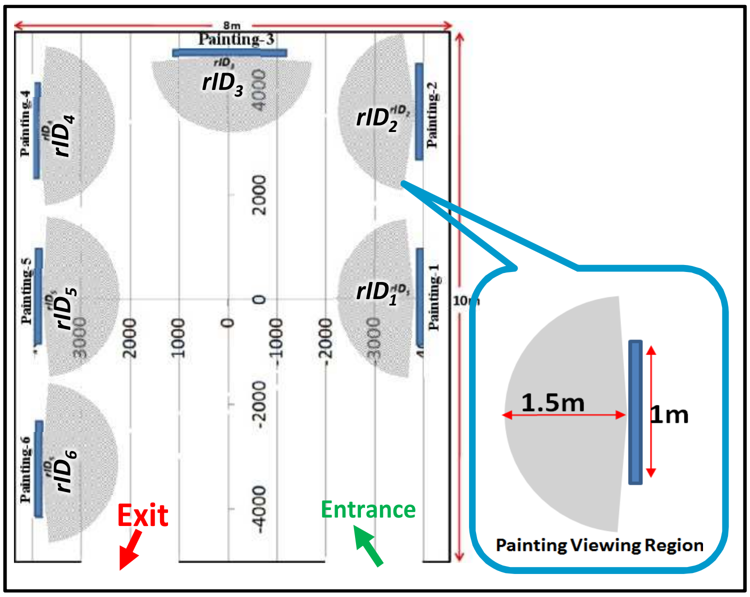

Figure 4.6: Schematic representation of the paintings, painting viewing regions and their rID assignment.

Table 4.2: Walking trajectory patterns recognition accuracy evaluation.

Trajectory Patterns	**Recognition rate[%]**
T1	80
T2	81

4.3.2 Client Sub-System (CSS)

CSS is an important part of the proposed network enabled HRI system. This sub-system does not have any self-functioning capabilities but it's functionalities are fully controlled by *SSS*. Based on the decision of *GBTU*, *SSS* triggers the painting specific *LBTU*.

4.3.2.1 Local Behavior Tracking Unit (LBTU).

To estimate the visual attention (*local behavior*) of the person (whether s/he is looking at the painting or not), in this unit, the Viola-Jones AdaBoost Haar-like frontal face detector [200] is employed to the continuously captured image plane from each USB video camera placed beneath each painting. This unit can detect the person's frontal face, when s/he really looks at the painting (as seen at our observational experiment, described in Section 4.2), thereby estimating that the visual attention of the person (*local behavior*) is high toward any given paintings. In the following paragraphs, a brief description of the Viola-Jones AdaBoost Haar-like cascade classifier for face detection is given. In this study, the face detection time is experimentally set to 5 frames. The face detector was appplied to the stored video footage from the observational experiments and found fruitful detection of frontal faces of the participants. Examples of the participants' visual attention tracking is illustrated in Figure 4.8. From the *local behavior* and *global behavior* information of the people, it can be stated that the proposed HRI system can accurately estimate the interest level of the target people toward the exhibit. For example, when the person's specific walking trajectory pattern is detected in *GBTU* but his/her frontal face is not detected in the *LBTU* by the exhibit specific USB video camera, then the person seems to be less interested to the painting. In such a case, the *CSS* will not alert the *RCU* to offer commentary about the painting to him/her.

Adaboost Haar-like cascaded Classifier: Numerous methods for detecting faces in video image has been proposed. Among them, the AdaBoost-based face detector using Haar-like features has been popular because of its accuracy and robustness against observation with low resolution or varying illumination conditions. The AdaBoost-based classifier consists of linearly connected week classifier. Viola and Jones arranged the classifier in a cascade structure and proposed an efficient computation technique for Haar-like features. Though the training of AdaBoost-based cascaded classifiers requires huge amount of time, the cascaded classifier rapidly detects a face because most of non-face target regions are rejected in an early stage of the cascade. This cascade is effective in the evaluation phase even in the particle filter framework.

In Figure 4.7(a), $\mathbf{H}_i$ represents a strong classifier. Each strong classifier classifies an input image into a positive or negative. Only positive images are used as the input of the next strong classifier. At each stage, a strong classifier is trained to detect almost all face images while rejecting a certain fraction of non-face images. For instance, the classifier at each stage is trained to eliminate 50

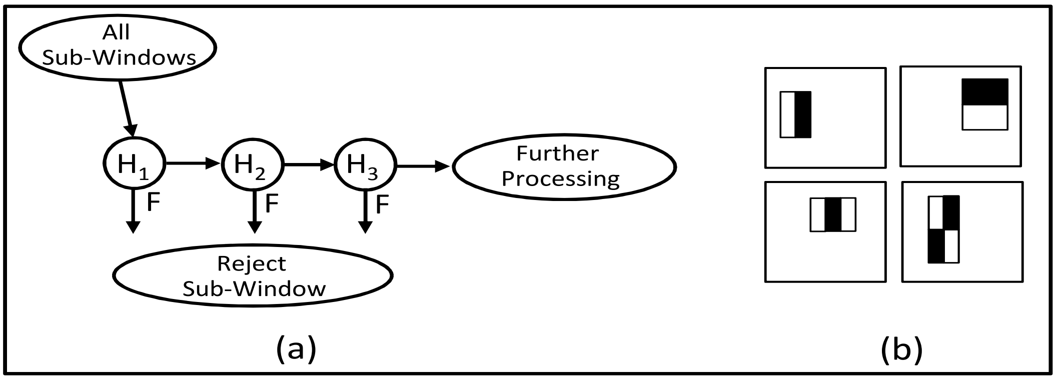

Figure 4.7: Cascaded classifier (Source: [115]): (a) Cascade of classifiers, (b) Example of features

% of the the non-face images while falsely eliminating is only 0.1 % of the face images. After passing 40 stages, we can then expect a false alarm rate about $0.5^{40} \approx 9.1 \times 10^{-13}$ and a hit rate about $0.999^{40} \approx 9.6$. Thus the face detector detects almost all the face images and rejects almost all the non-face images. A strong classifier $\mathbf{H}_i(x)$ at each stage of the cascade consists of many weak classifier $h_i(x)$ (Figure 4.7(b)). This can be described as follows:

$$\mathbf{H}_i(x) = \mathbf{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{4.3}$$

where T is the number of weak classifiers and $\alpha_t = log\frac{1-\varepsilon_t}{\varepsilon_t}$. It is noted that ε_t is an error rate specified in the training phase. Each weak classifier $h_i(x)$ evaluates a target image region by using Haar-like features. The weak classifier performs that the sum of the intensity of pixels located within the black rectangles is subtracted from the sum of the intensity of pixels located within the white rectangles. The AdaBoost algorithm selects efficient features to classify the target image region among huge variety of features.

4.3.2.2 Robot Control Unit (RCU).

After successfully detecting a target person using the *GBTU* and *LBTU*, our HRI system immediately sends signals to the *RCU* of the corresponding *CSS*. Then the *RCU* triggers the guide robot to proactively perform predetermined verbal and gestural actions to attract the attention of the target person. After finishing

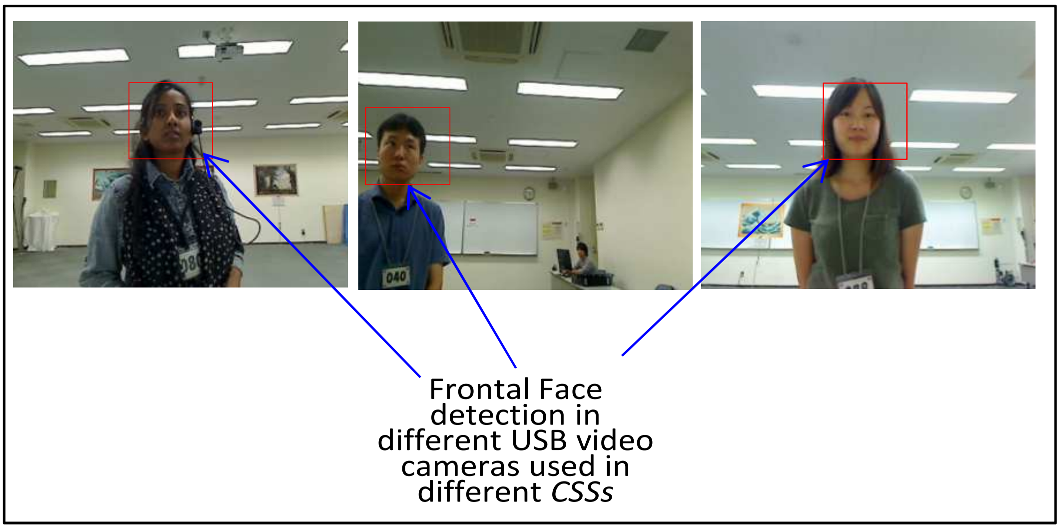

Figure 4.8: Examples of Visual Attention tracking.

its explanation about its assigned painting, the guide robot returns back to its idle state to receive future alert signals from the *RCU* to offer explanations about the painting for future interested people.

Naoko Desktop Robot. The *Naoko Desktop Robot* (see Figure 4.9) is employed as the type of guide robot for every *CSS*. Like, *Robovie-R3* (explained in Chapter 3) it does not have any built-in artificial intelligence. But, in the implementation of the proposed HRI system, *Naoko Desktop Robot* is programmed to perform various kinds of functions for holding conversation in a fixed location. After proposed programming, each robots were placed on a desk beside each of the paintings so it could proactively approach target people that are interested in particular paintings. The key properties of *Naoko Desktop Robot* is illustrated in Table 4.3. If the attention level of the person is detected as high in the *GBTU* and *LBTU* towards any specific painting, then the guide robot exhibits its verbal and gestural (head shaking, hand shaking and body base movement in between the target person and painting) actions to draw attention and offer commentary about the paintings.

Figure 4.9: *Naoko Desktop Robot.*

4.4 Experiments

In order to confirm the effectiveness and accuracy of the proposed HRI system, experiments were conducted by implementing it using four painting specific *Naoko-desktop robots* as guide robots for four paintings at the self-designed art museum in the laboratory where observational experiments were previously conducted. A schematic representation of the experiment setup of the proposed HRI system is illustrated in Figure 4.10. A panoramic view of the considered museum scenario

Table 4.3: Properties of *Naoko Desktop Robot.*

Parameter	**Specification**
Machine Body	KHR-1/2/3HV Normal.
Size	Height: about 40cm
Weight	1.5/ 1.6/ 1.8/ 2.0/3.0 Kg
Degree of Freedom	7 (Arms:4, Head: 2 (pitch and yaw), Body base:1)
Servomotor	KRS-788HV16
Battery	Li-Po 11.1V 1350mAh
Production Period	2006∼2010

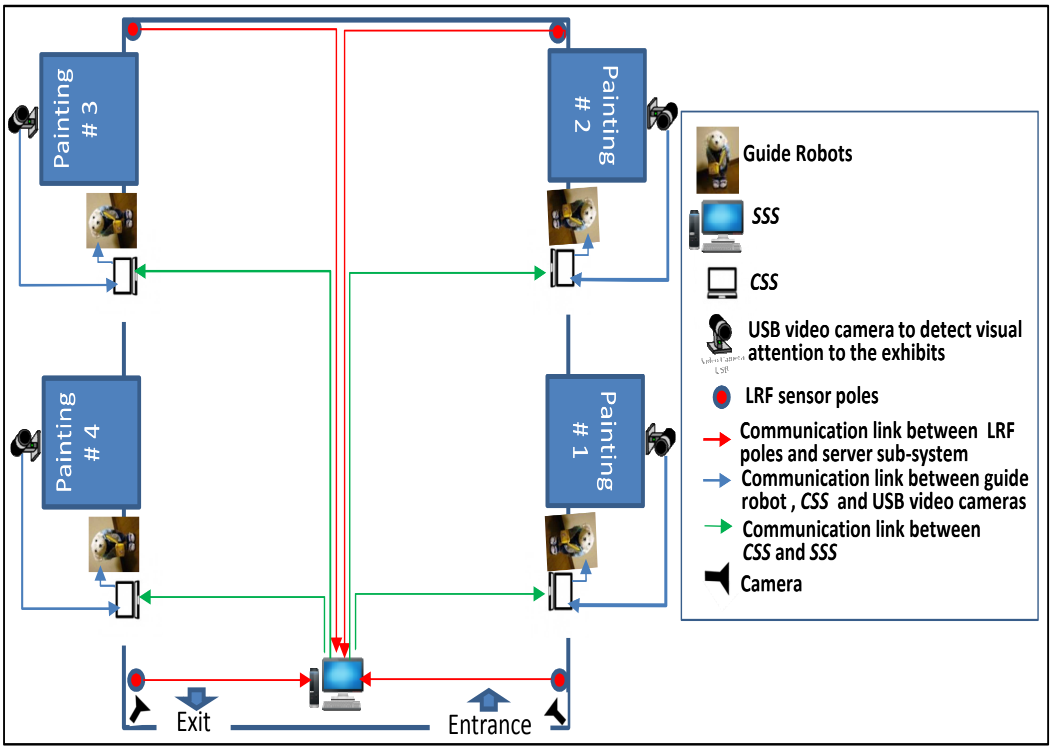

Figure 4.10: Experimental setup of the proposed HRI system.

under the proposed HRI system is illustrated in Figure 4.11.

The experiments were conducted where people moved inside the museum to view the paintings, while the system estimated their *local* and *global behavior* to determine their intention and interest in the paintings. If any desired *local* and *global behaviors* of the people are found then one of the network enabled guide robots will proactively offer guidance to him/her.

4.4.1 Demonstration using Guide Robots.

Demonstrative experiments were conducted with attendees from Saitama university, Japan. In the demonstration, the attendees were asked to visit that museum under the proposed HRI system according to the instructions from the observational experiments (stated in Section 4.2). Two video cameras were placed at appropriate positions inside the museum to capture all the activities of the attendee and the guide robots during the demonstration. Example scenes are shown in Figure 4.12(a) and Figure 4.13(a). In both figures, the lower rows show

Figure 4.11: Panoramic view of the considered museum scenario under the proposed HRI system.

the tracking results corresponding to the scenes in the upper row where each scene defines the potential steps of a typical demonstration with an attendee. Figure 4.12(b) and Figure 4.13(b) show the *local* and *global behavior* tracking results of two typical cases under the proposed HRI system.

4.4.1.1 Case-1.

Consider Figure 4.12(a) where an attendee came inside the museum and briefly viewed painting #1 (scene #1) then moved to the next painting #2 (scene #2). In this case, he briefly viewed all the other remaining paintings (i.e. painting #3, and #4, as shown in scene #3 and #4, respectively). It is seen from the recorded video that he viewed every painting for a very short time. After that, it is also seen in scene #5 that the attendee moved again to view painting #2, and viewed that painting carefully with much more time than earlier. In such a case, the proposed system detected the type of walking trajectory pattern (*global behavior*) of that attendee as *T1*, which is shown in the left side of Figure 4.12(b). Accordingly, his visual attention (frontal face) toward that painting was detected as his *local behavior* which is shown in the middle of Figure 4.12(b). Finally, the guide robot assigned to painting #2 approached him proactively offering more explanation about that painting by saying "*Excuse me, may I explain more details about the painting to you? ......Thank You! This painting is a famous painting.....*". A

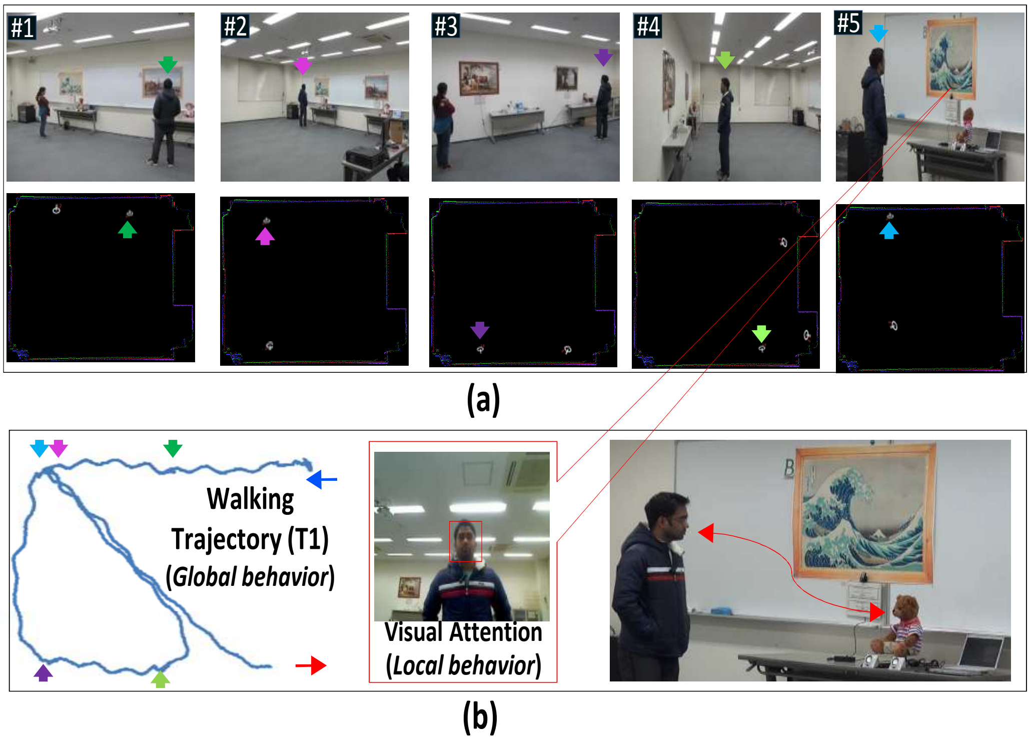

Figure 4.12: (a) Example scenes of demonstrative experiments: ***Case-1***, (b) Example of *global* and *local behavior* tracking result and session of a guide robot and attendee during demonstrative experiments: ***Case-1***.

snapshot of the session with the guide robot and the attendee is depicted in the right side of Figure 4.12(b).

4.4.1.2 Case-2.

In Figure 4.13(a), it is seen that an attendee just came inside the museum and did not go to the painting viewing regions of all the paintings (scene #1, and scene #2) but went to the painting viewing region of painting #3 (scene #3) and stayed there for while to view that painting carefully. After a few moments, the walking trajectory pattern (*global behavior*) of that attendee was detected as *T2* and this is shown in the left side of Figure 4.13(b). With a very short delay, his visual attention (frontal face) toward that painting was detected as his *local behavior* which is shown in the middle of Figure 4.13(b). Finally, the guide robot assigned for painting #2 approached him proactively offering extra commentary about that

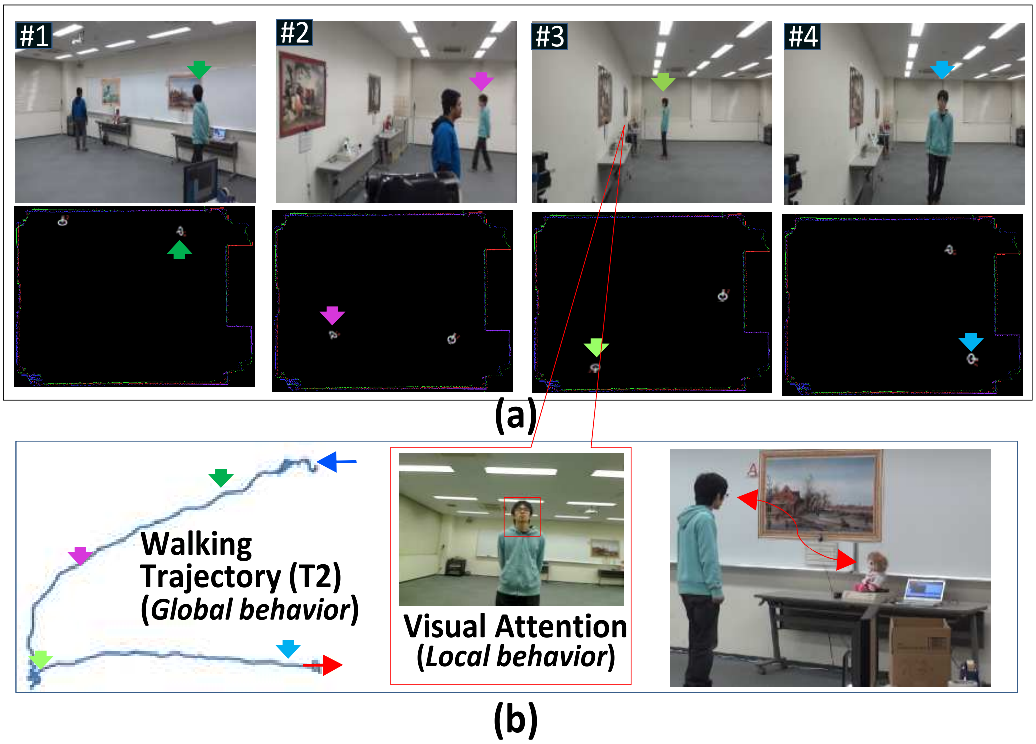

Figure 4.13: (a) Example scenes of demonstrative experiments: ***Case-2***, (b) Illustration of *global* and *local behavior* tracking results and a commentary session of between the guide robot and the attendee during demonstrative experiments: ***Case-2***.

painting by saying "*Hello, may I explain to you more details about the painting? ....*". A snapshot of the commentary session between the guide robot and the attendee is shown in the right side of Figure 4.13(b). After listening to the guide robot's commentary, the attendee left the museum room without making additional stops at other paintings (scene #4).

Conclusively, it can be said that the proposed HRI system is effective for detecting the interests and intentions of people to paintings in museums from their *local* and *global behaviors* using the local and global sensor systems and is able to offer proactive guidance to them accordingly using the guide robots in the network robot system.

4.5 Chapter Summary

In this chapter, a Human Robot Interaction(HRI) system is presented which can determine people's interest and intentions concerning exhibits in a museum, then proactively approach people that may want guidance or commentary about the exhibits. To do that, firstly an observational experiments has been conducted in a museum with participants. From these experiments, mainly three kinds of walking trajectory patterns has been found which characterize *global behavior*, and visual attentional information that indicates the *local behavior* of the participants. These behaviors ultimately indicate whether certain participants are interested in the exhibits and could benefit from the robot system providing additional details about the exhibits. Based on the findings, a network enabled guide robot system has been designed and implemented for the museum. Finally, the proposed HRI system has been demonstrated by experimenting with a set of *Desktop Robots* as guide robots to proof its viability. the experiments revealed that the proposed HRI system is effective for the network enabled *Desktop Robots* to proactively provide guidance.

4.5.1 Limitations

The proposed HRI system is implemented, tested, and demonstrated using participants inside a designed art gallery under laboratory controlled environments. But in real art galleries, it would be difficult to observe people's interests, intentions, and preferences towards the exhibits using the network enabled sensor system because the used *global behavior* tracking system is not robust enough to track people's positions and walking trajectory patterns in real art gallery like popular public spaces. The main drawback of the LIDAR based *global behavior* tracking system is that it is very sensitive to occlusion which will frequently happen in crowded public spaces. Thus, the next chapter focuses on developing a robust human tracking method for service robots to identify humans and track their positions, body orientation, and movement trajectories in crowded public spaces to read their various types of behavioral responses to surroundings.

Chapter 5

Robustly Tracking People with LIDARs in a Crowded Museum for Behavioral Analysis

5.1 Introduction

Nowadays, observing and understanding people's behavior is highly valuable within real social environments, such as shopping malls, hospitals, train stations, schools, and museums. To do so, we can sense and track people's various bodily actions. Commonly considered bodily actions include position, body orientation, walking trajectory patterns, head and/or gaze movements, facial expressions, and so on. Such types of behavioral information can be used for estimating their attention levels and intentions, and for extracting knowledge on their actual expectations from environments. For example, it is very important for museum curators and museum professionals (MPs) to observe and understand the visitors' various attention levels and intentions in a given museum gallery so that they can tell which exhibits visitors consider more attractive. Such types of information may help them to make a museum gallery more attractive by rearranging the exhibits in more appropriate ways and/or adapting various services for the visitors. However, making manual large scale observations of human behaviors using only MPs is a very difficult task.

To meet the demands, together with the growing acceptance of modern technology in our daily social environments, sensing technologies can play a crucial role to extracting such valuable information. So far, there has been much research on sensing technologies that are employed to track people in the fields of robotics and computer vision [26, 136], which can be used to extract knowledge on their behavior and social connections. In ubiquitous computing, positioning devices are often used. These include the use of GPS, or the signal strength of radios (GSM, WiFi, Bluetooth, RFID) [103]. These technologies all used wearable or mobile personal devices, but these approaches have a number of weaknesses for applications in large-scale social environments. For example, in the context of public social spaces, people may enter the space spontaneously, usually pass time based on their own interests, and may not be interested in actively engaging with the technology. Thus, in this chapter, we are interested in making use of people tracking for a wearable-free solution where people do not need to attach markers to themselves or carry special devices so that we may observe them in an unrestricted manner.

In the context of previous trackers, a new system is introduced in this chapter that is more robust than previous work such as [152] while being cheaper and easier to setup than [26]. As a result, it will be easy to deploy the presented tracking system in real crowded environments for studying human behavioral patterns. LIDAR is chosen as the primary remote sensing technology in this presented system, which can be distributed in public social spaces. Because of the simplicity and unobtrusiveness of the LIDAR's installations in any public space, the proposed sensing system does not detract from interests and intentions toward objects in the social space. In this study, my claims is tested by taking a real art gallery (at the Ohara Museum of Art, Kurashiki, Japan, see Figure 5.1) as an example of a public space and using the sensing system to track people's position, body orientation, and walking trajectory patterns to obtain their overall behavioral information. Such information could assist the MPs in introducing special services (for example, a museum guide robot [173]) to support future patrons.

5.1.1 Importance of Tracking Museum visitors

Museums are more important public spaces than ever before [54] where traditionally, MPs collect demographic data to understand their visitors [191]. This

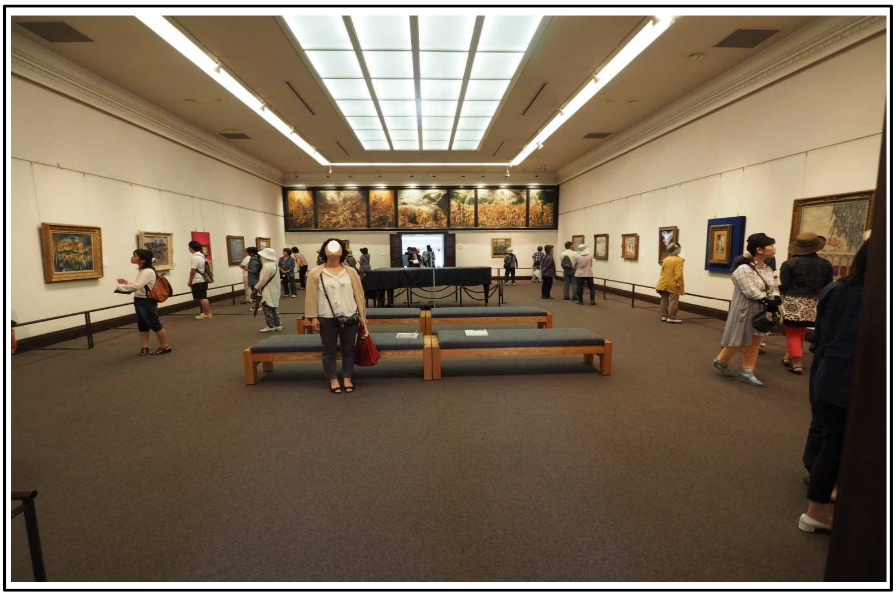

Figure 5.1: Art Gallery at Ohara Museum of Art, Kurashiki, Japan.

data typically includes information such as race, age, level of education and so on. While this data is easy for visitors to provide, it has been shown through years of visitor research that it is not predictive of visitor behavior or experience [195]. But, the importance of understanding how visitor behaviors in museums has been recognized for many years [184] because their behaviors reflect their internal states (e.g., interests and preferences toward exhibits). How visitors circulate through museums determines what visitors will see, where they focus their attention, and, ultimately, what they learn and experience [185]. Such types of information may help the MPs to make a museum more attractive by rearranging the exhibits in more appropriate ways or adapting various services (e.g. museum guide robot services) for the visitors. To achieve such objectives, while collecting the visitors' demographic data, MPs can sense and track visitors' behaviors in terms of their various bodily actions and activities. However, making manual large scale observations of human behaviors using only MPs is a very difficult and complicated task.

In recent years there have been a variety of initiatives involving the introduction of emerging technologies in museums [208]. Most of these introduced technologies (for example, multimedia technologies[48, 125, 155] and robotic technologies [29, 30, 175, 193, 208]) open the opportunities of various presentations of the exhibits to the visitors, which ultimately enhance their experiences in museums. But, most of these technologies are mostly not related to identifying the visitors' overall behaviors about the exhibits. But, to meet the current demands of MPs, modern sensing technologies can play a crucial role in supporting MPs in extracting the behavioral information of the visitors inside the museum.

Thus, it is important to develop a human-tracking system for museum with networks of relatively low cost LIDARs as opposed to a 3D range sensor platform to include better spatial coverage, robustness, modularity, and overcome the occlusion constraints. Due to the current limitations (stated in Section 5.2) of the system presented in [152], in this chapter, the system is extended using LIDAR to track a relatively large number of people even in occluded situations in large scale social environments. The sensing accuracy in the proposed approach can be improved by overlapping the fields of view of widely separated sensors deployed in real-world environments. The ultimate goal of the research is that users with access to the information provided by the set of LIDARs capable of detecting and tracking people in the public spaces, will allow for classifying behaviors and ultimately estimating their degree of interest towards their surroundings. It may also help in adapting new services for people in the near future.

5.2 Drawbacks of a Human Tracking Method

The human tracking method presented in [152] can only work reliably in laboratory environments where there is a small number of people and there are line-of-sight views from the LIDARs to the humans. Furthermore, it can not robustly handle partial occlusions and fails to handle full occlusions which occur frequently in between the LIDARs and the humans in crowded large-scale environments. Here, the basic reasons for those drawbacks are briefly discussed.

The previous system can effectively track humans with unique-IDs when all the LIDARs are accurately calibrated on the poles deployed in the observed environment, and if there are no occlusions in between the target human and all the deployed LIDARs, even when humans are moving about (see Figure 5.2(a)). But, due to the lack of accurate calibrations among all of the LIDARs as well as sensors' internal operational processing delays in time, the partially observed body outlines from all the LIDARs do not always fit to the contour observation model in the image mapped from the distance data (see Figure 5.2(b)). This results in inaccurate computation of weights of the evaluation points of the target person in the image mapped from the LIDARs and thereby degrades the performance of human body position and orientation measurement.

Again, in the case of crowded environments, it is not always possible to maintain line-of-sight views from all the LIDARs to the target people. Hence, partial occlusion may occur frequently between any of the LIDARs and the target person. So, it is quite difficult to assess the contour similarity between the observation model and the contour of the human body observed from all the deployed LIDARs due to insufficient observable evaluation points (see Figure 5.2(c)). Thus, the weight computation of the very small number of evaluation points of a given person is not enough to compute the expectation value across the target person over time. For this reason, this system can not track a partially occluded person's position and body orientation accurately with his/her unique-ID.

In addition, this system can not deal with situations where reassigning the same unique-ID to the persons that were temporarily lost by the tracker due to full occlusion from the LIDARs. Thus, system is not robust enough to track humans with their proper identity in crowded public spaces. For these reasons, in this chapter, this method is extended to make it usable not only in laboratory environments but also in crowded large-scale social environments to robustly track human position, body orientation, and walking trajectories even with partial and/or full occlusion, to thereby obtain their behaviors. Details of the extended system are presented in the next section.

5.3 Extended Human Tracking System: Proposed Approach

To estimate the location and body orientation of people in public spaces, we employ a system using a set of poles where each pole is equipped with a LIDAR, installed at the shoulder level of a typical adult (Figure 5.3(a)). A LIDAR can measure the distance to the person on a horizontal plane and then map the distance data on to the 2D image plane (what we call a "laser image"); shown in

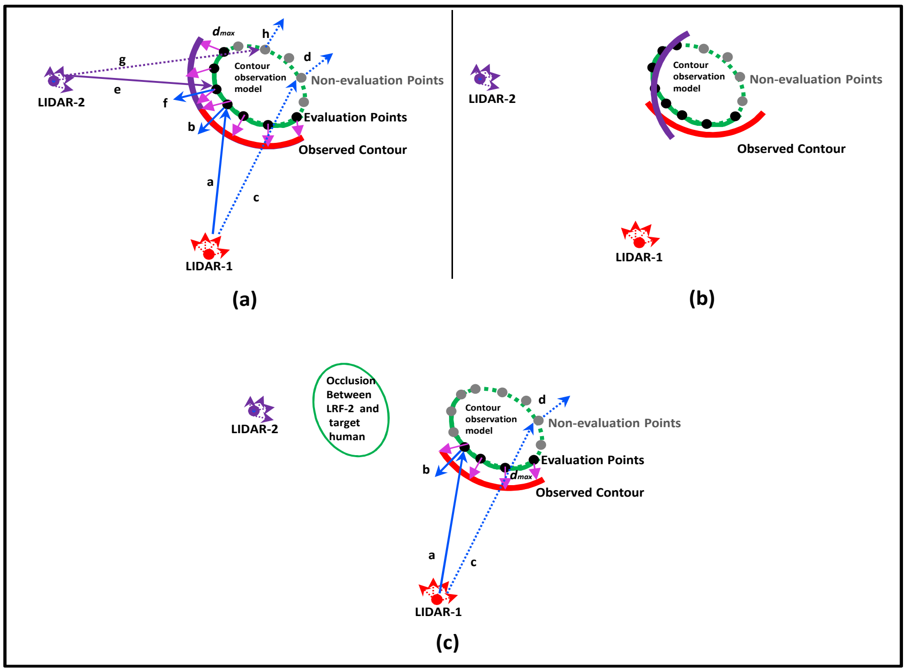

Figure 5.2: (a) Ideal Case: Evaluation model formed by fitting an ellipse to the shoulder outline obtained by LIDAR-1 and LIDAR-2, (b) Defective observed body outline by LIDAR-1 and LRF-2 as compared to the ideal case. (c) Fitted evaluation model is quite different from the shoulder outline obtained only from LIDAR-1, without LIDAR-2 due to occlusion.

Figure 5.3(b)(left). The outline shape of a human's shoulder can then be observed as shown in Figure 5.3(b)(right). This outline portion of the human's shoulders can be considered as a part of an ellipse. Thus, for likelihood evaluation of the people's observed samples, an ellipse is used for the tracking model (Figure 5.4(a)). The system tracks the locations and body orientations of visitors by using a particle filter framework [90]. A coordinate system is assumed with its 2D axes [X,Y] aligned on the ground plane. The center coordinates of the ellipse (x,y) and rotation of the ellipse θ are considered to represent the model of the visitor tracker. These parameters are estimated in each frame by the particle filter. Using regular particle filters, posterior distribution is represented by a set of weighted samples that are propagated by a motion model and evaluation by an observation model. A simple random walk model for state propagation is employed and samples are evaluated based on the observations of the LIDAR. The

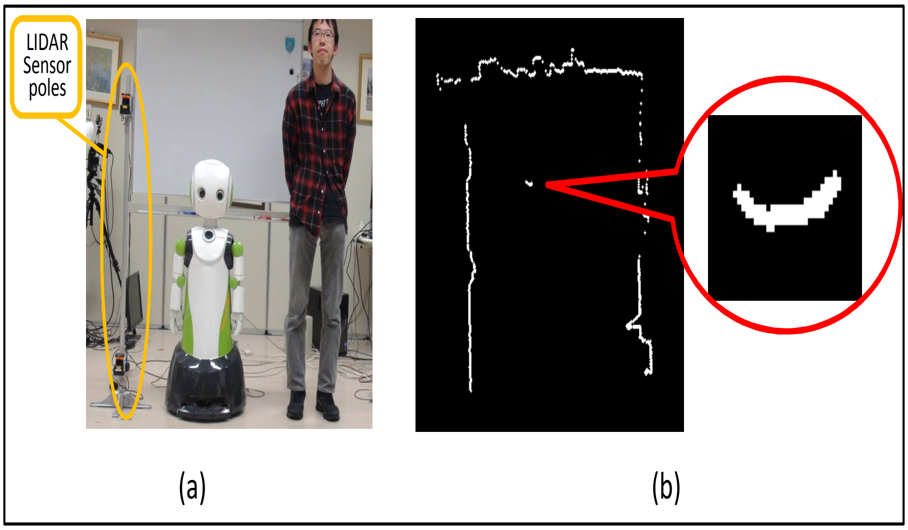

Figure 5.3: (a) The Sensor Pole consists of a LIDAR, (b) Distance-mapped image generated by the LIDAR.

likelihood evaluation model is discussed in the next section.

5.3.1 Likelihood Computing Model

To compute the likelihoods of the samples from the raw sensor data of the deployed LIDARs, proposed system incorporates two layers of computations: the *fundamental* and *integration* layers.

In the *fundamental* layer, for each of the deployed LIDARs, the weights of the samples are evaluated individually from assessing the contour similarity between the model (shown in Figure 5.4(b)) and the body outline partially observed by the individual LIDARs. To do so, data from individual LIDARs are considered as providing qualitative information which are useful for estimating human positions in the observation area under the following three scenarios: (a) *most probably exist*-indicating a contour which may correspond with the edge of a detected human and/or object, (b) *must not exist*-indicating certain points are empty, that is there is no any observable entities, (c) *undefined*-indicating that a certain area is occupied by some observable entity. In such an area, the existence of any person and/or objects is unclear. Figure 5.5 (a) and (b) illustrates the distinction between these three types of scenarios, obtained from the LIDARs provided information.

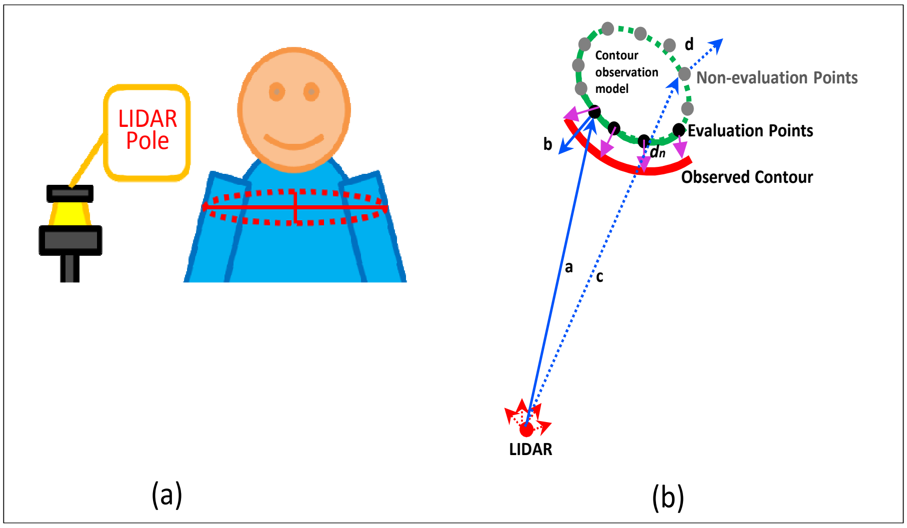

Figure 5.4: (a) The shoulder outline can be modeled as an ellipse, (b) Evaluation model formed by fitting an ellipse to the shoulder outline obtained by the laser range finder.

Now, it is important to design a weighting scheme for each of the samples. The design of our weighting function is illustrated in Figure 5.6 for the three mentioned scenarios. The weights are assigned as C_{HIGH}, C_{LOW}, and C_{MEDIUM} to the data provided by the LIDARs for the scenarios-*most probably exist*, *must not exist* and *undefined*, respectively. With these assigned weights, it can be easily identified the state of the observation area on whether any object/person exists or not.

Specifically, the *i-th* sample is generated at time t for each contour obtained from the image mapped from the distance data. The normal vectors of each point (the blue points in Figure 5.4(b)), such as 'b' and 'd', are assumed as shown in Figure 5.4(b). The vectors from the position of the laser range finder to the points are assumed to be 'a' and 'c'. The system then calculates the inner product of the vectors for each point. If the result of the inner product is negative ($a \cdot b < 0$), the point is able to be observed by the LIDAR. These observable points are dealt with as evaluation points (the deep black points in Figure 5.4(b)). Conversely, a positive inner product ($c \cdot d > 0$) indicates that the point is not able to be observed by the LIDAR. Next, the distance, d_n between each evaluation point and the observed contour is calculated. Then, the weights of the *i-th* sample at

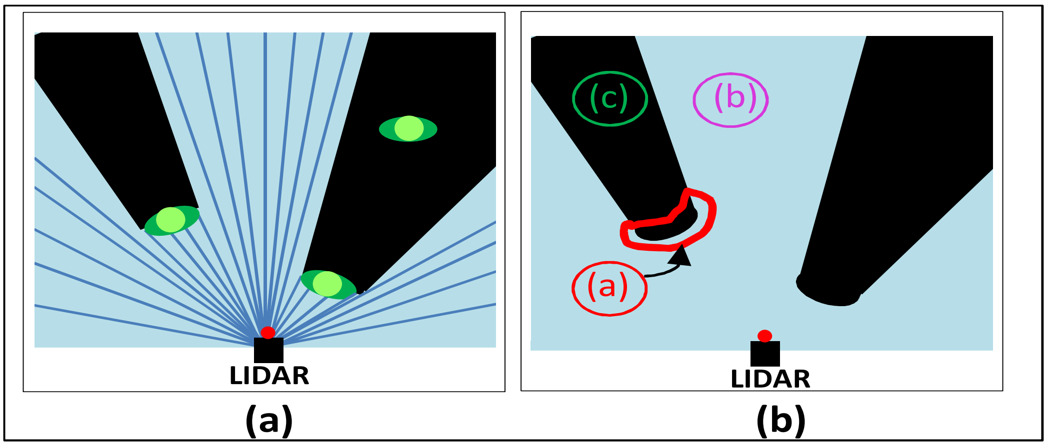

Figure 5.5: A typical LIDAR to scan for people. (a) The position of people relative to the LIDAR, (b) three types of scenarios.

time t for a single LIDAR is calculated by:

$$w_t^{l,i} = \prod_n V_n \tag{5.1}$$

Here, V_n=$f(d_n)$, where d_n is the distance from the LIDAR sensor to the samples of evaluation candidates, l is the LIDAR ID.

More specifically, V_n can be expressed as follows.

$$V_n = f(d_n) = \begin{cases} C_{\text{LOW}}, & \text{if } d_n < T_1 \\ C_{\text{HIGH}} \cdot \exp(-d_n^2), & \text{if } T_1 \leqq d_n \leqq T_2 \\ C_{\text{MEDIUM}}, & \text{if } T_2 \leqq d_n \end{cases} \tag{5.2}$$

Now in the *integration* layer, simply added those computed weights of individually observed samples from different LIDARs to get the aggregated weights for each of the samples. Thus, the aggregated weights of the *i-th* sample at time t from all the LIDARs can be obtained by the following simple calculation:

$$W_t^{\,i} = \sum_{l=0}^{n\text{-}1} w_t^{i,l} \tag{5.3}$$

By calculating the weight values across the samples, the system can estimate and track each person's position accurately with a unique-ID. The higher the

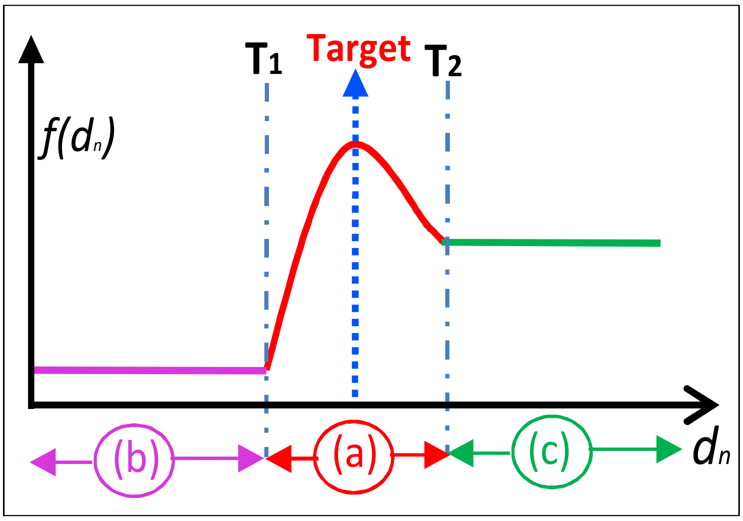

Figure 5.6: Representation of the weight function for the three observation scenarios.

aggregated weight value for a sample, the greater the tracker's stability to robustly detect and track the position and body orientation of the person.

Since, in this extended system, the computations of the weights of the observed sample are performed from each LIDAR individually in the *fundamental* layer, and in the *integration* layer, the LIDARs specific computed weights are integrated to finally calculate the expectation value across the samples to estimate each person's position and orientation. Thus, it does not need careful calibrations among the LIDARs, which was necessary in the system presented in [152] to fit the combined observed body outline from all the LIDARs to the contour observation model. In addition, proposed system can track humans even if partially occluded in a crowded large scale environment. This is because, system can track a human as long as s/he is detected by at least one LIDAR. In [152], detection and tracking was based on the observations of all the deployed LIDARs.

5.3.2 Reassigning Unique-ID to a Temporarily Lost Person

If the proposed human tracking system can not estimate the location of a person from all the LIDARs due to temporary full occlusion by the surroundings person, then the weight computation will not be applicable in maintaining the tracking

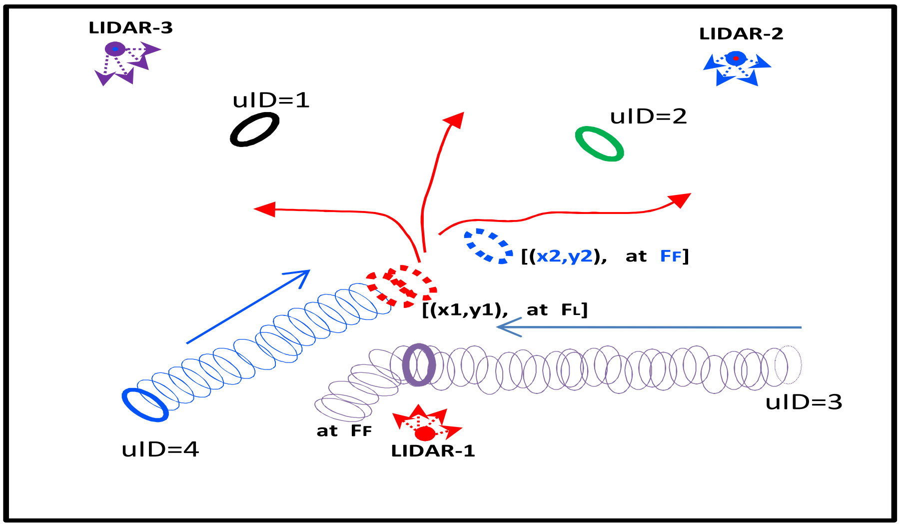

Figure 5.7: Illustration of full occlusion scenario.

of him/her. Thus, the likelihood would be below the threshold. In such a case, his/her tracker is immediately removed because the particle filter is assumed to no longer be tracking anything.

For example, in Figure 5.7, there are three LIDAR poles in the observation area to track people with their assigned unique-ID, uID=1,2,3, and 4 by the proposed system. At frame $\#F_L$, person with unique-ID uID=4 is fully in occlusion from all the LIDAR poles by the persons with unique-IDs 1,2, and 3. The the tracker of that person failed to track. In the proposed system, such situations are overcome with the following simple but effective strategy.

Once a person is fully under occlusion or not visible to the all LIDARs of the sensor system for a while at some location in the observation area, then the resulting likelihoods tend to zero by which the tracker for him/her will be removed by the system but the following valuable information will be preserved for a later attempt to reassign to the lost person:

- Lost unique-ID,
- Coordinate position where tracker failed, (x_1,y_1)
- Frame no. (when tracker failed), F_L

After few frames later (for example, at frame $\#F_F$ in Figure 5.7), if the fully occluded person becomes visible to any of the LIDARs of our sensor system, then the system will try to assign the previous tracker's unique-ID to that person if

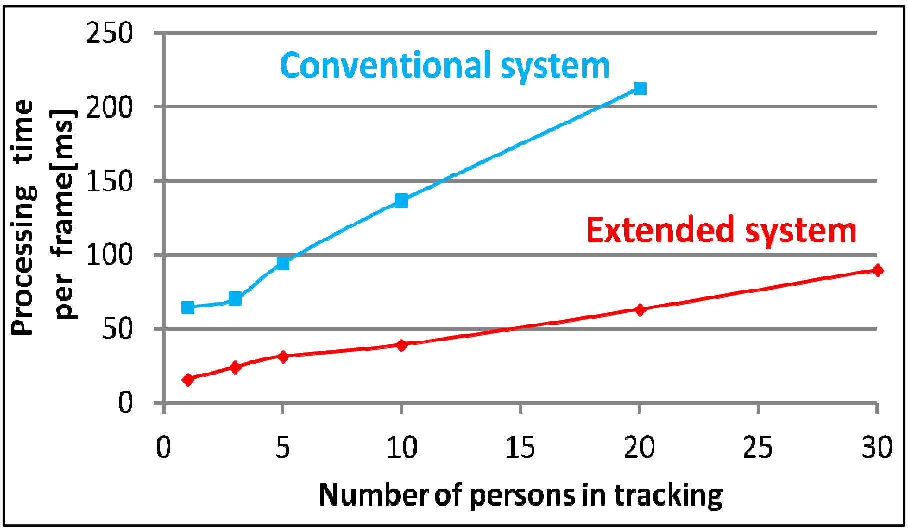

Figure 5.8: The processing time per frame compared to the number of persons being tracked. The blue and red lines indicate the time needed for the CPU and GPU respectively (Source: [98]).

meet the following two conditions are satisfied:

- *Condition-1.* The distance, **D** between the lost coordinate (x_1,y_1) and the coordinate of the newly visible position of the lost person (x_2,y_2) is less than some threshold distance, D_{th}.
- *Condition-2.* The frame difference, F_D between the frame where tracking failed, F_L and the frame where the person appeared, F_A is some reasonable amount,F_{th}.

Thus, it can be said that if **D** $<D_{th}$ and $F_D(=F_A\text{-}F_L)<F_{th}$ are satisfied for a person then the system will reassign the same unique-ID to the person who just lost that unique-ID previously due to full occlusion for a while.

In this presented system, parallel computation of each evaluation point in each sample can be performed by using CUDA (Compute Unified Device Architecture) [129]. Consequently, the system's performance will not be adversely affected by the number of visitors being tracked even in large-scale environments. Figure 5.8 shows the processing time of each frame as the number of tracking targets increases. As can be seen, the processing time is not significantly increased as the number of visitors increase. Additionally, the system can track multiple persons robustly in real time even with full occlusions for a while.

5.4 Art Gallery Installation

In the following the experimental setup and results of the human tracking system implementation in a art gallery of the "Ohara Museum of Art" in the Kurashiki area of Okayama, Japan is presented. First the system design for the real world environment is demonstrated and analyzed the obtained performance in terms of tracking accuracy. Second, an example is provided where the achieved large amount of tracking data can provide useful knowledge about the visitors to the MPs.

5.4.1 Tracking System Setup

Figure 5.9(a) illustrates location map of the art gallery (the red dashed line with size 20m $\times$ 11m) in the museum. That art gallery is exhibited with 21 masterpieces of the western arts from the 19$^{\text{th}}$ and 20$^{\text{th}}$ centuries. Among those, 19 paintings are hanged on the wall and two remaining large sized paintings are hanged over the doors in the north and south sides of the art gallery, see Figure 5.9(b). There are also four sofas and one piano permanently in the middle of the art gallery; see Figure 5.1. Additionally, for the visitors, there is a charming window view of the outside through the door in the north side of the art gallery. All the valuable paintings and piano are permanently protected against physical contact by the visitors using a physical barrier. Usually, in that art gallery, visitors view the paintings according to different patterns of behaviors. For example, some of the visitors view all the paintings and exit the art gallery, some visitors sit on the sofa to rest in between viewing the paintings, a few visitors go through the door on the north side of the art gallery to look out the charming window view of the outside before leaving the art gallery.

The partial goal of the art gallery installation was to make the sensor arrangement inside the art gallery as simple and visitor friendly as possible so that the sensors would be inconspicuous and not detract from the visitor's actual interests towards the painting inside the art gallery. To do so, the tracking system is realized as a combination of six LIDARs which are set up at the standard shoulder level of an adult human inside the art gallery; see Figure 5.10. Each of the LIDAR poles is equipped with a laser range finder (Model: UTM-30LX/LN and UTM-30LX-EW). This setup is well suited to installation at a real large-scale

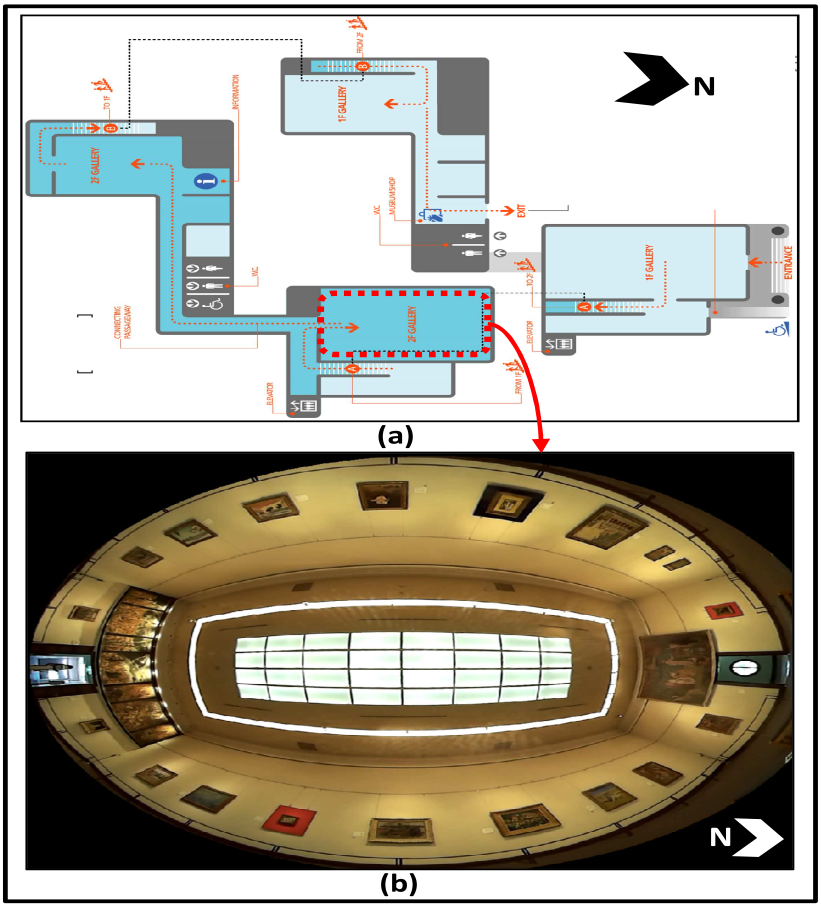

Figure 5.9: Visitor Tracking area in an art gallery of Ohara Museum of Art, Japan; (a) The red dashed line on the map shows the border of the area covered by the sensors; (b) Upward global view of the entire art gallery from a tripod mounted SP360 action camera.

environment such as an art gallery in a museum. This is because deployed sensors have a relatively large usable range: the maximum range at which correct and stable measurements can be obtained is around 0.1 meter to 30 meters [2]. The measurements are accurate with low noise, especially for close ranges. The number of missing measurements increases with distance, especially for dark and transparent objects. To overcome such types of technical constraints as well as to compensate the tracking error due to occlusion among visitors, in considered wide-scale art gallery six LIDARs are used. This setup ultimately assists the human tracking system to robustly track most of the visitors with their unique-ID

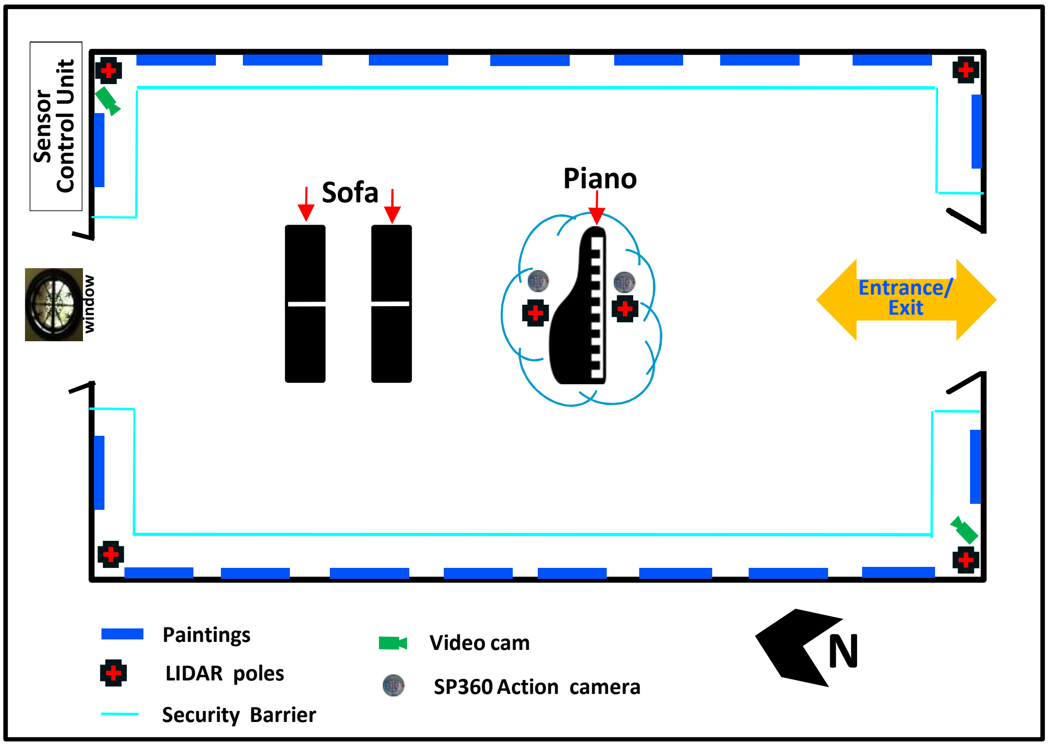

Figure 5.10: Illustration of the experimental setup.

from the beginning to the end of their painting viewing time inside the art gallery.

It is noted that, current setup can also continue tracking the visitors with their unique-ID, in cases where they sit on the sofas for a while, and/or who shortly go through the door of the north side of the art gallery to look out the charming window. In general, proposed system detects and tracks people by observing their shoulder positions only inside the art gallery. So, to continue tracking those visitors the method stated in Section 5.3.2 is utilized to reassign the same unique-ID to the visitors when they stand-up from the sofa or return back to the art gallery from the door of the north side of the art gallery, to continue viewing the paintings. In this case, *condition-2* is relaxed to reassign the lost tracking unique-ID to the lost person. Hence, this method compensates for tracking errors to continuously track visitors who view the paintings in different ways inside the art gallery.

The LIDARs are connected using USB extensions to the sensor control unit which was located at the outside of the art gallery; see Figure 5.10. For simplicity, only one PC (Intel core i7 CPU, 3.60 GHz, 8GB RAM) is used to receive and store the sensor data. In addition to the sensor system, four cameras (two handy cams, two SP360 action cameras) were additionally installed inside the art gallery to capture the visual video information of the whole tracking area; see Figure 5.10.

Experiment was conducted on 28^{th} August 2015 with the above mentioned setup and it has been using it to gather visitors' tracking data for 7 hours (09:00 to 16:00). Later, the stored data were applied to the proposed method to analyze the achieved performance of it in terms of tracking accuracy. Additionally, an example is provided where the obtained large amount of continuous human tracking data can provide very useful knowledge about the visitor's behaviors to the MPs.

5.4.2 Tracking Accuracy Evaluation

For the evaluation, the tracking accuracies of the proposed system is determined for different visitor densities, using manually labelled data. For the labeling, continuous tracking data is employed for three one-hour periods (11:00 to 12:00, 13:00 to 14:00 and 15:00 to 16:00). It can be called these periods *light-density*, *high-density* and *moderate-density*, respectively, depending on the number of visitors that entered the art gallery for each one hour period. During the time periods, each visitor would be labelled upon first entering the art gallery and there were a total of 344 visitors. Labeling was done by comparing the view from the video cameras' information and the tracking results. But, because of the manual labeling, the obtained evaluation result can only be considered approximate.

As the ultimate goal is to observe each individual's behaviors inside the art gallery using the proposed system, it is necessary to evaluate the accuracy by how long a visitor is tracked using the assigned unique-ID for the duration of his/her total time in the art gallery. For every labelled visitor, we compute the *Accuracy Index (AI)* using Equation 5.4, and if the value of the AI for a visitor is above 80-called AI_{80}, then it is assumed that the visitor is tracked for the majority of the time with his/her assigned unique-ID inside the art gallery. If the value of the AI for a visitor is 100-called AI_{100}, then that visitor was fully tracked by the proposed system with his/her assigned unique-ID. Thus, the AI matric is used to evaluate the tracking accuracy of the proposed system. Table 5.1 illustrates

Table 5.1: Tracking accuracy evaluation of our proposed system under different visitor densities.

Tracking Scenario	*Light Density* (11:00 to 12:00)	*Moderate Density* (15:00 to 16:00)	*High Density* (13:00 to 14:00)
Number of-labeled visitors	59	107	178
Average density-(visitors)	9.17	14.16	19.94
Maximum gathering-at a time	19	25	31
% of visitors secured-AI_{80}	96.61	78.50	71.34
% of fully tracked-visitors (AI_{100})	89.83	73.83	61.23

the accuracy evaluation of the proposed visitor tracker system in the considered art gallery under different visitor densities.

$$AI = \frac{\text{Tracking time with assigned unique-ID}}{\text{Total passed time inside the art gallery}} \times 100 \tag{5.4}$$

The accuracy of any tracking system is highly dependent on the coverage of the observation area using the deployed sensor system. It was seen that, by using six LIDAR poles inside the art gallery, the coverage of the deployed sensor system was satisfactory when there were about 20 people(a moderate level of density in that ar gallery); see Figure 5.11. With higher densities of people, there were sometimes fully occluded regions for short periods of time. Thus, it was planned to setup the sensor system inside the art gallery carefully to reduce the impact of occlusions (due to high density of people) on tracking people with unique-IDs in the art gallery.

The proposed human tracking system performs quite satisfactorily when the visitor density in the art gallery is low. But an increase in density leads to a decrease in tracking performance. Although the method stated in Section 5.3.2 is utilized to combat the full occlusion of visitors in the case of high visitor density, there were still errors. The main source of errors are lost tracks of some visitors

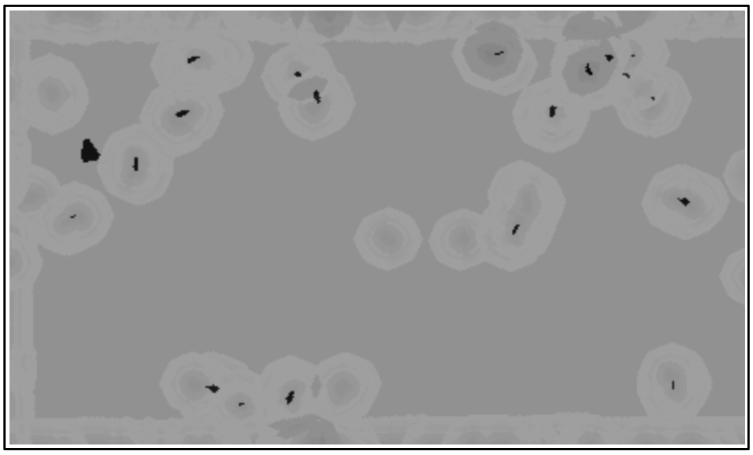

Figure 5.11: LIDAR emitted beam image inside the art gallery: Light gray circles with black spot indicate humans; and other gray colored space indicate the area covered by the laser beam; the black region indicates the only fully occluded region inside the art gallery with the density of people being more than 20.

who were in a group or family. It is observed that, in the case of a group or family of visitors, they often tend to move closely together from one painting to another. In such cases, in a particular small region in front of any painting, the density of the visitors would be very high, thereby causing the system to fail in identifying and tracking each of the group members with their initially assigned unique-ID. Besides that, the presented method described in Section 5.3.2 also failed to reassign unique-IDs to visitors that stood up from sofas while with their group, or those that returned back to the art gallery with a group from the door of the north side of the art gallery. In most cases, the system recovered the track, but there were also a number of ID changes. The evaluation for the *high-density* samples when the art gallery was mostly crowded, gives an insight into the influence of person density on the tracking.

It is noted that in general, the proposed system applied trackers to each visitors with a unique-ID at the entrance/exit of the art gallery one by one when their shoulders were detected by the sensory system, and finally removed those applied trackers automatically when they left the art gallery through the entrance/exit gate. Using shoulder height sensors prevented a major cause of false positives because the height of the most objects in the gallery such as the piano, below the shoulder height of a typical visitor. But false negatives did occur because it

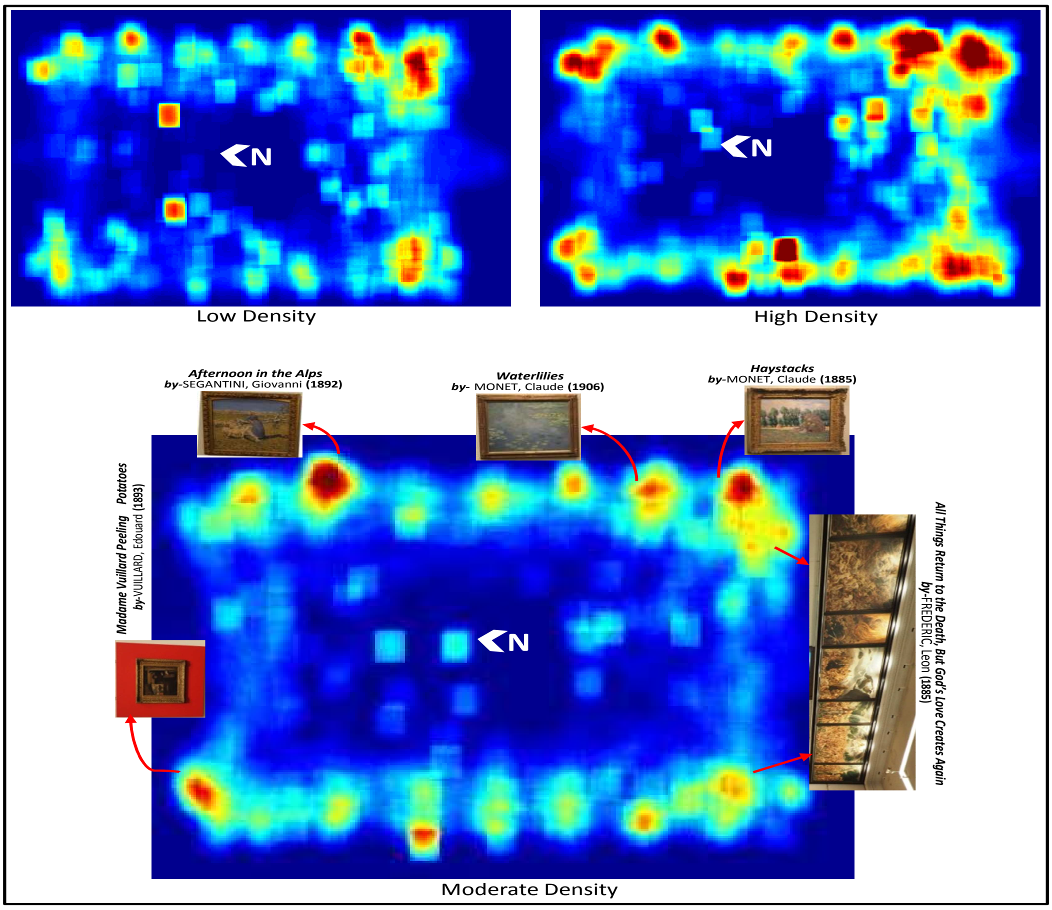

Figure 5.12: Visualization of visitors' movements and preferences to the paintings; (below) Afternoon Sample (Moderate Density); the photos show the most likeable paintings.

is observed from the captured video data that among the total labelled visitors, there were children, toddlers and wheelchair users who were not observable by the deployed LIDAR based sensing system since both of their heights were less than the height of the LIDAR poles. This is another source of errors.

5.4.2.1 Visualization of Visitors' Movement Patterns and Preferences to Exhibits

In this study, based on the recorded data, heat maps is used to generate meaningful and intuitive visualizations of the flow of visitors and their preferences to the paintings through the target art gallery. Figure 5.12 visualizes the positions of the visitors in the art gallery where fully tracked visitors often stopped in different density scenarios. The most stopped at positions (dark-red) of visitors inside the art gallery reflects the message that there are interesting paintings around those positions from where visitors usually view their most liked paintings. It is noted that, at the outgoing corridor of that art gallery, the visitors were often asked randomly to know about their most liked paintings. The top ranked most likeable paintings information from these surveys matched with the heat maps generated information obtained from the proposed human tracking system's provided data, which is illustrated in Figure 5.12 (bottom) in the case of moderate density scenario. It is also observed from all the figures in Figure 5.12 that because visitors usually move in an anti-clockwise pattern through the paintings in the gallery, the first paintings in this anti-clockwise path received relatively more attention than the ones at the end of the path. Melton [134] called this "exit gradient" as a special case when visitors move through a art gallery. Thus it can be stated that the proposed human tracking system is very much effective for the MPs to gather such types of valuable information autonomously. These types of information indicate whether certain people are interested in some selected exhibits and may help the MPs to rearrange the exhibits to make the art gallery more attractive and/or to adapt attractive services for providing additional details to the visitors about those exhibits in that art gallery.

5.4.3 Application of the proposed System for the MPs: Statistical Analysis

The proposed human tracking system permits to observe the undisturbed behaviors of the visitors inside the art gallery for extended periods of time. Thus, inside the art gallery, it is possible to gather knowledge on, how many visitors visited, which directions (clockwise or counter clockwise) they usually moved to view the paintings, how long they were inside the art gallery, which paintings are mostly liked by the visitors, how many people were seated on the sofa to rest, how many visitors were in special areas of the room for a particular time period. Thus, with the proposed system, the interests and intentions of the visitors can be naturally observed inside the art gallery, which will definitely be valuable information for

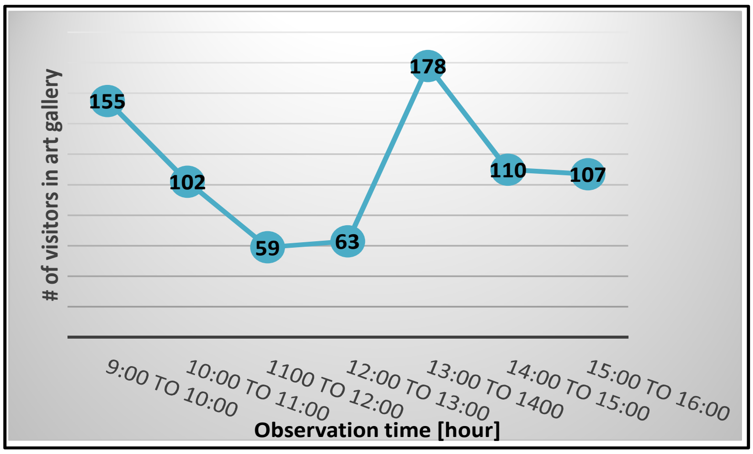

Figure 5.13: Total Number of Visitors to the Art Gallery on an Hourly Basis.

MPs to take into account for any further decisions to make the art gallery more attractive.

Here, some example statistics are illustrated that were able to extract from the aggregated data about visitors by the proposed environmentally deployed sensor system. Figure 5.13 shows the variations of the total number of visitors inside the art gallery on the experimental day. It is seen that the number of visitors after lunch was remarkably high. Before and after the lunch period, the average number of visitors were almost the same. It is also observed from the recorded data that 88.76% of the total visitors moved counter clockwise inside the art gallery, as is typical (>75%) in an art gallery which was claimed by Melton in [134, 135]. Furthermore, it is also observed that 24.54% of the visitors would sit on the sofa while 6.58% would go through the north gate of the art gallery to view the natural beauty of outside window. This is the type of information that the museum curator asked while the experiment was conducted in that art gallery.

To obtain more specific behaviors of the visitors inside the art gallery, the 1 hour afternoon sample (*Moderate Density*) is used in our system. The positional data of the total number of visitors were extracted who were completely tracked by the proposed human tracking system. From the extracted data, visitors' stereotypical movements can also be categorized into four categories in that art gallery as proposed in ethnographic studies by Veron and Levasseur in [198]. The four visiting styles are based on animal behaviors are ant, fish, grasshopper,

Table 5.2: Different categories of visitors based on their visiting style.

Visiting Style	[%]
Ant visitor	20.78
Fish visitor	6.5
Grasshopper visitor	38.96
Butterfly visitor	33.77

and butterfly. The ant visitor spends quite a long time to observe all the exhibits by walking closer to exhibits but avoids empty spaces. The fish visitors prefer to move and stop at empty spaces but avoid areas near exhibits. The grasshopper visitors spend long periods of time seeing selected exhibits but ignore the rest of exhibits. The butterfly visitors observe almost the exhibits but spend varied times to observe each exhibit. Identifying visitors' visiting styles by the MPs could be advantageous for setting up an effective guide system in museums as mentioned in [22, 56, 214]. The obtained results on categorization of visitors based on their visiting style is summarized in Table 5.2. Figure 5.14 shows the visualizations of the different category of visitors' walking trajectories. The positions in the art gallery where visitors how often stopped are illustrated in Figure 5.14(Left). Figure 5.14(Right) shows very clearly how different categories of visitors tend to move to view paintings.

5.4.4 Discussion

Practically tracking in public spaces has been considered difficult because the arrangement of static objects in such places is very variable over time. But, the layout of the contents in art galleries are quite well organized over long periods of time. So, the setup presented here does not show any big problem. Actually, the aim of this research was to setup the sensor arrangement inside the art gallery to be as simple and unobtrusive as possible to track the visitors so that it would not detract from the natural experience of attending the gallery. Thus, a LIDAR pole based sensor system was chosen. Due to its portability, it is more convenient to arrange such types of setups in any other art galleries within a very short amount of time with few technical professionals for the same purposes. LIDARs are much more cost effective and commercially available on the market as opposed us to

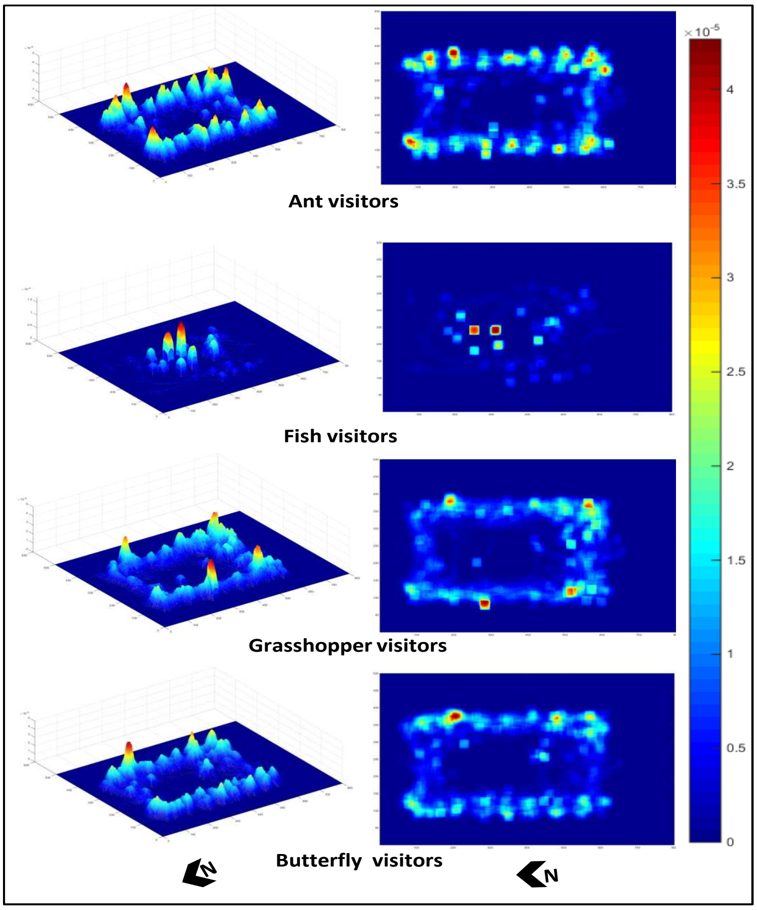

Figure 5.14: Heat map image of different types of visitors' movements in the art gallery.

3D range sensors because of their prohibitive increase in cost and necessities of maintenance with highly technical professionals.

5.5 Chapter Summary

This chapter introduces a method which uses LIDAR to identify humans and track their positions, body orientation, and movement trajectories in any public space to read their various types of behavioral responses to surroundings. We use a network of LIDAR poles, installed at the shoulder level of typical adults to reduce potential occlusion between persons and/or objects even in large-scale social environments. With this arrangement, a simple but effective human tracking method is proposed that works by combining multiple sensors' data so that large-scale areas can be covered. How valuable information related to people's behaviors can be autonomously collected and analyzed using this method is also described. Additionally, a solution to visualize people's movement patterns and preferences with respect to any social space is presented. Thereafter, the effectiveness of the proposed human detection and tracking method is evaluated in an art gallery of a real museum. Ultimately, results revealed good human tracking performance and provided valuable behavioral information related to the art gallery, which are very important for the MPs to take into account in making decisions on improving the attractiveness of any art gallery by introducing more services for the visitors. Additionally, it can be concluded that the obtained visitor's behavioral knowledge from these experiments could be very useful for the museum professionals and curators to achieve knowledge on visitor's undisturb experiences on the exhibits. Thus, the obtained experimental results could help them to adapt new services inside the art gallery and/or rearranging the paintings to make the art gallery more attractive for the future visitors.

5.5.1 Limitations

Although the aim of the presented robust human tracking system in this chapter is to utilize it for social robots to observe different types of behaviors from people to know their interests, intentions, and preferences in any public places, it has not been utilized yet for service robots. We primarily consider the behaviors of individuals but during the experiments, we found that people often visit the art gallery in groups. For future work, observing the behaviors in terms of groups of people inside the art gallery can be considered in addition to observing individual's behaviors. Additionally, based on the observed behaviors by the presented human tracking system, in the future, it can be used in any art gallery like popular public space in support of service robots in real time to offer proactive services.

Chapter 6

Conclusions

Nowadays, it is very important to introduce social robots in social spaces with the ability to proactively offer services to humans. To achieve such an ability, social robots should have the capabilities to observe human behaviors so that they can easily understand their internal states because human internal states reflect their interests, intentions, and preferences concerning surrounding environments. In this dissertation, we explored issues related to recognizing human interests, intentions, and preferences concerning surrounding environments using modern human sensing technologies to provide proactive services to the humans by social robots. This dissertation began with a review of the existing literature of the interdisciplinary research related to our considered issues.

Social robots play very important roles in human world. Understanding the human behaviors is a challenging task for social robots to offer their proactive services to humans in real world environments. We analyzed video camera footage and laser sensor based tracking system extracted human tracking data both in laboratory controlled environments and in an art gallery of a real museum (Ohara Museum of Art, Kurashiki, Japan) in order to find out about how the various types behaviors of the human concerning the exhibits. These types of analysis are thoughtful before introducing proactive guide robot services in museum scenario.

The primary focus of this dissertation is to develop a human behavior sensing system for social robots to estimate the interests, intentions, and preferences of people concerning surrounding environments in order to initiate proactive human-robot interaction by social robots in museum scenarios. Towards these larger goals, this work has made a set of methodological, theoretical, and practical contributions. The methodological contributions include an interdisciplinary, intergrade process for designing, building, evaluating, and demonstrating human behaviors sensing systems using modern sensing technologies to estimate their interests, intentions, and preferences to the exhibits in museum scenarios. These contributions are listed in Section 6.1. The theoretical contributions advance our understanding of human communicative mechanisms from a computational point of view and of peoples responses to theoretically based manipulations in these mechanisms when they are enacted by social robots. Section 6.2 summarizes these contributions. The technical contributions include the computational models of social behavior created for the empirical studies, which are described in Section 6.3. To designing human behavior sensing systems to estimate their interests, intentions, and preferences to the exhibits in museum scenarios, several technical and methodological challenges are faced which are remain significant bottlenecks in developing proactive behaviors of the museum guide robots. Section 6.4 discusses the central challenges and provides a vision for how future work might address them. The last section in this chapter provides my closing remarks.

6.1 Methodological Contributions

This dissertation presents an interdisciplinary research approach that combines techniques and methods from several research domains such as sociology, psychology, and human-robot interaction. Furthermore, it also presents an incremental process for studying and designing the methods to estimate human's interests, intentions, and preferences concerning surrounding environments. Additionally, it demonstrated and evaluated the effectiveness of the designed proactive communicative mechanisms of the social robots to interact with humans in museum scenarios. Table 6.1 lists these contributions.

6.2 Theoretical Contributions

The theoretical contributions of this work consist of a set of new knowledge extracted from a psychology, sociology, cognitive science and human-robot interaction fields that help a deeper understanding on human interests, intentions, and

Table 6.1: Methodological contributions of the dissertation.

Context	**Contributions**
All Studies	•A theoretically and empirically grounded, interdisciplinary process for designing, implementing and evaluating human behavior tracking system to estimate human's internal states to offer proactive services by social robots.
Chapter 3	•An experimental framework for studying how a museum guide robot offers proactive guidance to the visitors based on vision based visitors detection and *local behavior* tracking system in terms of subjective and quantitative evaluations.
Chapter 4	•An experimental framework for studying how a museum guide robot system offers proactive guidance to the visitors by judging their both *local* and *global behavior* inside the museum.
Chapter 5	•An experimental framework for designing, implementing, and evaluating a robust human behavior tracking system usable in an art gallery like real crowded public space where dealing with partial and full occlusion is a challenging task to tracking people continuously for deploying future guide robot systems to offer proactive services.

preference extracting mechanisms as applied to the guide robot system so that guide robots gain the ability to judge human behaviors and their social/cognitive outcomes. Theoretical contributions are listed in Table 6.2.

6.3 Technical Contributions

The technical contributions of this dissertation include a set of design, and behavioral variables for judging the behaviors to know the human's internal states, and computational models of a social robot's proactive behaviors that were created for empirical studies of this dissertation. Table 6.3 provides a detailed list of these contributions.

Table 6.2: Theoretical contributions of the dissertation.

Context	Contributions
All Studies	•Evidence that social robot's verbal and gestural (head shaking, hand waving and body movement) actions lead to significant social and cognitive outcomes, particularly better feeling of attraction toward the social robot while proactively offer guidance to the visitors in museum scenarios.
Chapter 3	•Evidence that social robot's successful initiation of proactive guidance to the visitors takes four key steps process: observe visitors' presence to select one as a target visitor, tracks head orientation of the target visitor to determine his/her VFOA as his/her *local behavior* toward the painting, proactively move toward the target visitor's social distance, finally, initiate proactive guidance.
Chapter 4	•Evidence that network enabled social robot's successful initiation of proactive guidance to the visitors takes three key steps process: observe visitors' *global behavior*, then observe their *local behavior* toward the paintings if typical *global behavior* found, finally, offer proactive commentary to the visitors.
Chapter 5	•Evidence that the robust human behavior tracking system for future museum guide robot system enabled the museum curators and museum professionals to achieve valuable undisturbed behavioral information related to the art gallery. •Evidence that visitors moved counter clockwise inside the art gallery as is typical. •Evidence that four categories of visitor's stereotypical movements (ant, fish, grasshopper, and butterfly) occur in an art gallery. •Evidence that an "exit gradient" exists as a special case when visitors move through an art gallery. In this case, the first paintings in the movement path in an art gallery received relatively more attention than the one at the end of the path.

Table 6.3: Technical contributions of the dissertation.

Context	Contributions
All Studies	•Conceptually designed and implemented smart art galleries of a typical museum by incorporating modern human behavior sensing system for museum guide robot to perform the proactive human-robot interaction studies. •Technically designed and implemented an environmentally distributed modern human tracking system useable in an art gallery of a real museum for future museum guide robot.
Chapter 3	•A computational model of identifying interested visitors through observing their *local behavior* is programmed in C, C++, and OpenCV library functions. •Additionally, a piece of dedicated software called "RobovieMaker2" is used for imposing AI to the Robovie-R3 for controlling its proactive behaviors toward the target visitor.
Chapter 4	•A computational model of identifying interested visitors through observing their both *local* and *global behavior* is programmed in C, C++, and OpenCV library functions. •Additionally, a piece of dedicated software called "RobovieMaker2" is used to implement the AI in the *Desktop Robot* to proactively initiate guidance to the visitors about the exhibits in the museum scenario.
Chapter 5	•An integrated computational model of tracking human positions, body orientations in real world environments is programmed in C, C++, and OpenCV library functions. •CUDA (Compute Unified Device Architecture) is utilized in parallel computation related programming tasks to track large numbers of people even in large-scale environments.

6.4 Future Work

In this dissertation, attempts have been made to solve a highly complex and unconventional design problem in HRI-designing a human behavior tracking system to estimate their interests, intentions, and preferences concerning surrounding environments with incorporation of the psychology, sociology, cognitive science, and human-robot interaction fields. The ultimate goal is to gain a better understanding of human behaviors to design an HRI system which will enable social robots to proactively provide services to the humans. There are still several issues that have not been addressed in the current model. Some of the issues are discussed in the following.

•**Generalizability:** We tested the proposed HRI model for a specific scenario where group of participants were asked to visit the designed art gallery of a typical museum with some instructions. Then their behaviors were analyzed to estimate their internal states. Therefore, its generalizability is limited. More studies are needed to judge the visitors' behaviors in a real museum before introducing the proactive services of the social robots in museums in real world spaces.

•**Limited Interaction:** In this dissertation, the social robot's actions designed for the proposed HRI systems are limited to use voice and gestural (head shaking, hand waving, and body movement in between the visitors and the exhibits) actions to draw attention and offer proactive commentary about the exhibits. Social robots may also need to use eye gaze movements or physical touch depending on the various complex situations. Therefore, it is needed to explore other possible situations and design appropriate actions for those situations to make the activities of the social robot more natural.

•**Controlled Laboratory Settings:** The HRI approach presented here used controlled laboratory experiments to understand the social and cognitive outcomes of the designed human intention, interest, and preference estimation model as well as social robot's proactive approaching model. To prove the current approaches will be effective beyond controlled laboratory settings, future work needs to also situate designed behaviors in real-world scenarios and contexts.

•**Mobility of Social Robots:** In this dissertation, the mobility of the social robots as museum guide robots is partially considered before proactively offer guidance to the human. Studies presented in Chapter 3 considered social robot's semi-autonomous movements where it can only move to some predefined location to offer proactive service, whereas in studies presented in Chapter 4, the social robots are considered to be stable in some exhibit specific fixed locations to offer proactive services. Thus, the mobility of the social robots in this study is limited. More studies are needed to make social robots with the capability to autonomously move anywhere to offer proactive services inside real art galleries.

6.5 Closing Remarks

We believe that it is a very challenging task to the HRI research community to implement the intelligence inside the social robot to estimate human interests, intentions, and preferences toward the surrounding public environments before introducing their proactive services for human activities because in this way humans can acquire exclusive proactive services from the social robots in the same manner as that from humans. To introduce such proactive services from social robots to humans, in this dissertation, we proposed more than one HRI system in which modern human sensing technology based methods are presented to judge the behaviors (*local* and *global behavior*) of the humans in the art gallery scenario to estimate their interests, intentions, and preferences towards the exhibits (paintings). It is revealed that the outcomes of the presented human sensing methods enabled the social robots to provide proactive services to the humans. Furthermore, based on our observation of human behaviors by using our designed, implemented, and deployed robust human tracking system in a real art gallery, it can be concluded that humans always exhibit some typical behaviors (e.g. stay for a long time in front of their most liked paintings).

In this process, we employed methods and knowledge from a variety of disciples (such as psychology, sociology, cognitive science, and robotics) and made a number of design decisions that were grounded in theory and empirical data. While further work remains in order to improve the validity of these decisions and the generalizability of the results, this dissertation provides a major step towards designing important social capabilities (such as proactively providing human social services to people) for social robots using a theoretically and empirically grounded methodology. We believe that our proposed HRI systems have the ability to provide proactive novel services to humans in real world environments.

Appendix A

Data Collection Techniques

This appendix includes the methods or techniques used in this dissertation to gather the data from the human-robot interaction studies. Additionally, it also includes the techniques and tools to gather the data about the preferences of the visitors in an art gallery of a real museum. To collect the data, we used questionnaires as well as observation methods. Table A.1 summarizes these techniques. The questionnaire data was collected in terms of the Likert scale.

A. Questionnaire Based Method: A questionnaire is a set of questions for gathering information from individuals. The project leader administers questionnaires by mail, telephone, using face-to-face interviews, as handouts, or electronically (i.e., by e-mail or through Web-based questionnaires). Questionnaires are a well-establish technique for collecting demographic data and users' opinions. Efforts and skills are needed to develop questions that clearly communicate what the information seeker wants to know. Questionnaires can be used on their own or in conjunction with other methods to clarify or deepen understanding. It is important that questions are specific; when possible, closed questions should be asked and a range of answers offered using clear and simple wording written at the reading level of the participants, including a 'no opinion' or 'none of these' option.

B. Likert Scale: Likert scales are used for measuring opinions, attitudes, and beliefs, and consequently they are widely used for evaluating user satisfaction with systems. The purpose of this rating scale is to elicit a range of responses to a question that can be compared across respondents. This is good for getting peo-

Table A.1: Overview of data gathering techniques used in this dissertation.

Technique	Good for	Data types	Pros	Cons
Questionnaire	✓Answering specific questions, ✓To gather data about individuals' knowledge, beliefs, attitudes, and behaviors	Quantitative and qualitative	✓Can reach many people with low resource, ✓Administers can disseminate questionnaires relatively inexpensively, ✓It is helpful in maintaining participants' privacy because participants' responses can be anonymous or confidential	✓The design is crucial. ✓Response rate may be low. ✓Responses may not be what administer want.
Observation	✓Capturing the detail of what individuals do, ✓When data collection from individual is not a realistic option through questionnaire	Quantitative and qualitative	✓Can focus on the details of a task without interruption, ✓Result may have greater use if conducted in real environments	✓Results may have limited use in the real environments if observation conducted in controlled artificial/designed environments.

Very ineffective	Ineffective	Somewhat ineffective	Undecided	Somewhat effective	Effective	Very effective
(1)	(2)	(3)	(4)	(5)	(6)	(7)

Figure A.1: 7-point Likert scale with corresponding rating features.

.

ple to make judgements about things, e.g., how easy, how effective, and such like. The success of Likert scales relies on identifying a set of statements representing a range of possible opinions and this scale is more commonly used because identifying suitable statements that respondents will understand easily. When designing the Likert scale, issues that need to be addresses include: How many points are needed on the scale, How should they be presented, and in what form? Many questionnaires use seven-or-five-point scales and there are also three-point scales. Longer range is better, when asking respondents to make subtle judgements. In this dissertation, we have used a 7-point Liket scale to collect the participants' impression on robot's behaviors using the questionnaire method where '1' stands for definite no/very ineffective, and '7' stands for definitely yes/very effective (see in Figure. A.1 as an example).

C. Observation Based Method: Observation is way of gathering data by watching behavior, events, or noting physical characteristics in their natural setting. Observations can be overt (everyone knows they are being observed) or covert (no one knows they are being observed and the observer is concealed). The benefit of covert observation is that people are more likely to behave naturally if they do not know they are being observed. However, typically conducting overt observations is suitable because of ethical problems related to concealing observation. Observations can also be either direct or indirect. Direct observation is when someone watch interactions, processes, or behaviors as they occur; for example, observing a teacher teaching a lesson from a written curriculum to determine whether they are delivering it with fidelity. Indirect observations are when someone watch the results of interactions, processes, or behaviors; for example, measuring the amount of plate waste left by students in a school cafeteria to determine whether a new food is acceptable to them.

Data collection through observation methods and techniques may take place in the real world environments, or in a controlled environment. In the former case, individuals can be observed in natural settings. For example, museum curators or professionals can observe whether the layout of the paintings in an art gallery is conducive to learning for visitors. In the latter case, individuals are observed performing specified tasks within a controlled environment. In this dissertation, we have used both recorded video data and laser sensor based tracking system extracted data in both controlled laboratory settings (see in Chapter 4) and an art gallery of a real museum (see in Chapter 5) to gather observational data. The number of successful attempts of the guide robot to initiate proactive interaction with humans (see in Chapter 3) is an example of quantitative measures collected by observation.

References

[1] ... Naoko x robot, available: http://www.naoko-robot.net, 2013. 55

[2] ... Scanning laser range finder uxm-30lx-ew specification. Hokuyo Automatic Co., LTD, 2014. 86

[3] E. Aguirre, M. García-Silvente, and D. Pascual. *A Multisensor Based Approach Using Supervised Learning and Particle Filtering for People Detection and Tracking*, pages 645–657. Springer International Publishing, Cham, 2016. 20

[4] E. Aguirre, M. Garcia-Silvente, and J. Plata. *Leg Detection and Tracking for a Mobile Robot and Based on a Laser Device, Supervised Learning and Particle Filtering*, pages 433–440. Springer International Publishing, Cham, 2014. 20

[5] A. O. Alexander Carballo and S. Yuta. Reliable people detection using range and intensity data from multiple layers of laser range finders on a mobile robot. *International Journal of Social Robotics*, 3(2):167–186, Apr. 2011. 21, 24

[6] B. Ali, Y. Ayaz, M. Jamil, S. Gilani, and N. Muhammad. Improved method for stereo vision-based human detection for a mobile robot following a target person. *South African Journal of Industrial Engineering*, 26:102 – 119, 05 2015. 19, 20

[7] B. Ali, K. F. Iqbal, Y. Ayaz, and N. Muhammad. Human detection and following by a mobile robot using 3D features. In *2013 IEEE International Conference on Mechatronics and Automation*, pages 1714–1719, Aug 2013. 20

[8] I. Ali and M. N. Dailey. Multiple human tracking in high-density crowds. *Image and Vision Computing*, 30(12):966 – 977, 2012. 19

[9] I. G. Alonso, editor. *International Technology Robotics Applications-Proceedings of the 2nd INTERA Conference*, volume 70. Springer International Publishing, 2014. 13

[10] R. C. Arkin. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture part i: Motivation and philosophy. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 121–128, March 2008. 16

[11] K. Arras and S. Vestli. Hybrid, high-precision localisation for the mail distributing mobile robot system mops. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 4, pages 3129–3134, May 1998. 15, 21, 54

[12] H. Asoh, S. Hayamizu, I. Hara, Y. Motomura, S. Akaho, and T. Matsui. Socially embedded learning of the office-conversant mobile robot jijo-2. In *In Proceedings of 15th International Joint Conference on Artificial Intelligence (IJCAI-97*, pages 880–885, 1997. 15, 54

[13] C. Astua, R. Barber, J. Crespo, and A. Jardon. Object detection techniques applied on mobile robot semantic navigation. *Sensors*, 14(4):6734, 2014. 16

[14] S. Ba and J.-M. Odobez. Recognizing visual focus of attention from head pose in natural meetings. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):16–33, Feb 2009. 38

[15] S. O. Ba. *Joint Head Tracking and Pose Estimation for Visual Focus of Attention Recognition*. PhD thesis, Ãcole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 3 2007. 36, 38, 39

[16] C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 591–594, Sept 2004. 9, 31

[17] M. Begum, R. W. Serna, D. Kontak, J. Allspaw, J. Kuczynski, H. A. Yanco, and J. Suarez. Measuring the efficacy of robots in autism therapy: How informative are standard hri metrics'. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 335–342, New York, NY, USA, 2015. ACM. 12, 54

[18] N. Bellotto and H. Hu. Multisensor-based human detection and tracking

for mobile service robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):167–181, Feb 2009. 20

[19] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke. Towards a humanoid museum guide robot that interacts with multiple persons. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, pages 418–423, Dec 2005. 15, 20, 40

[20] N. Bergstrom, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita. Modeling of natural human-robot encounters. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 2623–2629, Sept 2008. 54

[21] A. Bevilacqua, L. D. Stefano, and P. Azzari. People tracking using a time-of-flight depth sensor. In *2006 IEEE International Conference on Video and Signal Based Surveillance*, pages 89–89, Nov 2006. 22

[22] A. Bianchi and M. Zancanaro. Tracking users movements in an artistic physical space. In *Proceedings of the The i3 Annual Conference*, pages 103–106, Siena, Italy, October 20-22 1999. 94

[23] J. Blanco, W. Burgard, R. Sanz, and J. Fernandez. Fast face detection for mobile robots by integrating laser range data with vision. In *Proc. of the International Conference on Advanced Robotics (ICAR)*, 2003. 20, 23

[24] C. Breazeal. *Designing Sociable Robots*, page 256. MIT Press, Cambridge, MA, USA, 2002. 8, 9

[25] A. Brooks and S. Williams. Tracking people with networks of heterogeneous sensors. In *In the proceedings of the Australian Conference on Robotics and Automation*, pages 1–7, 2004. 21

[26] D. Brscić, T. Kanda, T. Ikeda, and T. Miyashita. Person tracking in large public spaces using 3-D range sensors. *Human-Machine Systems, IEEE Transactions on*, 43(6):522–534, Nov 2013. 20, 21, 22, 23, 25, 33, 54, 74

[27] D. Brscić, H. Kidokoro, Y. Suehiro, and T. Kanda. Escaping from children's abuse of social robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 59–66, New York, NY, USA, 2015. ACM. 20, 27

[28] R. Brunelli, O. Lanz, A. Santuari, and F. Tobia. *PEACH - Intelligent Interfaces for Museum Visits*, chapter Tracking Visitors in a Museum, pages 205–225. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 33

[29] W. Burgard, A. B. Cremers, D. Fox, D. Hanel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. The Interactive Museum Tour-Guide Robot. In *Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, 1998. 1, 20, 27, 34, 37, 75

[30] W. Burgard, A. B. Cremers, D. Fox, D. Hhnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(12):3 – 55, 1999. 15, 54, 75

[31] J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied. Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, 5(4):593–618, 2013. 12

[32] G. Chen, S. A. King, and M. Scherger. *Informatics in Control, Automation and Robotics: Volume 2*, chapter Robot Remote Control Using Bluetooth and a Smartphone Augmented System, pages 453–460. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 20

[33] W. Choi, C. Pantofaru, and S. Savarese. Detecting and tracking people using an RGB-D camera via multiple detector fusion. In *Workshop on Challenges and Opportunities in Robot Perception, at the International Conference on Computer Vision (ICCV)*, 11/2011 2011. 20, 23

[34] W. Choi, C. Pantofaru, and S. Savarese. Detecting and tracking people using an rgb-d camera via multiple detector fusion. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1076–1083, Nov 2011. 23

[35] C. T. Chou, J.-Y. Li, M.-F. Chang, and L. C. Fu. Multi-robot cooperation based human tracking system using laser range finder. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 532–537, May 2011. 20

[36] T. M. Christopher and D. M. Marc. Development of fast food service robot, 2006. 1

[37] W. Chung, H. Kim, Y. Yoo, C. B. Moon, and J. Park. The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder. *IEEE Transactions on Industrial Electronics*, 59(8):3156–3166, Aug 2012. 20

[38] M. G. Cline. The perception of where a person is looking. *The American Journal of Psychology*, 80(1):41–50, 1967. 38

[39] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Laser-based interacting people tracking using multi-level observations. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 1799–1804, Oct 2006. 20, 21

[40] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Laser-based detection and tracking of multiple people in crowds. *Comput. Vis. Image Underst.*, 106(2-3):300–312, May 2007. 20

[41] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Multi-modal tracking of people using laser scanners and video camera. *Image and Vision Computing*, 26(2):240 – 252, 2008. 23

[42] D. Das, M. Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno. Vision-based attention control system for socially interactive robots. In *RO-MAN, 2012 IEEE*, pages 496–502, Sept 2012. 20

[43] D. Das, M. G. Rashed, Y. Kobayashi, and Y. Kuno. Supporting human-robot interaction based on the level of visual focus of attention. *IEEE Transactions on Human-Machine Systems*, 45(6):664–675, Dec 2015. 18, 36

[44] K. Dautenhahn and I. Werry. A quantitative technique for analysing robot-human interactions. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, volume 2, pages 1132–1138 vol.2, 2002. 18, 54

[45] M. Díaz, D. Paillacho, C. Angulo, O. Torres, J. González, and J. Albo-Canals. A week-long study on robot-visitors spatial relationships during guidance in a sciences museum. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 152–153, New York, NY, USA, 2014. ACM. 34

[46] P. Duan, G. Tian, and W. Zhang. Human localization based on distributed laser range finders. *International Journal of Hybrid Information Technology*, 7(3):311–324, 2014. 21

[47] N. Eagle and A. (Sandy) Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, Mar. 2006. 22

[48] M. Economou. The evaluation of museum multimedia applications: lessons

from research. *Museum Management and Curatorship*, 17(2):173 – 187, 1998. 75

[49] S. K. Elizabeth, P. Rhea, S. Frederick, and S. Brian. Bridging the research gap: Making hri useful to individuals with autism. *Journal of Human-Robot Interaction*, 1(1):26–54, 2012. 12

[50] T. Ellersiek, G. Andrienko, N. Andrienko, D. Hecker, H. Stange, and M. Mueller. Using bluetooth to track mobility patterns: Depicting its potential based on various case studies. In *Proceedings of the Fifth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, ISA '13, pages 1–7, New York, NY, USA, 2013. ACM. 22, 33

[51] J. F. Engelberger. Health-care robotics goes commercial: the helpmate experience. *Robotica*, 11:517–523, 11 1993. 15, 54

[52] A. Fod, A. Howard, and M. Mataric. A laser-based people tracker. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 3, pages 3024–3029, 2002. 21

[53] T. W. Fong, I. Nourbakhsh , and K. Dautenhahn. A survey of socially interactive robots: Concepts, design, and applications. Technical Report CMU-RI-TR-02-29, Robotics Institute, Pittsburgh, PA, December 2002. 8, 9

[54] B. S. Frey and S. Meier. *The Economics of Museums*, volume 1 of *Handbook of the Economics of Art and Culture*, pages 1017 – 1047. Elsevier, 2006. 74

[55] I. Fujimoto, T. Matsumoto, P. R. S. De Silva, M. Kobayashi, and M. Higashi. Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. *International Journal of Social Robotics*, 3(4):349–357, 2011. 12

[56] F. Gabrielli, P. Marti, and L. Petroni. The environment as interface. In *Proceedings of the The i3 Annual Conference*, pages 44–47, Siena, Italy, October 20-22 1999. 94

[57] D. L. Gameiro and J. S. Sequeira. *Robot 2015: Second Iberian Robotics Conference: Advances in Robotics, Volume 2*, chapter RFID-Based People Detection for Human-Robot Interaction, pages 395–406. Springer International Publishing, Cham, 2016. 20

[58] A. Garrell and A. Sanfeliu. Model validation: Robot behavior in people guidance mission using dtm model and estimation of human motion behav-

ior. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5836–5841, Oct 2010. 3

[59] A. Garrell, M. Villamizar, F. Moreno-Noguer, and A. Sanfeliu. Proactive behavior of an autonomous mobile robot for human-assisted learning. In *2013 IEEE RO-MAN*, pages 107–113, Aug 2013. 20, 27

[60] J. M. Gascuea and A. Fernndez-Caballero. Agent-oriented modeling and development of a person-following mobile robot. *Expert Syst. Appl.*, 38(4):4280–4290, 2011. 19

[61] T. Germa, F. Lerasle, P. Danes, and L. Brethes. Human / robot visual interaction for a tour-guide robot. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3448–3453, Oct 2007. 20

[62] T. Germa, F. Lerasle, N. Ouadah, V. Cadenat, and M. Devy. Vision and rfid-based person tracking in crowds from a mobile robot. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5591–5596, Oct 2009. 20

[63] J. L. Giesbrecht, H. K. Goi, T. D. Barfoot, and B. A. Francis. A vision-based robotic follower vehicle. In *Proc. of The International Society for Optical Engineering*, volume 7332, pages 733210–733212, 2009. 19, 20

[64] D. Glas, T. Kanda, H. Ishiguro, and N. Hagita. Simultaneous people tracking and localization for social robots using external laser range finders. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 846–853, Oct 2009. 20, 21

[65] D. Glas, T. Miyashita, H. Ishiguro, and N. Hagita. Laser tracking of human body motion using adaptive shape modeling. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 602–608, Oct 2007. 34

[66] D. Glas, Y. Morales, T. Kanda, H. Ishiguro, and N. Hagita. Simultaneous people tracking and robot localization in dynamic social spaces. *Autonomous Robots*, 39(1):43–63, 2015. 20, 21, 25

[67] D. F. Glas. *The network RObot System: Enabling Social Robots in the Real World*. PhD thesis, The Graduate School of Engineering and Science, Osaka University, 03 2013. 11

[68] D. F. Glas, D. Brcic, T. Miyashita, and N. Hagita. SNAPCAT-3D: Calibrating networks of 3D range sensors for pedestrian tracking. In *2015 IEEE*

International Conference on Robotics and Automation (ICRA), pages 712–719, May 2015. 20

[69] D. F. Glas, S. Satake, F. Ferreri, T. Kanda, H. Ishiguro, and N. Hagita. The network robot system: Enabling social human-robot interaction in public spaces. *Journal of Human-Robot Interaction*, 1(2):5–32, 2012. 2, 3, 12, 16, 20, 27, 30, 53

[70] R. Gockley, J. Forlizzi, and R. Simmons. Interactions with a moody robot. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 186–193, New York, NY, USA, 2006. ACM. 1

[71] R. Gockley, J. Forlizzi, and R. Simmons. Natural person-following behavior for social robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, HRI '07, pages 17–24, New York, NY, USA, 2007. ACM. 19

[72] C. Grazia, A. Carmela, G. Cataldo, and T. D'Orazio. A kinect-based gesture recognition approach for a natural human robot interface. *Int J Adv Robot Syst*, 12(22):1–11, 2015. 20

[73] A. P. Gritti, O. Tarabini, J. Guzzi, G. A. D. Caro, V. Caglioti, L. M. Gambardella, and A. Giusti. Kinect-based people detection and tracking from small-footprint ground robots. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4096–4103, Sept 2014. 20

[74] H. Gross and H.-J. Boehme. Perses-a vision-based interactive mobile shopping assistant. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 1, pages 80–85 vol.1, 2000. 15, 20

[75] H. M. Gross, H. J. Boehme, C. Schroeter, S. Mueller, A. Koenig, C. Martin, M. Merten, and A. Bley. Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 3471–3478, Oct 2008. 12

[76] Y. Gu, H. Do, Y. Ou, and W. Sheng. Human gesture recognition through a kinect sensor. In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pages 1379–1384, Dec 2012. 20

[77] M. Gyoda, T. Tabata, Y. Kobayashi, and Y. Kuno. Mobile care robot accepting requests through nonverbal interaction. In *Frontiers of Computer*

Vision (FCV), 2011 17th Korea-Japan Joint Workshop on, pages 1–5, Feb 2011. 15, 54

[78] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, and M. Philipose. Mapping and localization with rfid technology. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, volume 1, pages 1015–1020 Vol.1, April 2004. 20

[79] E. T. Hall. A system for the notation of proxemic behavior. *American Anthropologist*, 65(5):1003–1026, 1963. 17

[80] E. T. Hall. *The Hidden Dimension*, chapter Distance in Man. Anchor, oct 1990. 39, 40

[81] S. Han, H. Lim, and J. Lee. An efficient localization scheme for a differential-driving mobile robot based on rfid system. *IEEE Transactions on Industrial Electronics*, 54(6):3362–3369, Dec 2007. 20

[82] D. W. Hansen, H. S. Hansen, M. Kirschmeyer, R. Larsen, and D. Silvestre. Cluster tracking with time-of-flight cameras. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–6, June 2008. 22

[83] A. Haro, M. Flickner, and I. Essa. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 163–168 vol.1, 2000. 38

[84] F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer. Understanding social robots. In *Advances in Computer-Human Interactions, 2009. ACHI '09. Second International Conferences on*, pages 169–174, Feb 2009. xii, 10

[85] N. Hirai and H. Mizoguchi. Visual tracking of human back and shoulder for person following robot. In *Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference on*, volume 1, pages 527–532 vol.1, July 2003. 19, 20

[86] M. Hofmann, D. Wolf, and G. Rigoll. Hypergraphs for joint multi-view reconstruction and multi-object tracking. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3650–3657, June 2013. 19

[87] T. Horiuchi, S. Thompson, S. Kagami, and Y. Ehara. Pedestrian tracking from a mobile robot using a laser range finder. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 931–936, Oct 2007. 20, 21

[88] S. Iengo. *Human Gesture Recognition and Robot Attentional Regulation for Human-Robot Interaction*. PhD thesis, Department of Electronic Systems, Aalborg University, Denmark., 4 2014. 14

[89] T. Ikeda, H. Ishiguro, and T. Nishimura. People tracking by cross modal association of vision sensors and acceleration sensors. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4147–4151, Oct 2007. 21

[90] M. Isard and A. Blake. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998. 56, 78

[91] Y. Iwamura, M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 449–457, 3 2011. 12

[92] A. D. P. James J. Gibson. Perception of another person's looking behavior. *The American Journal of Psychology*, 76(3):386–394, 1963. 38

[93] O. Jenkins, G. Serrano, and M. Loper. Recognizing human pose and actions for interactive robots. *N. Sarkar, Human Robot Interaction. I-Tech Education and Publishing*, 2007. 14

[94] H. Jeonghye, J. Miheon, J. Vicki, and H. J. Jun. Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, 4(4):159–168, 2008. 13

[95] S. Jia, L. Zhao, X. Li, W. Cui, and J. Sheng. Autonomous robot human detecting and tracking based on stereo vision. In *2011 IEEE International Conference on Mechatronics and Automation*, pages 640–645, Aug 2011. 19, 20

[96] G. Joshua, M. Joyce, and B. Meyer. EAGER: An indoor positioning system (IPS) for informal learning experiences. In *Proceedings of the Advancing Informal STEM Learning (AISL) Principal Investigator (PI) meeting*, 2014. 22, 33

[97] E.-J. Jung, J. H. Lee, B.-J. Yi, J. Park, S. Yuta, and S.-T. Noh. Development of a laser-range-finder-based human tracking and control algorithm for a marathoner service robot. *IEEE/ASME TRANSACTIONS ON MECHATRONICS*, 19(6):1963–1976, 2014. 20

[98] A. Kanda, M. Arai, R. Suzuki, Y. Kobayashi, and Y. Kuno. Recognizing groups of visitors for a robot museum guide tour. In *Human System Interactions (HSI), 2014 7th International Conference on*, pages 123–128, June 2014. xiv, 84

[99] T. Kanda, D. Glas, M. Shiomi, and N. Hagita. Abstracting people's trajectories for social robots to proactively approach customers. *Robotics, IEEE Transactions on*, 25(6):1382–1396, Dec 2009. 16, 19, 20, 26, 27, 30

[100] T. Kanda, D. F. Glas, M. Shiomi, H. Ishiguro, and N. Hagita. Who will be the customer?: A social robot that anticipates people's behavior from their trajectories. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 380–389, New York, NY, USA, 2008. ACM. 11, 22, 29

[101] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Comput. Interact.*, 19(1):61–84, 6 2004. 3, 12, 54

[102] T. Kanda and H. Ishiguro. *Human-Robot Interaction in Social Robotics*, volume 70. Tailor and Francis Group, 2013. 21

[103] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 173–180, New York, NY, USA, 2009. ACM. xii, 2, 3, 11, 16, 20, 23, 27, 30, 53, 54, 74

[104] Y. Kato, T. Kanda, and H. Ishiguro. May i help you?: Design of human-like polite approaching behavior. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 35–42, New York, NY, USA, 2015. ACM. 2, 16, 20, 25, 26, 27, 30, 53, 54

[105] O. Keisuke, L. Jae Hoon, and O. Shingo. People tracking by collaboration between a mobile robot and static laser scanner. In *Proceedings of the*

International MultiConference of Engineers and Computer Scientists 2014, volume 2, Hong Kong, March 2014. 19, 20

[106] R. Kelley, M. Nicolescu, A. Tavakkoli, M. Nicolescu, C. King, and G. Bebis. Understanding human intentions via hidden markov models in autonomous mobile robots. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 367–374, March 2008. 20, 26, 27

[107] N. Kemper. Effects of proactive behavior and physical interaction with a social robot. Master's thesis, Faculty of Natural Sciences, Mathematics and Computing, University of Amsterdam, 2009. 2, 31, 32

[108] E. S. Kim, L. D. Berkovits, E. P. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, 43(5):1038–1049, 2013. 12

[109] H. Kim, W. Chung, and Y. Yoo. Detection and tracking of human legs for a mobile service robot. In *Advanced Intelligent Mechatronics (AIM), 2010 IEEE/ASME International Conference on*, pages 812–817, July 2010. 21

[110] K. L. Koay, E. A. Sisbot, D. S. Syrdal, M. L. Walters, K. Dautenhahn, 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*, pages 18–24. AAAI, 2007. 18

[111] K. L. Koay, M. L. Walters, and K. Dautenhahn. Methodological issues using a comfort level device in human-robot interactions. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 359–364, Aug 2005. 18

[112] Y. Kobayashi, Y. Hoshi, G. Hoshino, T. Kasuya, M. Fueki, and Y. Kuno. Museum guide robot with three communication modes. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3224–3229, Sept 2008. 30, 34

[113] Y. Kobayashi and Y. Kuno. People tracking using integrated sensors for human robot interaction. In *Industrial Technology (ICIT), 2010 IEEE International Conference on*, pages 1617–1622, March 2010. 20, 23

[114] Y. Kobayashi, T. Shibata, Y. Hoshi, Y. Kuno, M. Okada, and K. Yamazaki. "i will ask you" choosing answerers by observing gaze responses using integrated sensors for museum guide robots. In *19th International Symposium*

in Robot and Human Interactive Communication, pages 652–657, Sept 2010. 30

[115] Y. Kobayashi, D. Sugimura, K. Hirasawa, N. Suzuki, H. Kage, Y. Sato, and A. Sugimoto. 3d head tracking using the particle filter with cascaded classifiers. In *Proc. BMVC*, pages 5.1–5.10, 2006. doi:10.5244/C.20.5. xiii, 40, 65

[116] S. Kodagoda and S. Sehestedt. Simultaneous people tracking and motion pattern learning. *Expert Systems with Applications*, 41(16):7272 – 7280, 2014. 20

[117] Y. Koide, T. Kanda, Y. Sumi, K. Kogure, and H. Ishiguro. An approach to integrating an interactive guide robot with ubiquitous sensors. In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2500–2505 vol.3, Sept 2004. 20, 27, 32

[118] S. I. Kourosh Meshgi. The state-of-the-art in handling occlusions for visual object tracking. *IEICE Transactions on Infromation and Systems*, E98-D(7):1260–1274, 2015. 24

[119] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 341–346, Aug 2005. 1, 54

[120] G. Kreiman. Biological object recognition. 3(6):2667, 2008. revision 91063. 16

[121] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1191–1194, New York, NY, USA, 2007. ACM. xii, 11

[122] M. C. Lam, A. S. Prabuwono, H. Arshad, and C. S. Chan. *A Real-Time Vision-Based Framework for Human-Robot Interaction*, pages 257–267. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. 20

[123] S. R. H. Langton. The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology*, 53A:825845, 2000. 38

[124] S. R. H. Langton, H. Honeyman, and E. Tessler. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics*, 66(5):752–771, 2004. 38

[125] F. le Coz and F. Lemessier. Multimedia technology at the orsay museum: Institutional experience and future prospects. In *ICHIM*, pages 377–383, 1993. 75

[126] D. Lees and P. Lepage. Robots in education: The current state of the art. *Journal of Educational Technology Systems*, 24(4):299–320, 1996. 1

[127] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pages 75–82, New York, NY, USA, 2015. ACM. 12, 54

[128] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational markov networks. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, IJCAI'05, pages 773–778, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc. 22

[129] Lozano and K. Otsuka. Real-time visual tracker by stream processing. *J. Signal Processing Systems*, 57(2):285–295, 2009. 84

[130] M. Luber, L. Spinello, and K. O. Arras. People tracking in RGB-D data with on-line boosted target models. In *In IEEE/RSJ Int. Conf. on*, 2011. 23

[131] P. Martin, B.-J. Ho, N. Grupen, S. Muñoz, and M. Srivastava. An ibeacon primer for indoor localization: Demo abstract. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, pages 190–191, New York, NY, USA, 2014. ACM. 22, 33

[132] K. Maruyama and M. Endo. The effect of face orientation upon apparent direction of gaze. *Tohoku Psychologica Folia*, 42:126138, 1983. 38

[133] Y. Matsumoto, T. Wada, S. Nishio, T. Miyashita, and N. Hagita. Scalable and robust multi-people head tracking by combining distributed multiple sensors. *Intelligent Service Robotics*, 3(1):29–36, 2010. 20, 21, 23

[134] A. Melton. Some behavior charateristics of museum visitors. In *Phychological Bulletin*, volume 30, pages 720–721, 1933. 92, 93

[135] A. Melton. Visitor bevavior in museum: some early research in environmental design. In *Human Factors*, volume 14, pages 393–403, 1972. 93

[136] T. B. Moeslund, A. Hilton, and V. Krger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(23):90 – 126, 2006. 19, 54, 74

[137] D. Montemerlo, S. Thrun, and W. Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 1, pages 695–701, 2002. 21, 34

[138] A. Morate. People detecting and tracking using laser and vision. Master's thesis, Royel Institute of Technology, Stockholm,Sweden, 2005. 21

[139] T. Mori, T. Matsumoto, M. Shimosaka, H. Noguchi, and T. Sato. Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filters. In *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008*, Marseille, France, Oct. 2008. Andrea Cavallaro and Hamid Aghajan. 23

[140] T. Moussouri. Automated visitor tracking and data analysis for research and evaluation in museums. In *Proceedings of the Visitor Studies Association Annual Conference*, Philadelphia, PA, 2005. 22, 33

[141] O. Mubin, C. J. Stevens, S. Shahid, A. Al Mahmud, and J.-J. Dong. A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1:209–0015, 2013. 13

[142] M. Mucientes and W. Burgard. Multiple hypothesis tracking of clusters of people. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 692–697, Oct 2006. 21

[143] L. Muller and T. Kalin. Analyse von besucherbewegungen in museen mittels rfid. In *Museum Aktuell*, 2009. 22, 33

[144] T. Murakita, T. Ikeda, and H. Ishiguro. Human tracking using floor sensors based on the markov chain monte carlo method. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 917–920 Vol.4, Aug 2004. 21, 23

[145] E. Murphy-Chutorian and M. Trivedi. Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):607–626, April 2009. 38

[146] B. Mutlu and J. Forlizzi. Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, HRI '08, pages 287–294, New York, NY, USA, 2008. ACM. 3

[147] S. Nair, G. Panin, M. Wojtczyk, C. Lenz, T. Friedlhuber, and A. Knoll. A multi-camera person tracking system for robotic applications in virtual reality tv studio. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3990–3996, Sept 2008. 20

[148] L. E. Navarro-Serment, C. Mertz, and M. Hebert. *Field and Service Robotics: Results of the 7th International Conference*, chapter Pedestrian Detection and Tracking Using Three-Dimensional LADAR Data, pages 103–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. 22

[149] I. R. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. An affective mobile robot educator with a full-time job. *Artificial Intelligence*, 114(12):95 – 124, 1999. 34

[150] I. R. Nourbakhsh, C. Kunz, and T. Willeke. The mobot museum robot installations: a five year experiment. In *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Nevada, USA, October 27 - November 1, 2003*, pages 3636–3641, 2003. 54

[151] K. Okada, S. Kagami, M. Inaba, and H. Inoue. Walking human avoidance and detection from a mobile robot using 3d depth flow. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 3, pages 2307–2312 vol.3, 2001. 19, 20

[152] T. Oyama, E. Yoshida, Y. Kobayashi, and Y. Kuno. Tracking visitors with sensor poles for robot's museum guide tour. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 645–650, June 2013. 20, 21, 25, 37, 55, 56, 74, 76, 82

[153] O. Ozturk, T. Yamasaki, and K. Aizawa. Estimating human body and head orientation change to detect visual attention direction. In *Proceedings of the 2010 International Conference on Computer Vision - Volume Part I*, ACCV'10, pages 410–419, Berlin, Heidelberg, 2011. Springer-Verlag. 19

[154] 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, pages 315–320, Sept 2006. 17

[155] F. L. Panagiotis .P, Ian D. et al. The herbert virtual museum. *Journal of Electrical and Computer Engineering*, pages 1–8, 2013. 75

[156] A. Panangadan, M. Mataric, and G. Sukhatme. Detecting anomalous human interactions using laser range-finders. In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2136–2141 vol.3, Sept 2004. 20

[157] E. Papaefstathiou, R. Sanchez, and N. Clinton. Providing remote gestural and voice input to a mobile robot, 2012. US Patent App. 13/154,468. 53

[158] E. PARK, Q. JENKINS, and X. JIANG. Measuring trust of human operators in new generation rescue robots. *Proceedings of the JFPS International Symposium on Fluid Power*, 2008(7-2):489–492, 2008. 1

[159] J. PARSONS. Would you eat in a restaurant with a robot waiter? this droid works 8-hour shifts and doesn't expect a tip. *The daily Mirror*, 2016. xii, 11

[160] S. Piérard, D. Leroy, J.-F. Hansen, and M. Van Droogenbroeck. *Advanced Concepts for Intelligent Vision Systems: 13th International Conference, ACIVS 2011, Ghent, Belgium, August 22-25, 2011. Proceedings*, chapter Estimation of Human Orientation in Images Captured with a Range Camera, pages 519–530. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. 22

[161] S. Piérard and M. Van Droogenbroeck. Estimation of human orientation based on silhouettes and machine learning principles. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, volume 2, pages 51–60, Vilamoura, Portugal, February 2012. 19

[162] D. Prattichizzo, M. Malvezzi, and A. Bicchi. On motion and force controllability of grasping hands with postural synergies. In *In Proceedings of Robotics: Science and Systems*, 2010. 15

[163] C. Premebida, O. Ludwig, and U. Nunes. Lidar and vision-based pedestrian detection system. *Journal of Field Robotics*, 26(9):696–711, 2009. 20, 23

[164] D. B. R. *The Social Robot*. PhD thesis, Department of Computer Science, Univ. College Dublin, 2000. 9

[165] C. Raghavachari, V. Aparna, S. Chithira, and V. Balasubramanian. A comparative study of vision based human detection techniques in people counting applications. *Procedia Computer Science*, 58:461 – 469, 2015. Second International Symposium on Computer Vision and the Internet (VisionNet15). 19

[166] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, New York, NY, USA, 1996. 31

[167] J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaeillie, and D. Lefeber. Expressing emotions with the social robot probo. *International Journal of Social Robotics*, 2(4):377–389, 2010. 1

[168] B. Scassellati, H. Admoni, and M. Mataric. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14(1):275–294, 2012. 12

[169] O. C. Schrempf, U. D. Hanebeck, A. J. Schmid, and H. Worn. A novel approach to proactive human-robot cooperation. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 555–560, Aug 2005. 32

[170] D. SCHULZ. People tracking with a mobile robot using sample-based joint probabilistic data association filters. *Int. J. of Robotics Research*, 22(2):99–116, 2003. 19, 20, 21

[171] F. Schler, J. Behley, V. Steinhage, D. Schulz, and A. B. Cremers. Person tracking in three-dimensional laser range data with explicit occlusion adaption. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1297–1303, May 2011. 22, 25

[172] J. Seyama and R. S. Nagayama. The effect of torso direction on the judgment of eye direction. *Visual Cognition*, 12:103116, 2005. 38

[173] M. Shiomi et al. Interactive humanoid robots for a science museum. *Intelligent Systems, IEEE*, 22(2):25–32, March 2007. 16, 20, 27, 34, 74

[174] M. Shiomi, T. Kanda, D. F. Glas, S. Satake, H. Ishiguro, and N. Hagita. Field trial of networked social robots in a shopping mall. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2846–2853, Oct 2009. 20, 27

[175] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 305–312, New York, NY, USA, 2006. ACM. 16, 20, 22, 33, 75

[176] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. A larger audience, please!: Encouraging people to listen to a guide robot. In *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, pages 31–38, Piscataway, NJ, USA, 2010. IEEE Press. 20, 27

[177] M. Shiomi, D. Sakamoto, T. Kanda, C. T. Ishi, H. Ishiguro, and N. Hagita. Field trial of a networked robot at a train station. *International Journal of Social Robotics*, 3(1):27–40, 2011. xii, 11, 15, 16, 20, 27, 54

[178] M. Shiomi, F. Zanlungo, K. Hayashi, and T. Kanda. Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *International Journal of Social Robotics*, 6(3):443–455, 2014. 19

[179] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah. Part-based multiple-person tracking with partial occlusion handling. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1815–1821, June 2012. 19

[180] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(12):140 – 164, 2005. 34

[181] R. Siegwart, K. Arras, B. Jensen, R. Philippsen, and N. Tomatis. Design, Implementation and Exploitation of a New Fully Autonomous Tour Guide Robot. In *None*, 2003. 15

[182] X. Song, H. Zhao, J. Cui, X. Shao, R. Shibasaki, and H. Zha. Fusion of laser and vision for multiple targets tracking via on-line learning. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 406–411, May 2010. 23

[183] L. Spinello, M. Luber, and K. O. Arras. Tracking people in 3D using a bottom-up top-down detector. In *Proc. of the Int. Conf. on Robotics & Automation (ICRA)*, Shanghai, China, 2011. 22

[184] B. Stephen. Chapter 17: Problems in visitor orientation and circulation. *Visitor Studies*, 1(1):155–170, 1988. 75

[185] B. Stephen. An analysis of visitor circulation: Movement patterns and the general value principle. *Curator: The Museum Journal*, 49(4):463–475, 2006. 75

[186] R. Stiefelhagen and J. Zhu. Head orientation and gaze direction in meetings. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 858–859, New York, NY, USA, 2002. ACM. 37, 38

[187] M. A. Svenstrup. *Mobile Robots in Human Environments; towards safe, comfortable and natural navigation*. PhD thesis, Computer Science and Automation Engineering, Universit degli Studi di Napoli Federico II, 06 2011. 14, 15

[188] Y. Takemasa, I. Kazuya, H. Yuki, and K. Shinichi. Attentional capture by change in direct gaze. *Journal of Perception*, 40:785–797, 2011. 37

[189] F. Tanaka, A. Cicourel, and J. R. Movellan. Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, 104(46):17954–17958, 2007. 1, 12, 53, 54

[190] M. Tappin. Getting annoyed by people invading your personal space is a genuine psychological problem, available: http://www.viralthread.com/getting-frustrated-by-people-encroaching-your-personal-space-is-a-genuine-disorder-2/, 2016. xii, 18

[191] C. A. Templeton. Museum visitor engagement through resonant,rich and interactive experiences. Master's thesis, School of Design, Carnegie Mellon University, 2011. 74

[192] T. THIAGO and D. . A. S. GERSHON. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. *Tech. Rep. 09-2010*, 1(1):1–41, September 2010. 19

[193] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Creemers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *International Journal of Robotics Research*, 19(11):972–999, 2000. 54, 75

[194] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Minerva: a second-generation museum tour-guide robot. In *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, volume 3, pages 1999–2005, 1999. 16, 20, 27, 34

[195] K. Tinworth. Identity and the museum visitor experience. by john h. falk. *Museum Anthropology*, 34(1):83–85, 2011. 75

[196] D. Todorovic. Geometrical basis of perception of gaze direction. *Vision Research*, 46(21):3549 – 3562, 2006. 37

[197] C. C. Tsai, S. M. Hsieh, Y. P. Hsu, and Y. S. Wang. Human-robot interaction of an active mobile robotic assistant in intelligent space environments. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 1953–1958, Oct 2009. 20

[198] E. Véron and M. Levasseur. *Ethnographie de lExposition.*, page 220. Bibliothèque publique d'information, Paris. Centre Georges Pompidou, 1983. 93

[199] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743 – 1759, 2009. Visual and multimodal analysis of human spontaneous behaviour:. 16

[200] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 40, 63

[201] K. Wada, T. Shibata, T. Saito, and K. Tanie. Effects of robot-assisted activity for elderly people and nurses at a day service center. *Proceedings of the IEEE*, 92(11):1780–1788, Nov 2004. 1

[202] J. Wainer, B. Robins, F. Amirabdollahian, and K. Dautenhahn. Using the humanoid robot kaspar to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development*, 6(3):183–199, Sept 2014. 12

[203] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. Boekhorst, C. Nehaniv, I. Werry, and D. Lee. Close encounters: spatial distances between people and a robot of mechanistic appearance. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, pages 450–455, Dec 2005. 17

[204] M. Wojtczyk, G. Panin, C. Lenz, T. Röder, S. Nair, E. Roth, A. Knoll, R. Heidemann, K. Joeris, C. Zhang, M. Burnett, and T. Monica. A vision based human robot interface for robotic walkthroughs in a biotech laboratory. In *Proceedings of the 4th ACM/IEEE International Conference on*

Human Robot Interaction, HRI '09, pages 309–310, New York, NY, USA, 2009. ACM. 20

[205] S. S. Yalowitz and K. Bronnenkant. Timing and tracking: Unlocking visitor behavior. *Visitor Studies*, 12(1):47–64, 2009. 33

[206] A. Yamazaki, K. Yamazaki, T. Ohyama, Y. Kobayashi, and Y. Kuno. A techno-sociological solution for designing a museum guide robot: Regarding choosing an appropriate visitor. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 309–316, March 2012. 30

[207] K. Yamazaki, M. Kawashima, Y. Kuno, N. Akiya, M. Burdelski, A. Yamazaki, and H. Kuzuoka. *ECSCW 2007: Proceedings of the 10th European Conference on Computer-Supported Cooperative Work, Limerick, Ireland, 24-28 September 2007*, chapter Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings, pages 61–78. Springer London, London, 2007. 31, 53

[208] K. Yamazaki, A. Yamazaki, M. Okada, Y. Kuno, Y. Kobayashi, Y. Hoshi, K. Pitsch, P. Luff, D. vom Lehn, and C. Heath. Revealing gauguin: Engaging visitors in robot guide's explanation in an art museum. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '09, pages 1437–1446, 2009. 30, 34, 75

[209] H. A. Yanco and J. Drury. A taxonomy for human-robot interaction. In *AAAI Technical Report FS-02-03*, Oct 2002. 13

[210] M. Yoda and Y. Shiota. Basic study on avoidance motions for human behaviors. In *Robot and Human Communication, 1995. RO-MAN'95 TOKYO, Proceedings., 4th IEEE International Workshop on*, pages 318–322, Jul 1995. 17

[211] M. Yoda and Y. Shiota. The mobile robot which passes a man. In *Robot and Human Communication, 1997. RO-MAN '97. Proceedings., 6th IEEE International Workshop on*, pages 112–117, Sep 1997. 17

[212] E. Yong. What are you looking at? people follow each others gazes, but without a tipping point, available: http://blogs.discovermagazine.com/notrocketscience/2012/04/23/what-

are-you-looking-at-people-follow-each-others-gazes-but-without-a-tipping-point/.V611N9J95ph, 2012. xii, 28

[213] M. A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki. *Intelligent Computing Technology: 8th International Conference, ICIC 2012, Huangshan, China, July 25-29, 2012. Proceedings*, chapter Development of a Mobile Museum Guide Robot That Can Configure Spatial Formation with Visitors, pages 423–432. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 34

[214] M. Zancanaro, T. Kuflik, Z. Boger, D. Goren-Bar, and D. Goldwasser. Analyzing museum visitors behavior patterns. In C. Conati, K. McCoy, and G. Paliouras, editors, *User Modeling 2007*, volume 4511 of *Lecture Notes in Computer Science*, pages 238–246. Springer Berlin Heidelberg, 2007. 94

[215] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, and L.-Q. Xu. Crowd analysis: a survey. *Machine Vision and Applications*, 19(5-6):345–357, 2008. 19

[216] H. Zhao and R. Shibasaki. A novel system for tracking pedestrians using multiple single-row laser-range scanners. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(2):283–291, March 2005. 20, 21